ROBUST MEDICAL IMAGE RECOGNITION AND SEGMENTATION

BY ZHENNAN YAN

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Computer Science

Written under the direction of

Dimitris N. Metaxas

and approved by

———————————————

———————————————

———————————————

———————————————

New Brunswick, New Jersey

OCTOBER, 2016

ABSTRACT OF THE DISSERTATION

ROBUST MEDICAL IMAGE RECOGNITION AND SEGMENTATION

by ZHENNAN YAN

Dissertation Director: Dimitris N. Metaxas

In recent decades, with increasing amount of medical data, clinical trials are designed and conducted to explore whether a medical strategy, treatment, or device is safe and effective for humans. In clinical trials, due to the large variance of collected image data and limited golden standard training samples, designing a robust and automated algorithm or framework for quantitative medical image analysis is still challenging and active field of research. Many state-of-the-art algorithms are designed/trained for specific anatomies or tasks with corresponding prior knowledge. In this dissertation, we focus on robust and easy-to-use solutions for two fundamental and key modules in medical image analysis, specifically, anatomy recognition and segmentation.

The medical image recognition is formulated as a classification problem to identify the body section from which the image is taken. The problem is solved by a patch-based convolutional neural network (CNN). The proposed method can utilize the image-level label to discover discriminative local patches without local annotations and train classifier using these local features. Its performance in our application is superior to conventional models using ad-hoc designed features as well as standard CNN. Accurate and efficient image recognition serves as a reliable initialization module for anatomy segmentation algorithms.

In medical image segmentation, precise labeling usually relies on prior knowledge due to ambiguous visual clues of different anatomies and between-subject variance. We use Gaussian Mixture Model and Markov Random Field to model the appearances and spatial relationships of voxels in medical image. To finely utilize the prior knowledge from training atlases (medical image and its corresponding label image), we design an adaptive statistical atlas based method to segment new subjects which could be very different from the training samples. The method is shown robust and accurate in brain segmentation and can be easily applied in other applications.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor, Professor Dimitris N. Metaxas, for his guidance and support of my Ph.D study and related research. He provided excellent research environment and collaborative opportunities to explore challenging tasks in medical image analysis. He taught me to think independently and guided me with his immense knowledge and extensive experiences to solve problems. None of the work in this dissertation would have been achieved without his support.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Kostas Bekris, Prof. Konstantinos Michmizos, and Prof. Xiaolei Huang, for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives.

My sincere thanks also goes to Prof. Shaoting Zhang for his mentorship and helps during my research. Besides, I would like to thank Dr. Yiqiang Zhan and Dr. Xiang Sean Zhou (Siemens Healthcare), who provided me an opportunity to join their team as intern. They have contributed many ideas and helps in my research. I also appreciate the helps from Prof. Albert Montillo and Prof. Kang Li during the collaborations. Without their precious support it would not be possible to conduct this research. Also thanks to other collaborators, especially to Chaowei Tan, Dr. Zhigang Peng, Dr. Shu Liao, Dr. Yoshihisa Shinagawa, Dr. Xiaofeng Liu, Dr. Boubakeur Belaroussi, Dr. Hui Jing Yu and Dr. Colin Miller.

I thank my fellow labmates in CBIM for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had over the last six years.

Last but not the least, I would like to thank my family: my parents and my wife for supporting me spiritually throughout writing this thesis and my life in general.

Dedication

This dissertation is dedicated to my wife, my parents and parents-in-law, for their endless love and support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

With the development of many advanced imaging and computing technologies, computer aided medical image analysis is fast growing in recent decades. Medical imaging, a process of creating visual representation of the interior of a body, is broadly used in many clinical diagnosis and surgeries. There is a new computerized medical image modality been introduced into clinical practice about every 10 years. In 1970s, computed tomography (CT) [1] was introduced. Then, magnetic resonance imaging (MRI) [2] was developed in the 80s, and PET [3] in the 90s. These days, the combined modalities, e.g. PET/CT [4], PET/MRI [5, 6] etc., are used in some applications to produce better visual representation, or new ways to use existing hardware for functional analysis, e.g. functional MRI. Therefore, the problem of medical image analysis is been redefined every 10 years and requires a fresh look based on the new advantages/limitations that each new modality brings to medical diagnosis.

Numerous algorithms have been developed [7, 8, 9, 10] to assist clinicians or researchers to interpret and assess medical images in different applications, from fundamental tasks, e.g. anatomical landmark detection [11, 12, 13] and organ segmentation [14, 15, 16, 17], to complicated computer aided diagnosis (CAD) systems [18, 19, 20] in clinical studies and also in surgery. However, since different organ systems have highly diverse characteristics and different imaging techniques are used in studies, medical image analysis models are usually trained or designed for specific anatomies or applications to incorporate appropriate prior knowledge. In clinical research, where a large number of medical image data is collected with large variance and different artifacts, designing a robust and automated framework or pipeline [21] is still a challenging and active field of research.

1.1 Background

There are many different diseases and disorders affecting people worldwide. For instance, Alzhemimer's disease (AD) is the sixth leading cause of death in the United States, and an estimated 5.4 million Americans of all ages have AD in 2016 (of which, about 5.2 million people are age 65 and older) [22]. The report [22] also states that it is the only disease among the top 10 causes of death in America that cannot be prevented, cured or even slowed. Another example, osteoarthritis, the most common chronic condition of the joints (e.g. knee, elbow, etc.), is a leading cause of pain and disability and affects many people particularly older adults. According to the Agency for Healthcare Research and Quality, more than 600,000 people undergo knee replacement surgeries every year in United States [23]. By 2030, the annual demand for primary total knee arthroplasties is projected to grow to 3.48 million procedures in the U.S. [24]. Accurate and timely diagnosis and proper treatment is critical. Therefore, to explore whether a medical strategy, treatment, or device is safe and effective for humans, researchers conduct studies involving a large set of medical data. Such research studies are called clinical trials.

In clinical trials, structural medical images are usually collected for diseases associated with changes in the size and shape of anatomical structures. The anatomical changes can be quantified after delineating the structures in the medical images. For example, ventricles and cortex measurements in brain can be used in Alzheimer's disease analysis [25, 26, 27, 28]; the bone and cartilage measurements can help in knee osteoarthritis studies [29, 30]. X-ray CT and MRI are two of the most commonly used techniques to acquire image scans as a series of cross-sectional image slices. Each has its own advantage and disadvantage. CT can produce quick and painless scan for a wide range of conditions, in particular for lung and bone related analysis. However, it uses radioisotopes which may cause other risks or problems. MRI is particularly sensitive to soft tissue injuries and neurologic pathologies, although it can be a lengthy and noisy procedure. It is ideal to help evaluate structural pathologies because of a high quality definition. It is also able to evaluate muscle, fat, tendon, cartilage, and bone marrow

pathology [31, 32, 29, 33]. Accurate delineation of anatomy of interest from the medical images is prerequisite for quantitative assessment in clinical researches, especially large data analysis.

Although doctors or clinical experts can manually delineate regions of interest (ROI) in medical images, it is low-efficient, high-cost and error-prone in a clinical trial which may involve a huge number of samples. Therefore, an automated, robust and accurate medical image analysis framework is essential for such large data analysis. Due to various characteristics of different organ systems, different framework involving multiple processing steps is designed for specific anatomy.

Among different processing steps, efficient anatomy recognition and robust image segmentation are two fundamental steps in most medical image analysis frameworks. Anatomy recognition module is used to identify anatomies contained in medical images. Through recognition, algorithms can understand the image content and locate the region of interest in images to initialize the segmentation modules for object delineation. The image segmentation methods can utilize intensities, shape, relative position and other prior knowledge to extract desired object boundaries from the ROI. Although different recognition and segmentation algorithms have been proposed in the recent decades, it is still challenging to get robust and accurate results due to several major concerns. First, medical image data set in clinical trials has large variance because of its collection from different institutes using equipments from different vendors and different imaging protocols. Second, all medical images contain different visual artifacts [34, 35, 36], which can affect the accuracy of intensity based algorithms. Third, each screened body is different. The same anatomy may have different size, shape, location in different bodies. What's more, even for the same patient, anatomical changes can be observed at different age or in pathological conditions. Moreover, limited amount of manually labeled ground truth can not provide sufficient training samples. Owing to the variety of data in clinical trials and limited amount of training samples, many learning based segmentation methods are hardly applied.

1.2   Organization

In the following chapters, we focus on two fundamental modules in typical medical image analysis framework, namely, anatomy recognition and segmentation. The methods are designed to be robust and scalable for two commonly used medical image modalities: CT and MRI.

In Chapter 2, we describe the main contribution of our first work, which is a deep learning based anatomy recognition algorithm. It aims to efficiently identify the anatomies contained in the cross-sectional medical images. To make the method robust to large variances in medical images, we use deep learning to learn hierarchical features from the data. Considering that the learning based methods require large amount of consistent training samples, we design a 2D slice based recognition algorithm, which could be easily extended to process volumetric images. To avoid the tedious labeling, training and error-prone online inference requirements of conventional object detection based approaches, we formulate the problem as a multi-class image classification task and use a two-stage deep learning framework to discover and utilize the discriminative local information to train the classifier.

With the result of the recognition module, we can easily initialize appropriate segmentation algorithm to delineate desired organs or structures. In Chapter 3, we propose a 3D image segmentation method, which is based on statistical and adaptive atlas. Considering the limited training samples in clinical trials, our adaptive atlas is designed to finely utilize the prior information in the training dataset and adapt to the target subject to obtain a robust and accurate segmentation result. The segmentation is formulated as an optimization task to label each image voxel as one of the classes. It is solved by an expectation-maximization (EM) algorithm. We can use this method to label the multiple structures simultaneously in human brains with Alzhemimer's disease using healthy training samples. However, the proposed method assumes the intensity of each class follows a Gaussian distribution, which may not hold in single object segmentation or binary classification where the background class needs complex modeling.

In Chapter 4, another application of liver segmentation is discussed. In this single

object segmentation, we utilize the atlas-based prior knowledge and incorporate into an improved deformable mode for robust result. We demonstrate that even with limited training data, the proposed method can still achieve promising results for robust segmentation in clinical researches. Based on the this framework, we also develop an intuitive correction method to refine the segmentation result interactively. An expert can use it to efficiently edit the segmentation result for quality assurance. At last, we conclude the proposed modules in Chapter 5, and discuss future work to produce robust results towards large scale medical image data analytics.

Chapter 2

Deep Learning based Medical Image Recognition

2.1 Introduction

CT and MR are two common forms of medical imaging scans. A CT/MR sequence is typically a 3D volume image consisted of a series of cross-sectional 2D slices. With the availability of more and more intelligent medical image analysis algorithms, radiologists hope that medical images could have been "pre-processed" by all applicable automatic algorithms before being loaded for visual evaluation. In this way, the automatic results can be displayed instantaneously in the reading room to speed up the reading process. Since automatic algorithms are usually trained or designed for specific anatomies, a robust image recognition algorithm becomes important to gate the intelligent algorithms properly for meaningful results. In this scenario, it is important to have a reliable automatic module that can identify the human body part contained in the medical image in the first place. Such a program can act like a medical professional to tell quickly different anatomies and their rough positions in an image at a glance. With the correct first impression, automated work flows can conduct meaningful pre-processing and other higher level tasks (e.g. anatomy detection or segmentation algorithms) using appropriate methods or models. Moreover, given the body part information, the search range of the following detection/segmentation algorithms can be reduced, hence, the algorithm speed and robustness are improved.

In fact, medical image recognition can also help the anatomy-based query in picture archiving and communication system (PACS). A PACS allows a healthcare organization to store and retrieve medical images conveniently. It is worth noting that although DICOM header includes body part information, text-based retrieval is not an ideal choice due to three major challenges. First, it may contain about 15% errors in DICOM
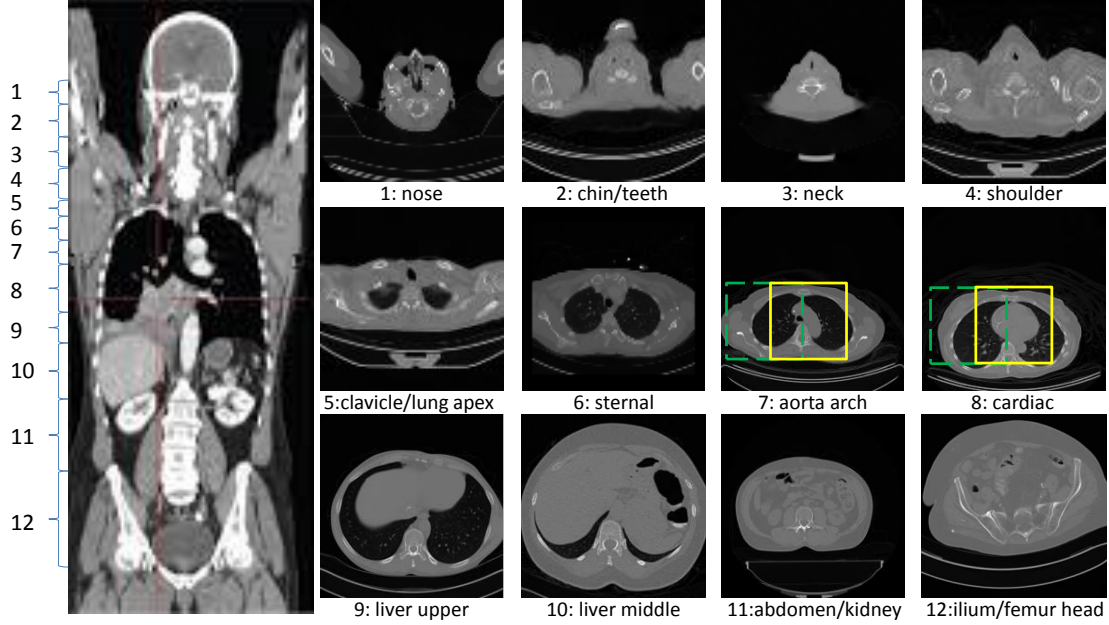
Figure 2.1: Definition of body sections. Human body is divided into 12 continuous parts. Each parts may cover different ranges due to the variability of anatomies. Yellow boxes indicate the discriminative local regions in aorta arch class and cardiac class, while green boxes indicating ambiguous local regions for this classification task.

headers [37]. Second, text information in DICOM is sometimes too abstract to precisely describe the anatomies contained in the scan. For example, it is difficult to tell if a scan with DICOM tag (0018,0015)="TSPINE" includes the mid-part of the liver. Third, the multi-language supporting of DICOM header becomes another barrier for text-based retrieval. On the contrary, a reliable image-based anatomy recognition algorithm can tackle all these three challenges by learning the intrinsic anatomical appearance information to enable content-based image retrieval [38, 39] and improve the retrieval precision.

Although organ segmentation and landmark detection topics have been extensively investigated, the automatic medical image recognition (identify the anatomies contained in the medical image) is not well explored. In this chapter, we introduce the human body parts recognition task in 2D image slice, namely "slice-based body part recognition" [40]. Specifically, we divide human body into continuous sections according to anatomical context as shown in Fig. 2.1. Given a 2D transversal slice, the task of a slice-based body part recognition algorithm is to identify which section the slice belongs

to. This is basically a multi-class image classification problem. Although 3D volume always contain more comprehensive anatomy information, based on which the body part recognition could be more accurate, this study only aims at slice-based body part recognition for three reasons. First, 2D slice-based body part recognition provides the foundation of 3D body part identification. Given the body part identities of all slices of a 3D volume, the 3D body part can be straightforwardly derived. Second, training a 3D model may need a large number of consistent and labeled 3D data, which is usually unavailable in clinical studies. Third, in some real-world systems, 3D volume is not always accessible. For example, in a client-server application, the server end might only receive the 3D volume data slice-by-slice due to the limited network speed but need to output body part information instantaneously.

Multi-class image classification has been extensively studied for decades in computer vision and machine learning communities. In general, image classification algorithms consist of feature extraction and classification modules. Conventionally, different hand-craft features based on prior or domain knowledge are used with various classifiers, e.g. SVM [41], logistic regression [42], random forest [43], etc. In slice-based body part recognition, it is difficult to "design" common features that work well for different body sections, due to diverse appearances in different body sections and large variability between subjects. Thus, deep learning technology, which learns features and classifiers simultaneously, becomes a promising solution. Instead of designing any features, those algorithms [44, 45] aim to learn both the features and classifiers jointly from the data. As the features are learned for specific image classification tasks, they often have more discriminative power, hence, achieve better classification performance than those ad-hoc designed ones.

However, slice-based body part recognition has its unique challenge which might not be solved by standard deep learning. As shown in Fig. 2.1, although image 7 and 8 come from aorta arch and cardiac sections, respectively, their global appearance characteristics are quite similar. For these two slices, the discriminative information only resides in the local mediastinum region (indicated by the yellow boxes). The rest areas are just "non-informative" for classification. Although the standard deep learning

framework is able to learn some low-level and abstract features from global context, it cannot learn local patches that are most discriminative for body part recognition. The "non-informative" regions here may mislead the classifier to recognize these two sections as identical. Hence, the classification power of the learned deep network may be limited.

To leverage local information but avoid the tedious local annotations, we present a two-stage deep learning method for body part recognition task [40, 46] in this chapter. Specifically, in the first stage (namely pre-training stage), a convolutional neural network (CNN) is learned in a multi-instance learning fashion to "discover" the most discriminative and non-informative local patches. In the second stage (fine-tuning stage), we select discriminative and non-informative local patches for each class based on the responses of the pre-trained CNN. These selected local instances are used to fine tune the pre-trained CNN. At run-time, a sliding window approach is employed to apply the fine-tuned CNN to the subject image. As the CNN is sensitive to the discriminative local patches, it essentially identifies body part by focusing on the most distinctive local information and discarding non-informative local regions.

2.2 Related Work

Several body part recognition systems in medical imaging domain has been introduced. Park et al. [47] proposed an algorithm to determine the body parts using energy information from Wavelet Transform. Look-up tables are designed to classify the imaging modality and body parts. Hong et.al. [48] proposed a framework to identify different body parts from a whole-body scan. The method starts from establishing global reference frame and the head location. After determining the bounding box of the head, other body parts including neck, thorax cage, abdomen and pelvis, are localized one by one using different algorithms. In general, these approaches employ ad-hoc designed features and algorithms to identify major body parts which have globally variant appearances. Recently, more learning-based approaches are proposed for body part recognition. All of these methods essentially resort to the detection of specific organs or landmarks. In [49], Zhan et al. trained multiple, organ-specific classifiers and optimize the schedule of organ detections based on information theory. In [50], Criminisi et al.

utilized regression forests for anatomy detection and localization and obtained better accuracy than their classification approach in [51]. In [52], Donner et al. also trained regressor for anatomical landmark detection. However, since these organ/landmark-based recognition methods rely on a number of organ/landmark detectors, large efforts of manual annotation are required in the training stage. Besides, complex inferences needed to resolve the correlation between results from different detectors may prone to error.

Technically, slice-based body part recognition is an image classification problem, which has been extensively studied in recent decades. Feature extraction from images is essential for the subsequent learning and generalization steps. Conventionally, features, such as SIFT [53], Histogram of Oriented Gradients (HOG) [54] and their variants, are carefully designed or selected for specific applications with domain knowledge. With the latest advances of machine learning technology, deep learning [55, 56] methods with automatic feature learning ability has shown successful outcomes in different applications, including signal processing [57], image/object recognition [44, 58], natural language processing [59], etc. The deep network architecture uses multiple layers of simple but nonlinear activation functions to transform the input data to multiple levels of feature representation, from low (detail) level to high (abstract) level. The network is able to learn such hierarchical feature representations in unsupervised or supervised way from a large amount of training data by itself. Among different deep learning methods, convolutional neural network (CNN) [60] based algorithms are more suitable in image related tasks, since images have highly correlated intensities in local regions and some local signals or statistics are invariant to location. So far, different CNN based methods [44, 61, 45] have shown their superiority in image classification tasks compared to conventional approaches with carefully designed/selected features. Despite its success, its application in medical image analysis remains to be fully explored.

Recently, deep learning methods have been applied in CAD tasks, including detection [62, 63, 64], segmentation [65, 66, 67, 68], disease classification [69], etc. They are proposed for specific anatomies and tasks. As discussed before, it would be useful to

have an image-based anatomy recognition method in the CAD system as a preprocessing module. Roth et.al. [70] proposed a method for anatomy-specific classification of medical images using CNN. They applied a standard deep CNN on 2D axial CT images to classify 5 anatomies (neck, lungs, liver, pelvis and legs) and obtained the state-of-the-art accuracy (5.9% error). In this slice-based anatomy recognition, the standard CNN is conducted as a *global* learning scheme, which takes the entire image as input. The CNN successfully learned feature representations to capture the diverse appearances in the five body sections. However, this standard CNN is not easily scalable to handle more detailed anatomy recognition (as shown in Fig. 2.1) effectively, since distinctive information often comes from *local* patches and these local patches are distributed "inconsistently" at different positions of the slices. As shown in Fig. 2.1, aorta arch section and cardiac section have globally similar appearance, while the discriminative information only resides in the local mediastinum region (indicated by the yellow boxes). The rest areas are just "non-informative" or misleading for classification purpose. One may argue that CNN can still learn local features through its convolutional layers. However, this situation only holds while local features always appear at the similar location across different images, which is not the case of body part recognition.

In fact, this problem also exists in general image classification/recognition applications in computer vision. Researchers are trying to leveraging local region information to train CNN for recognition or classification. For example, in face recognition [71], the authors first detect and align face regions properly before training CNN. In [61], Szegedy et al. utilized CNN for local object detection and recognition and achieves state-of-the-art performance on Pascal VOC database. Similarly, Girshick [72] trained a region-based CNN (fast R-CNN) based on existing object proposals and used a multi-task loss function to learn the classifier and bounding-box regressor for efficient object detection. Then, Ren et.al. [73] proposed a faster R-CNN detection by using a region proposal network which shares convolutional features with the detection network. Despite promising results these methods generate, they all require manual annotations of local regions of objects in images for training. However, the discriminative local regions for body section recognition are not easy to define, not to mention that the effort to

build these local detectors might be quite large.

To avoid explicit local region or object annotation, different approaches have been emerged. Singh et al. [74] used unsupervised clustering method and one-vs-all linear SVM to train classifier for each cluster to discover the discriminative patches which can be used as visual words in spatial pyramid based classification. In another pioneer work, Wei et al. [45] applied an existing objectness detector [75] to produce some local region proposals from a given image, and used them to train multi-label CNN classifier. Recently, several studies incorporate multi-instance learning (MIL) [76, 77, 78] with CNN to better utilize local information in weakly supervised learning fashion. For example, Wu et al. proposed a deep multi-instance learning framework in a weakly supervised setting for image classification and auto-annotation based on object and keyword proposals [79]. Pinheiro et al. combined CNN and MIL to do pixel labeling on Pascal VOC dataset [80]. However, their methods are built over other existing models.

Here, we present a two-stage multi-instance deep learning framework for fine-grained body part recognition [40, 46]. Compared with above mentioned approaches, our method does not require local annotation or any existing model. It only requires "weak" supervision at a global level, i.e., body part labels at image level. It is designed to discover the discriminative "local" information by itself. In this way, the annotation efforts in the training stage are dramatically reduced. This is in particular meaningful for medical image applications, since the annotations in medical images often require clinical expertise and high cost.

2.3 Methodology

We present a two-stage deep learning framework for slice-based body part recognition in this section. We start by the problem statement and notation definitions. The first learning stage is introduced in Sec. 2.3.2, which aims to learn representative local image features in a supervised multi-instance fashion. Then we describe the second learning stage in Sec. 2.3.3, in which some discriminative and non-informative local patches are extracted from images and used to fine-tune a patch-based CNN classifier initialized

by the first learning stage. In Sec. 2.3.4, we discuss the run-time image classification strategy using the find-tuned CNN. At last, implementation details are discussed in Sec. 2.3.5.

2.3.1 Problem Statement

Definitions: Slice-based body part recognition is a typical multi-class image classification problem for a learning algorithm. Denote $\mathbf{X}$ as the input slice/image, K as the number of body sections (classes), and $l \in \{1, ..., K\}$ as the corresponding class label of $\mathbf{X}$. The learning algorithm aims to find a function $\mathcal{O} : \mathbf{X} \to l$. In traditional image classification frameworks, $\mathcal{O}$ is often defined as $\mathbb{C}(\mathbb{F}(\mathbf{X}))$, where $\mathbb{F}(\mathbf{X})$ and $\mathbb{C}(.)$ denote the feature extractors and classifiers, respectively.

In the context of convolutional neural network (CNN), $\mathcal{O}$ becomes a multi-layer neural network. An example of standard CNN is shown in Fig. 2.2 (similar to LeNet [60]), it has two convolutional layers (C1, C3), each followed by a max-pooling layer (S2, S4), one fully connected hidden layer (H5) receiving outputs of the last pooling layer, and one logistic regression (LR) layer (O6) as the output layer. In CNN, $\mathbb{F}(\mathbf{X})$ becomes multiple nonlinear layers, which aim to extract image features in a local-to-global fashion. $\mathbb{C}(.)$ is implemented by the LR layer, whose output is a K-dimension vector $R(k), k \in \{1, ..., K\}$ representing the probability of $\mathbf{X}$ belonging to each class k. Mathematically, $R(k)$ can be described as a conditional probability as $R(k) = \mathbf{P}(k|\mathbf{X}; \mathbf{W})$. Here, $\mathbf{W}$ denote the CNN parameters, including the weights in all layers. The final predicted label l is determined by the argument of the maximum element (class with the highest probability) in R.

Given a set of training images $\mathcal{T} = \{\mathbf{X}_m, m = 1, ..., M\}$, with corresponding discrete labels $l_m \in \{1, ..., K\}$, the training algorithm of CNN aims to minimize the loss function:

$$L_1(\mathbf{W}) = \sum_{\mathbf{X}_m \in \mathcal{T}} -\log(\mathbf{P}(l_m|\mathbf{X}_m; \mathbf{W})). \tag{2.1}$$

where $\mathbf{P}(l_m|\mathbf{X}_m; \mathbf{W})$ indicates the probability of image $\mathbf{X}_m$ being correctly classified as class l_m using parameters $\mathbf{W}$.
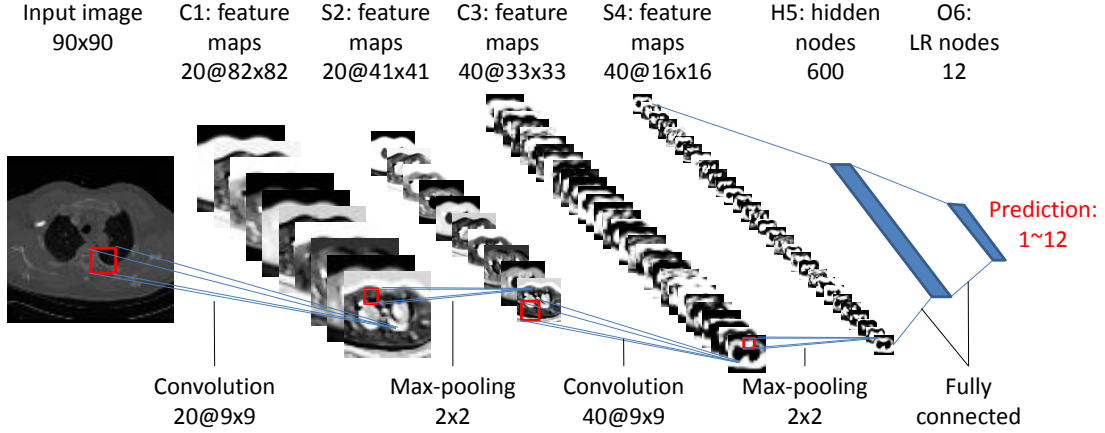
Figure 2.2: Illustration of one standard CNN architecture and the outputs of each layer.

CNN has shown impressive performance in image classification tasks [44, 45]. In these successful applications, standard CNN is conducted as a *global* learning scheme, which takes the entire image as input. To leverage the local information, more important, to automatically "discover" the discriminative local patches for different body sections, we design a two-stage CNN learning framework. It consists of pre-training and fine-tuning stages, which will be detailed next.

2.3.2   Learning Stage I: Multi-instance CNN Pre-training

In order to exploit the representative features for specific object which is usually located in a local area of an image, CNN should take *discriminative* local patches instead of the entire slice as its input. Here, the key problem is how to automatically *discover* these local patches through learning. This is the major task of our first learning stage. A multi-instance learning strategy is designed to achieve this goal.

Given a training set $\mathcal{T} = \{\mathbf{X}_m, m = 1, ..., M\}$, each training image $\mathbf{X}_m$, with corresponding labels l_m, is divided into a set of local patches defined as $\mathcal{L}(\mathbf{X}_m) = \{\mathbf{x}_{mn}, n = 1, ..., N\}$. These local patches become the basic training samples of the CNN and their labels are inherited from the original images, i.e., all $\mathbf{x}_{mn} \in \mathcal{L}(\mathbf{X}_m)$ share the same label l_m. While the structure of CNN is still the same as the standard one, the loss function is adapted as:

$$L_2(\mathbf{W}) = \sum_{\mathbf{X}_m \in \mathcal{T}} -\log(\max_{\mathbf{x}_{mn} \in \mathcal{L}(\mathbf{X}_m)} \mathbf{P}(l_m | \mathbf{x}_{mn}; \mathbf{W})), \qquad (2.2)$$
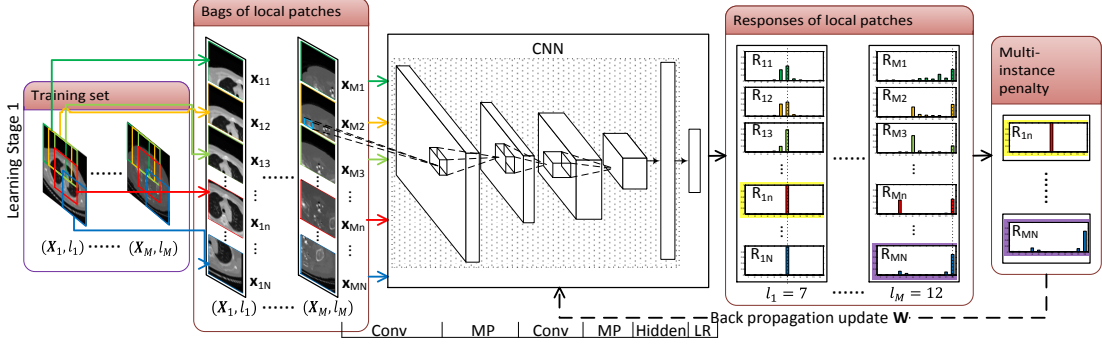
Figure 2.3: Illustration of pre-train stage. In this stage, the CNN is trained in a multi-instance fashion. Here, yellow highlighted response of instance $\mathbf{x}_{1n}$ from image $\mathbf{X}_1$ and purple highlighted response of instance $\mathbf{x}_{MN}$ from image $\mathbf{X}_M$ are selected to compute the loss to update the CNN parameters in backward propagation. They are picked because they have higher response on the correct label than other instances from the same image. Those local patches which can be easily and correctly classified are considered as the discriminative information for image classification.

where $\mathbf{P}(l_m|\mathbf{x}_{mn}; \mathbf{W})$ is the probability that the local patch $\mathbf{x}_{mn}$ is correctly classified as l_m by CNN with parameters $\mathbf{W}$.

The new loss function is different from Eq. (2.1) by adopting a multi-instance learning criterion. Specifically, each original slice $\mathbf{X}_m$ is treated as a bag consisting of multiple instances (local patches), $\{\mathbf{x}_{mn}\}$. The CNN receives a batch of labeled slices (bags) for training. The loss function Eq. (2.1) is adapted that as long as one local patch (instance) is correctly labeled, the class of corresponding slice (bag) is considered to be correct. Thus, within each bag (slice), only the instance with the highest probability to be correctly classified is counted in the loss function. Such instance is considered as the most discriminative local patch of the image slice. Let R_{mn} be the output vector of the CNN on local patch $\mathbf{x}_{mn}$. The l_mth component of R_{mn} represents the probability of $\mathbf{x}_{mn}$ being correctly classified. As illustrated in Fig. 2.3, for each training image $\mathbf{X}_m$, only the local patch that has the highest response at the l_mth component of R_{mn} (indicated by the yellow and purple boxes for two training images, respectively), contributes to the loss function and drives the update of network parameters $\mathbf{W}$ during the backward propagation. In this way, the pre-trained CNN will be more sensitive to the discriminative local patches than other local regions. In other words, the most discriminative local patches for each image class can be automatically *discovered* after
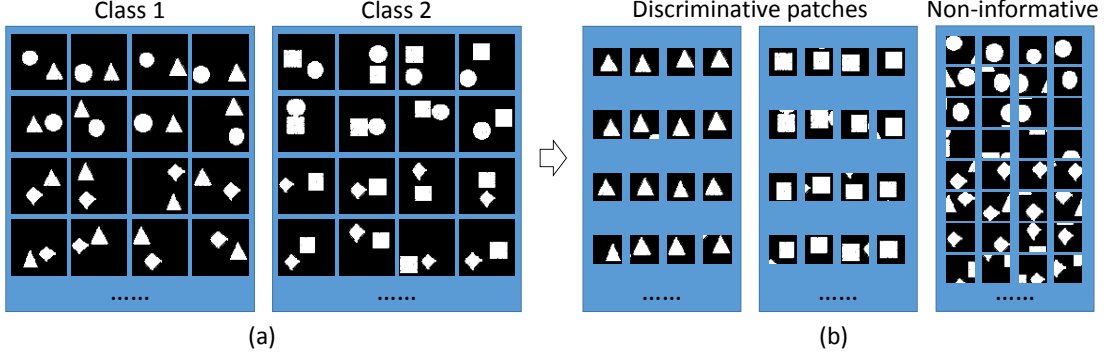
Figure 2.4: A synthetic toy example. (a) Synthetic images of two classes. (b) The discriminative and non-informative local patches selected by the pre-trained CNN model. Note that we never "tell" the algorithm that these two classes are differentiable by triangle and square.

the CNN training.

We design a toy example to illustrate this discovery ability. As shown in Fig. 2.4(a), four types of geometry elements, square, circle, triangle and diamond are randomly positioned and combined to generate two classes of gray images. While circle and diamond are allowed to appear in any classes, triangle and square are exclusively owned by Class1 and Class2, respectively (ref Sec. 2.4.1 for more details). Fig. 2.4(b) shows the discovered discriminative patches (containing triangle or square) for the image classification task in toy example. This is exactly in accordance to the fact that these two classes are only distinguishable by "triangle" and "square". It proves that our method is able to *discover* the key local patches without manual annotation. Of course, this problem would become trivial if we have the prior knowledge of the discriminative local patches and build specific classifiers on them. However, in real-world recognition tasks, it is not easy to figure out the most discriminative local patches for different classes. In addition, even with *adhoc* knowledge, annotating local patches and training local classifiers often requires large effort. The trivial solution thus becomes non-scalable when body-sections are re-defined or the imaging modalities are changed.

To ensure that the learned CNN will have stable high responses on discriminative local patches, a spatial constraint is further incorporated into the loss function as:

$$L_3(\mathbf{W}) = \sum_{\mathbf{X}_m \in \mathcal{T}} -\log(\max_{\mathbf{x}_{mn} \in \mathcal{L}(\mathbf{X}_m)} \sum_{\mathbf{x} \in \mathfrak{N}(\mathbf{x}_{mn})} \mathbf{P}(l_m | \mathbf{x}; \mathbf{W})), \quad (2.3)$$
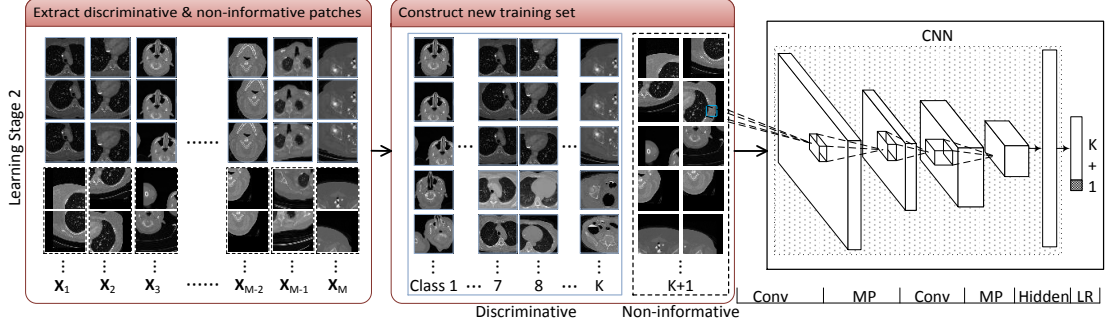
Figure 2.5: Illustration of fine-tuning stage. In this stage, CNN architecture is modified by adding a non-informative class in output layer. The parameters are inherited from the pre-trained CNN and fine-tuned using the discriminative and non-informative local regions extracted from each class by the pre-trained CNN.

Here, $\mathfrak{N}(\mathbf{x}_{mn})$ denotes the local neighboring patches of $\mathbf{x}_{mn}$. Based on Eq. (2.3), for each training slice, the local patch to be considered in the loss function is not the most *individually* discriminative one, but the one whose neighboring patches and itself are *overall* most discriminative. In this way, the selected discriminative local patches will be more robust to image translations and artifacts.

2.3.3 Learning Stage II: CNN Fine-tuning

In the second learning stage, the main task is to fine-tune the pre-trained CNN using selected local patches, which is illustrated in Fig. 2.5.

The first type of selected local patches are the discriminative ones, i.e., these local patches on which the pre-trained CNN have high responses at the corresponding classes. For each image $\mathbf{X}_m$, we select D discriminative local patches as:

$$\mathbf{A}_m = \operatorname*{argmax}_{\mathbf{x}_{mn} \in \mathcal{L}(\mathbf{X}_m)} \mathbf{P}(l_m | \mathbf{x}_{mn}; \hat{\mathbf{W}}). \tag{2.4}$$

Here, $\hat{\mathbf{W}}$ is the parameter set of the pre-trained CNN. $\operatorname{argmax}_D(.)$ is the operator that returns the arguments of the largest D elements.

We noticed that apart from the discriminative local patches, the remaining regions cannot be completely ignored for two reasons. First, only selecting discriminative patches to boost classifier may lead to overfitting problems. Second, some "confusing" local patches may mislead the body part recognition. For example, the patches

containing lung regions (green dashed boxes in Fig. 2.1) appear in both aortic arch and cardiac sections. For these "confusing" patches, CNN may generate similarly high responses for both aortic arch and cardiac classes. (Note that since the pre-trained CNN is only ensured to correctly classify one local patch per slice, the responses of the remaining patches are not guaranteed.) At run-time, when CNN is applied to the confusing patches, the high responses on multiple classes may cause wrong body part identification. Therefore, the algorithm should collect these "confusing" regions as the second type of local patches in fine-tune stage to suppress their responses for *all* desired classes (body sections).

To this end, we introduce a new "non-informative" class (patches in dashed box in Fig. 2.5) besides the existing training classes. This class includes two kinds of local patches: 1) local patches misclassified by the pre-trained CNN, and 2) local patches where the pre-trained CNN has "flat" responses across all classes. Denote $\mathbf{P}(k|\mathbf{x}_{mn}; \hat{\mathbf{W}})$ as the kth output of the pre-trained CNN on $\mathbf{x}_{mn}$, i.e., the probability of $\mathbf{x}_{mn}$ belonging to class k, the non-informative local patches of a training slice $\mathbf{X}_m$ are defined as:

$$
\begin{aligned}
\mathbf{B}_m = \{\mathbf{x}_{mn}| \operatorname*{argmax}_{k \in \{1,\dots,K\}} \mathbf{P}(k|\mathbf{x}_{mn}; \hat{\mathbf{W}}) \neq l_m\} \\
\cup \{\mathbf{x}_{mn}| \operatorname*{entropy}_{k \in \{1,\dots,K\}} \mathbf{P}(k|\mathbf{x}_{mn}; \hat{\mathbf{W}}) > \theta\}
\end{aligned}
\tag{2.5}
$$

Recall the toy example, Fig. 2.4(b) shows the selected discriminative and non-informative local patches. The non-informative patches including circle, diamond or black background are just common and misleading components for the classification problem.

After introducing the additional non-informative class, we keep the CNN structure same as the pre-trained CNN, except adding an additional output (shadowed box in the rightmost part of Fig. 2.5) in the LR layer and the corresponding connections to the hidden layer. Since the pre-trained CNN already captured some discriminative local appearance characteristics, all network layers except the last one are initialized by inheriting their parameters from the pre-trained CNN. These parameters are fine

tuned by minimizing Eq. (2.6):

$$L_4(\mathbf{W}) = \sum_{\mathbf{x} \in \mathbf{A} \bigcup \mathbf{B}} -\log(\mathbf{P}(l|\mathbf{x}; \mathbf{W})), \qquad (2.6)$$

Here, $\mathbf{A} = \bigcup_{\{m=1,...,M\}} \mathbf{A}_m$ and $\mathbf{B} = \bigcup_{\{m=1,...,M\}} \mathbf{B}_m$ denote the discriminative and non-informative local patches selected from all training images, respectively. Note that since the non-informative local patches are not belonging to any body section class now, their responses on any body section class can be effectively suppressed during the CNN fine-tuning stage.

2.3.4 Run-time Classification

The two-stage CNN learning algorithm is summarized as follows.

Algorithm 1 Two-stage multi-instance deep learning

Input:
   Scalars M, N, K, dataset $(\mathbf{X}_m, l_m), \forall m \in \{1, \cdots, M\}$, CNN architecture
Output:
   Fine-tuned CNN parameters $\mathbf{W}^{opt}$
   1: Partition $\mathbf{X}_m$ into N overlapping local regions $\mathbf{x}_{mn}$
   2: Pre-train CNN on $(\mathbf{x}_{mn}, l_m)$ using multi-instance loss function (2.3), and obtain optimized $\hat{\mathbf{W}}$
   3: Extract $\mathbf{A}_m$ and $\mathbf{B}_m$ according to Eq. (2.4) and (2.5)
   4: Assign label l_m to each instance of $\mathbf{A}_m$, and label $K+1$ to each of $\mathbf{B}_m$
   5: Modify pre-trained CNN by adding one unit to LR; layers except LR inherit parameters from $\hat{\mathbf{W}}$
   6: Fine-tune CNN on image patch set $\mathbf{A} \cup \mathbf{B}$ using loss function (2.6), and obtain optimized $\mathbf{W}^{opt}$

At runtime, the fine-tuned CNN with parameter $\mathbf{W}^{opt}$, optimized from Eq. (2.6), is applied for body part recognition in a sliding window fashion. The sliding window partitions a testing image $\mathbf{X}$ into N overlapping local patches $\mathcal{L}(\mathbf{X}) = \{\mathbf{x}_n, n = 1, ..., N\}$. For each local patch $\mathbf{x}_n$, the CNN outputs a response vector with $K+1$ components $\{\mathbf{P}(k|\mathbf{x}_n; \mathbf{W}^{opt})\}$. The class of the local patch $\mathbf{x}_n$ is then determined as:

$$c(\mathbf{x}_n) = \underset{k \in \{1,...,K+1\}}{\operatorname{argmax}} \mathbf{P}(k|\mathbf{x}_n; \mathbf{W}^{opt}). \qquad (2.7)$$

Since the class $K+1$ is an artificially constructed non-informative one, local patches belong to this class are simply ignored in body section determination. The class of the

testing slice $\mathbf{X}$ can be predicted by its most discriminative patch $\mathbf{x}_{n^*}$, which has the most peaky response on some label (excluding the non-informative):

$$C(\mathbf{X}) = c(\mathbf{x}_{n^*}), \qquad (2.8)$$

$$\mathbf{x}_{n^*} = \operatorname*{argmax}_{\mathbf{x}_n \in \mathcal{L}(\mathbf{X}); c(\mathbf{x}_n) \neq K+1} \mathbf{P}(c(\mathbf{x}_n)|\mathbf{x}_n; \mathbf{W}^{opt}). \qquad (2.9)$$

To generate reliable classification of the image $\mathbf{X}$, the effect of possible outlier $\mathbf{x}_{n^*}$ with different prediction of its neighbors can be suppressed by a label fusion of patches to label the image. We use a simple way by combining the class probabilities in the neighborhood around the most discriminative patch:

$$C(\mathbf{X}) = \operatorname*{argmax}_{k \in \{1, \dots, K\}} \sum_{\mathbf{x}_n \in \mathfrak{N}(\mathbf{x}_{n^*})} \mathbf{P}(k|\mathbf{x}_n; \mathbf{W}^{opt}), \qquad (2.10)$$

2.3.5 Implementation Details

In this study, we assume one middle level discriminative patch from an image is enough for the image classification. To discover the patches which are discriminative and representative for their image categories, the patch size should not be too small to include semantic information for the discriminative objects. The image patches are extracted by fixed-size sliding window with overlapping. The patch size and step size is specified in experimental settings. We further analyze the sensitivity of patch size later. In learning stage II, since the patches per image are overlapping, the non-informative patches are selected with a spatial constraint that they should not appear neighboring to the discriminative patches. The Eq. 2.9 can be implemented by manually setting $\mathbf{P}(K+1|\mathbf{x}_n; \mathbf{W}^{opt}) = 0$ to make sure all $c(\mathbf{x}_n) \neq K+1$.

In each of the two training stages, we train a CNN model similar to Fig. 2.2. The following strategies are employed to improve the performance of the learned CNN. First, Rectified Linear Units (ReLUs) are used to map the neurons' output in convolutional layers. As shown in [81, 44], ReLUs demonstrates faster convergence and better performance than sigmoid functions. Second, to incorporate larger variability in our training samples, hence, increase the robustness of the CNN, we augment data using label-preserving transformations like in [58]. Specifically, we apply random translation,

rotation and scaling to increase training data samples. Third, the "dropout" strategy [82] is employed to reduce the risk of "over-fitting". Dropout rate is 0.5. It forces half of the neurons randomly "dropped out" at each training iteration. In this way, the complex correlation of neurons is reduced and more robust features can be learned. Finally, as the training set may be too large to load into memory at one time, we trained our model using a mini batch of samples at each iteration. The optimization is implemented by stochastic gradient descent with a momentum term β [83] and a weight decay term γ. For a weight $\omega \in \mathbf{W}$, its update at iteration t is defined as

$$\omega^{(t)} = \omega^{(t-1)} + \Delta\omega^{(t)}, \tag{2.11}$$

where

$$\Delta\omega^{(t)} = \beta \cdot \Delta\omega^{(t-1)} - \epsilon \cdot (\delta\omega^{(t-1)} + \gamma * \omega^{(t-1)}). \tag{2.12}$$

$\delta\omega^{(t-1)}$ is the gradient of weight based on current batch of samples.

The learning process is conducted on a training subset and a validation subset. It won't stop until the error rate on validation subset is smaller than a threshold ξ or a predefined maximum number of epochs (1000 epochs in both stages) is reached. Besides, the learning may stop earlier if it cannot reach smaller error since the current smallest one after a number of patient iterations. Our algorithm is implemented in Python using Theano library [84]. To leverage the highly parallelable property of CNN training, we trained our models on a 64-bit desktop with i7-2600 (3.4GHz) CPU, 16GB RAM and NVIDIA GTX-660 3GB GPU.

2.4 Experiments

The evaluation is conducted on a synthetic data set and a CT data set. Classification accuracies are reported in terms of recall, precision and F_1 score as

$$recall = \frac{TP}{TP + FN}, \tag{2.13}$$

$$precision = \frac{TP}{TP + FP}, \tag{2.14}$$

$$F_1 = 2\frac{precision \cdot recall}{precision + recall}, \tag{2.15}$$

where TP (true positive) denote the number of samples belonging to class k and correctly classified; FN (false negative) denote the number of samples belonging to class k but misclassified; FP (false positive) denote the number of samples not belonging to class k but misclassified as class k.

2.4.1 Image Classification on Synthetic Data

We first validate our method on a synthetic data set, which has been briefly introduced as a toy example in Sec. 2.3.2. It is constructed by 4 types of geometry elements: triangle, square, circle and diamond. Each synthetic image (60×60) contains two of the geometry elements at random positions on black background (intensity value 0). The basic geometry elements are roughly 20×20 with up to 10% variance in height and width. They have random intensity values in $[1, 255]$. In constructing the two image classes, we ensure that the triangle and square are the "distinctive" element and only appear in Class1 and Class2, respectively. Besides the distinctive element, Circle or diamond is evenly picked as the second element in each image. (Some examples of the synthetic images are shown in Fig. 2.4(a).) Overall, we create 2000 training, 2000 validation and 2000 for testing samples (balanced distribution for each class).

A comparison study is conducted using: (1) logistic regression (LR); (2) SVM; (3) standard CNN, similar to LeNet[60], trained on whole image (SCNN); (4) local patch-based CNN without boost, i.e., the CNN trained by pre-train stage only (PCNN); (5) local patch-based CNN refined without additional non-informative class (RPCNN1); (6) local patch-based CNN refined with both discriminative and non-informative patches (RPCNN2). Methods (1)-(3) represent conventional learning (using image intensities directly as features) and deep learning approaches. Methods (4),(5) are two variants of our proposed one (6), which are presented to verify the effects of each component of our method. The parameters of LR and SVM are optimized using grid search with cross-validation. All CNN-related methods use the same intermediate architecture: one convolutional layer with 10 5×5 filters, one max-pooling layer with 2×2 kernel, one hidden layer of 300 nodes, and finally followed by a LR layer to output response. The patch size for all patch-based CNNs is 30×30. There are 36 patches extracted from

Table 2.1: Classification accuracies (%) on synthetic data set as shown in Fig. 2.4. Class 1 contains triangle; class 2 contains square.

	Triangle and square								
	Recall			Precision			F_1		
Class	1	2	Total	1	2	Total	1	2	Total
LR	78.7	83.4	81.1	82.6	79.7	81.1	80.6	81.5	81.1
SVM	84.5	81.2	82.9	81.8	84.0	82.9	83.1	82.6	82.9
SCNN	84.2	82.4	83.3	82.7	83.9	83.3	83.5	83.2	83.3
PCNN	99.6	99.7	99.7	99.7	99.6	99.7	99.7	99.7	99.7
RPCNN1	98.4	99.7	99.1	99.7	98.4	99.1	99.0	99.1	99.1
RPCNN2	**100**	**100**	**100**	**100**	**99.9**	**100**	**100**	**100**	**100**

each 60×60 image through a sliding window with 6-pixel step size.

As shown in Table 2.1, standard deep learning method (SCNN) is better than LR, which indicates deep learning can learn good features from raw data. By leveraging the local discriminative information, PCNN gets $\approx 16\%$ improvement from SCNN. It implies that standard CNN does not fully discover and learn the discriminative local patches: "triangle" and "square". On the contrary, the most discriminative and non-informative local patches are effectively discovered by the proposed CNN with multi-instance learning as shown in Fig. 2.4(b). Among our local patch-based CNNs (PCNN, RPCNN1 and RPCNN2), RPCNN1, which is trained on extracted discriminative (without non-informative) patches, is worse than PCNN due to overfitting (because the parameters of RPCNN1 are initialized from PCNN, and refined by training with the extracted discriminative patches only). RPCNN2, similar to RPCNN1 but refined by training with discriminative as well as non-informative patches, achieves the best performance.

To further prove the adaptivity of our algorithm, we re-labeled the synthetic data using Diamond and circle as distinctive elements in class 1 and class 2, respectively (see Fig 2.6(a)). In other words, although the synthetic data are exactly the same, the local patches to distinguish the two classes become different. This is in analogy to real-world problems where the datasets are identical but the classification goal is changed. After conducting the pre-train algorithm, the extracted local patches from the learned model are shown in Fig. 2.6(b). Again, the extracted local patches contain the most
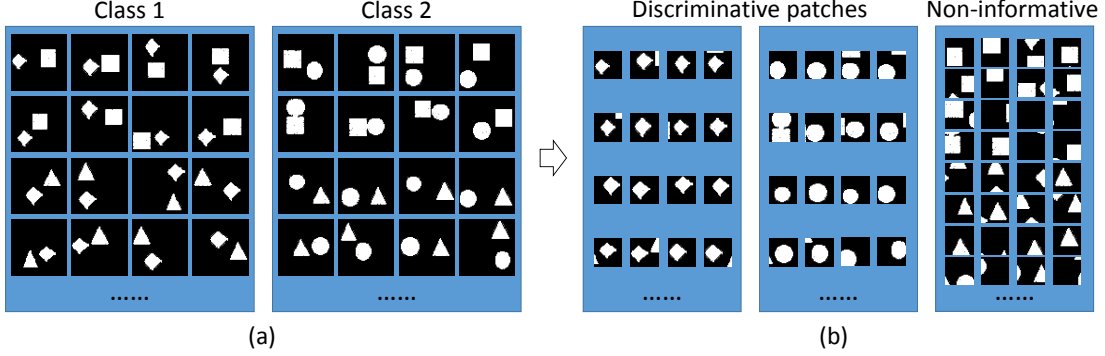
Figure 2.6: The second toy example. (a) Synthetic images of two classes distinguished by diamond and circle. Note that we use the same image samples as in Fig. 2.4, but assign different labels. (b) The discriminative and non-informative local patches discovered by the pre-trained CNN.

Table 2.2: Classification accuracies (%) on synthetic data set as shown in Fig. 2.6. Class 1 contains diamond; class 2 contains circle.

	Diamond and circle								
	Recall			Precision			F_1		
Class	1	2	Total	1	2	Total	1	2	Total
LR	70.3	62.5	66.4	65.2	67.8	66.5	67.7	65.0	66.5
SVM	69.0	63.1	66.1	65.2	67.1	66.1	67.0	65.0	66.1
SCNN	91.8	94.2	93	94.1	92.0	93.0	92.9	93.1	93.0
PCNN	99.6	**100**	99.8	**100**	99.6	99.8	99.8	99.8	99.8
RPCNN1	95.6	**100**	97.8	**100**	95.8	97.9	97.8	97.9	97.9
RPCNN2	**99.9**	99.9	**99.9**	99.9	**99.9**	**99.9**	**99.9**	**99.9**	**99.9**

discriminative information, diamond and circle. The classification accuracies are shown in Table 2.2. Our two-stage learning framework RPCNN2 achieves the best performance among all comparison methods again. This result demonstrates that our multi-instance CNN learning can *adaptively* learn discriminative local regions for specific classification tasks without any local level annotations.

2.4.2 Body Part Recognition on CT Data

We applied our method in body part recognition of transversal CT slices [40, 46]. As shown in Fig. 2.1, transversal slices of CT scans are categorized into 12 body sections (classes). A dataset of 7489 2D transversal images was extracted from whole body CT scans of 675 patients with very different ages (1-90 years old). The imaging protocols

were different: 31 different reconstruction kernels, $0.281mm - 1.953mm$ in-slice pixel resolution. The images are divided into 2413 (225 patients) training, 656 (56 patients) validation and 4043 (394 patients) testing subsets. In this experiment, we augment data by applying up to 10% random translations (relative to the image size) in training and validation subsets to make them 3 times larger.

The preprocessing includes two different steps: image sub-sampling and image cropping. First, all images are re-sampled to have $4mm \times 4mm$ pixel resolution and 90×90 in size. Then, cropping operation extracts 50×50 local patches from each image with 10-pixel step size. Thus, 25 local patches are extracted per image. Our CNN has similar structure as in Fig. 2.2. C1 layer has 20 9×9 filters. C3 layer has 40 9×9 filters. Two sub-sampling layer, S2 and S4, use 2×2 max-pooling. H5 layer has 600 hidden nodes. The LR layer, O6, has 12 output nodes in pre-train stage, or 13 output nodes in fine-tune stage.

As shown in the "RPCNN2" row of Table 2.5, our method can achieve the classification accuracy (F_1 score) at 92.23%. Fig. 2.7 shows more detailed classification performance by the confusion matrix. Most errors appear close to the diagonal line, which means most mis-classifications happen in the neighboring body sections. Quantitatively, the classification error is 7.79%, and $\approx 92\%$ of which are "less-than-one neighboring class error" (within the red line corridor of Fig. 2.7). In practice, this kind of errors are already acceptable for some use cases. The remaining gross error ($< 0.7\%$) can be further suppressed by a simple label smoothing after classifications of a series of continuous slices for 3D bodypart identification.

Table 2.3: Classification accuracies on CT data in terms of Recall (%).

Recall

Class	1	2	3	4	5	6	7	8	9	10	11	12	Overall
LR1	48.54	63.64	33.33	69.95	39.09	43.37	47.66	81.43	25.82	53.68	41.62	88.79	63.37
LR2	67.96	64.50	42.59	74.24	38.64	42.17	56.25	78.51	37.09	65.79	60.41	91.48	68.71
SVM1	41.75	64.07	46.30	76.26	36.36	45.18	47.27	78.66	31.46	52.90	50.25	88.57	64.63
SVM2	76.70	81.39	39.82	79.55	54.09	63.86	69.92	84.06	44.60	64.47	75.64	96.53	76.75
CaffeNet	71.85	64.94	**87.96**	85.10	74.09	80.72	57.81	94.59	80.75	78.95	85.28	97.53	84.74
SCNN	84.47	93.51	72.22	88.89	80.46	80.12	86.72	95.47	77.93	77.63	78.43	96.30	87.73
SCNN_a	81.55	96.54	43.52	92.68	79.09	**90.96**	93.75	88.45	51.17	64.47	**90.36**	92.60	84.76
PCNN	87.38	92.21	**87.96**	93.18	**90.00**	74.70	**94.53**	95.47	81.69	81.05	**90.36**	92.49	90.21
RPCNN1	54.37	83.12	69.44	**94.95**	75.91	82.53	93.36	95.03	**84.98**	72.63	85.53	96.75	87.78
RPCNN2	**88.35**	**96.97**	80.56	91.67	86.82	87.35	93.75	**95.61**	79.81	**87.11**	87.82	**99.33**	**92.21**

Table 2.4: Classification accuracies on CT data in terms of Precision (%).

Precision

Class	1	2	3	4	5	6	7	8	9	10	11	12	Overall
LR1	72.46	63.09	63.16	64.42	55.48	54.14	54.46	72.15	55.00	57.14	47.40	67.87	62.21
LR2	65.42	64.78	61.33	68.53	53.80	46.36	56.92	76.06	59.40	64.43	62.80	78.92	67.74
SVM1	75.44	61.93	65.79	64.81	53.69	51.37	53.78	74.00	56.30	58.26	49.62	72.15	63.72
SVM2	89.77	72.03	72.88	72.25	56.94	57.92	72.76	82.62	62.50	70.81	71.64	90.54	76.39
CaffeNet	98.67	**98.04**	41.49	86.41	**93.68**	83.23	87.57	84.25	66.15	77.32	84.42	99.09	86.84
SCNN	87.88	87.45	82.11	87.35	87.62	83.13	84.73	92.36	76.50	75.06	83.51	96.73	87.72
SCNN_a	96.55	78.80	77.05	81.19	89.69	79.06	80.54	94.38	83.85	74.70	66.17	98.33	85.75
PCNN	96.77	93.83	81.20	87.86	85.35	**96.88**	90.30	95.33	78.73	**83.92**	77.73	**99.76**	90.69
RPCNN1	**100.00**	94.58	54.75	76.27	91.26	93.20	91.92	95.59	77.68	82.64	78.74	97.08	88.62
RPCNN2	96.81	91.80	**88.78**	**90.30**	90.52	89.51	**91.95**	**95.75**	**80.95**	82.13	**91.78**	98.66	**92.25**

Table 2.5: Classification accuracies on CT data in terms of F_1 score (%).

Class							F_1 score						
	1	2	3	4	5	6	7	8	9	10	11	12	Overall
LR1	58.14	63.36	43.64	67.07	45.87	48.16	50.83	76.51	35.14	55.36	44.32	76.93	62.78
LR2	66.67	64.64	50.27	71.27	44.97	44.16	56.58	77.27	45.67	65.10	61.58	84.74	68.22
SVM1	53.75	62.98	54.35	70.07	43.36	48.08	50.31	76.26	40.36	55.45	49.94	79.52	64.17
SVM2	82.72	76.42	51.50	75.72	55.48	60.75	71.32	83.33	52.06	67.49	73.58	93.44	76.57
CaffeNet	83.15	78.13	56.38	85.75	82.74	81.96	69.65	89.12	72.73	78.13	84.85	98.31	85.78
SCNN	86.14	90.38	76.85	88.11	83.89	81.60	85.71	93.89	77.21	76.33	80.89	96.52	87.73
SCNN_a	88.42	86.77	55.62	86.56	84.06	84.59	86.64	91.32	63.56	69.21	76.39	95.38	85.25
PCNN	91.84	93.01	84.44	90.44	87.61	84.35	92.37	95.40	80.18	82.46	83.57	95.99	90.45
RPCNN1	70.44	88.48	61.22	84.59	82.88	87.54	92.64	95.31	**81.17**	77.31	82.00	96.91	88.20
RPCNN2	**92.39**	**94.32**	**84.47**	**90.98**	**88.63**	**88.42**	**92.84**	**95.68**	80.38	**84.55**	**89.75**	**98.99**	**92.23**
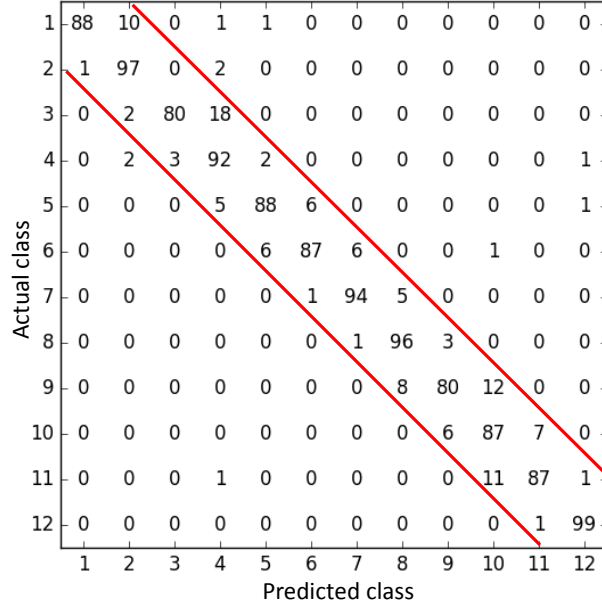
Figure 2.7: Confusion matrix of RPCNN2 on CT data. Values are normalized to $0 \sim 100$ in each row.

For quantitative comparison, tested image classification methods include: (1) LR1, (2) LR2, (3) SVM1, (4) SVM2, (5) CaffeNet, (6) SCNN, (7) SCNN_a, (8) PCNN, (9) RPCNN1, and (10) our proposed RPCNN2. In LR1 and SVM1 methods, we use bag-of-word model with dense SIFT features to train logistic regressor and SVM classifier, respectively. While LR2 and SVM2 methods simply replace SIFT by HOG features. Same as the previous experiment, the LR and SVM parameters were optimized using grid search with cross-validation on the same training and validation sets as other comparison methods. Then, the optimized models were applied on the same testing set to produce results for fair comparisons. SCNN method is the standard CNN that takes the whole slice as input. SCNN_a method is the same as SCNN except trained by extra 6 times more augmented data samples by random transformations, rotations and scalings. Method (8),(9) are two variants of our proposed one (10), which are presented to verify the effects of each component of our method. Similar network structure is used in all CNN-based methods, (6)-(10), except different input and output sizes in patch-based CNNs (8)-(10). CaffeNet [85] has the similar structure as AlexNet [44] with a minor variation, which is trained on whole images without cropping. We noticed that training of CaffeNet with 50×50 cropping doesn't converge. This observation shows that

our proposed method is not merely a kind of data augmentation via image cropping. The discriminative and non-informative patches discovered by multi-instance learning are the keys to success. RPCNN1 is trained on extracted discriminative (without non-informative) patches from learning stage I. Although the trained classifier focuses more on discriminative patches, ambiguous local patches across different classes (e.g. upholding arms may look similar to neck) are completely ignored and thereby mislead the classifier at runtime. Thus, the performance of RPCNN1 is worse than PCNN and close to the SCNN. Compared to its variants, the proposed RPCNN2 achieves the best performance (even better than much deeper CNN, CaffeNet), which proves the necessity of using all strategies designed in our method.

In addition, we noted that the SCNN_a trained with more augmented data is even inferior to the SCNN due to overfitting (training error: SCNN_a 4.4% vs. SCNN 5%; testing error: SCNN_a 14.7% vs. SCNN 12.3%). It shows that the global CNN cannot learn the anatomy characteristics from more augmented data but overfit them. As shown in Table 2.5, the overfitting problem is more severe in neck (column 3) and liver upper (column 9) sections. These two sections happen to have subtle global appearance differences compared to their neighboring sections and are thus prone to overfitting. It is proved that the success of the proposed method against the standard CNN does not result from more augmented training samples but its capability of *discovering* local characteristics of different body parts. The online classification time of each method is about (1) $4ms$, (2) $3ms$, (3) $5ms$, (4) $4ms$, (5) $3ms$, (6) $4ms$, (7) $4ms$, (8) $10ms$, (9) $11ms$, (10) $11ms$ per image, respectively.

In this application of body part recognition, the discovered discriminative patch samples and non-informative (kind of misleading) patches for each class in the CT dataset are shown in Fig. 2.8. From this figure, we observe that the proposed method "magically" extracts meaningful local patches for each class without any prior information, and these representative and discriminative local patches can significantly improve the classification task comparing with the global image information. It is also noted that the discriminative patches of liver middle class contain only a small part of the liver and a narrow band of the left lung bottom. The corresponding non-informative
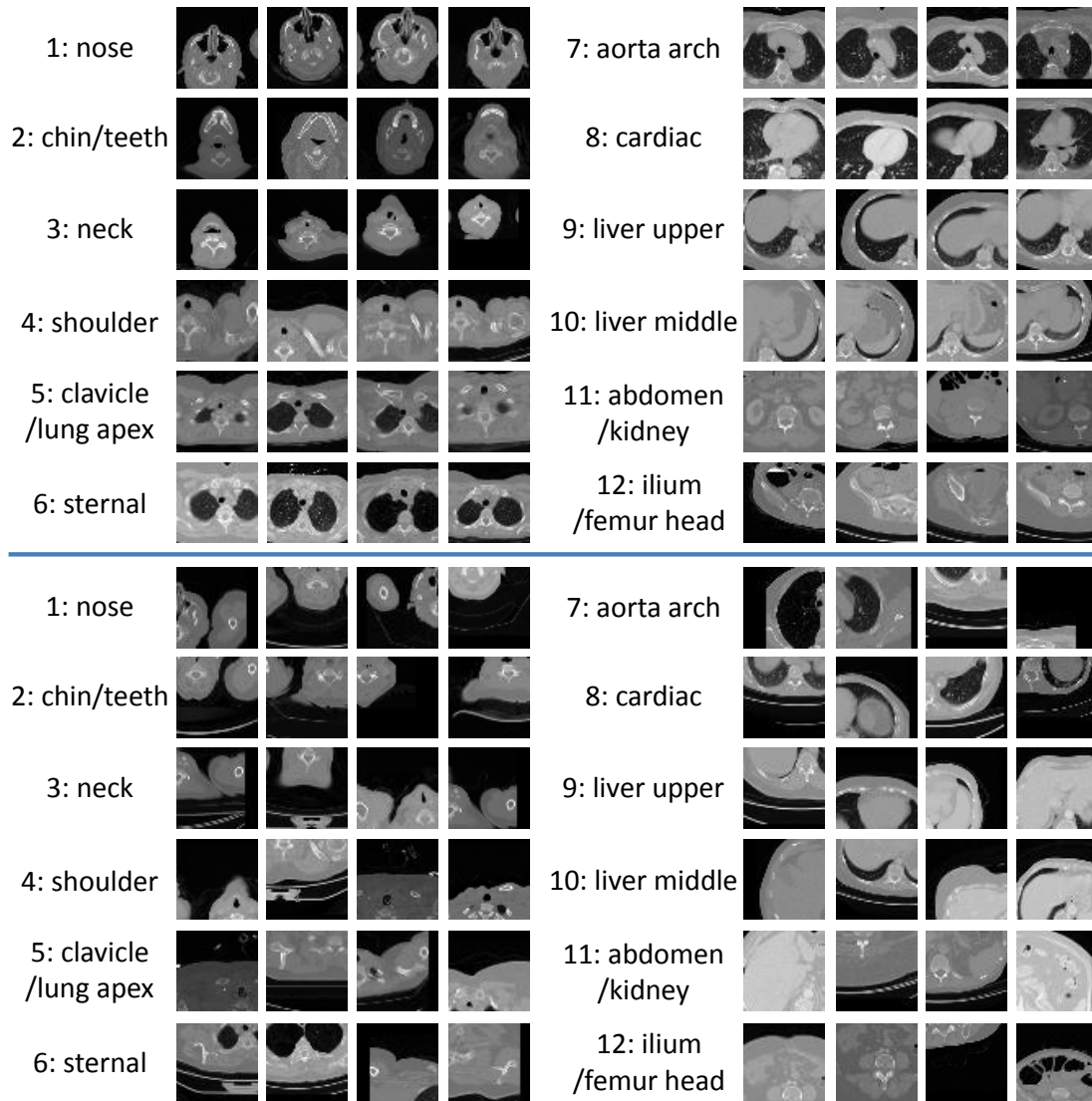
Figure 2.8: Automatically discovered discriminative and non-informative patches from each class through multi-instance learning.

patches may contain a large piece of the liver, since it can appear in different classes (e.g. liver upper and abdomen/kidney). This observation indicates that the proposed method just finds the discriminative information from the data, and does not guarantee that some particular "desired" element will be considered as discriminative.

2.4.3   Discussions

To investigate whether the standard CNNs can discover the required discriminative features at some intermediate layers or they completely miss them, we did extra experiments to train linear SVM classifier on the learned hidden activation on each layer of the baseline CNN (SCNN). Totally 5 classifiers were learned from features on layers (C1, S2, C3, S4, H5). The feature sizes are 134480, 33620, 43560, 10240 and 600, respectively. The F_1 scores on testing set are 0.75, 0.77, 0.86, 0.86 and 0.88, respectively. Compared with reported F_1 score of SCNN ($\approx$0.88), we conclude that (1) features on higher layers are better for the classification task; (2) although the learned features in SCNN are discriminative to some extent, the more representative and discriminative local features can only be discovered in our proposed patch based learning algorithm.

To evaluate the robustness of the trained models, we apply different scales of random linear translation on the testing data and compute the classification error rates. The Fig. 2.9a shows the results. From the plots, we can see that our proposed method RPCNN2 has the best robustness regarding to the random translation of testing samples. Although the training and validation subsets have been augmented using up to 11% random translation, the other approaches do not perform as well as the proposed method when the testing samples have larger translations. In this situation, retraining the models on augmented dataset with larger translation could be a solution. However, the re-training costs cannot be overlooked and fixing it after the fact is not efficient in practice.

As one of the important parameters in RPCNN2 method, step size of sliding window testing is investigated regarding to the accuracies (shown in Fig. 2.9b). The running times for step sizes 1, 5, 10, 15, 20, 25, and 30 pixels are 541.1, 30.6, 11.7, 5.3, 5.2, 3.6 and 3.4 ms per image respectively. Considering the balance of running time and
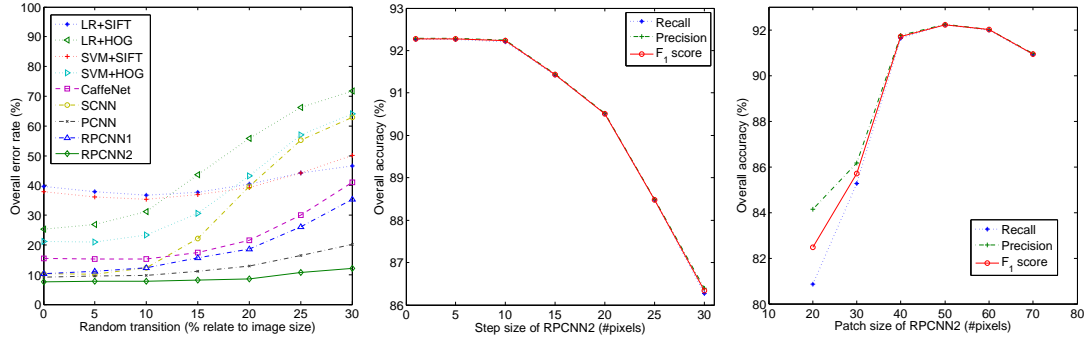
(a) Classification error rates vs. scales of random translations on testing data.

(b) Classification accuracies vs. step sizes of sliding window in RPCNN2.

(c) Classification accuracies vs. patch sizes in RPCNN2.

Figure 2.9: Performance analyses on the sensitivity of parameters.

accuracy, step size 10 or 15 should be a reasonable choice in this experiment. The effect of patch size to the classification accuracy is also investigated as shown in Fig. 2.9c. We can see from the plot that (1) the patch size should not be too small to capture the discriminative information (size 20 or 30); (2) the performance is not very sensitive to the local patch size once it is big enough to include discriminative information (sizes from 40 to 60 in this task).

We also conducted two extra experiments to test other variants of our proposed method. First, we use Eq. (2.2) with a bit larger patch size (70x70) rather than the Eq. (2.3) to accommodate for the neighbors information. The final classification accuracy in terms of F_1 score becomes 91.67%, a little worse than that of the proposed RPCNN2 (92.23%). Second, instead of using the run-time classification strategy by Eq. (2.10), we can simply use the Eq. (2.8) as in [40], or majority voting of predictions from all partitioned patches in the slice to predict image classification. The F_1 score drops $\approx 2\%$ and $\approx 4\%$, respectively.

One limitation of this recognition method is the identical patch size for different classes. Although the algorithm is not that sensitive to patch size, it requires prior knowledge to ensure that the chosen patch size is appropriate to include the discriminative information in a fixed size region. It could be better to incorporate strategies like multi-scale convolution [86] or multi-scale image patch [87] to discover and recognize different-size discriminative local regions in different classes. Another limitation is the

sliding-window and multiple stage pipeline. Like the improvement from R-CNN [88] to fast R-CNN [72] and faster R-CNN [73], it may be possible to simplify the multi-stage pipeline to single-stage training, and use part of the convolutional features of image for local regions to avoid the sliding-window strategy in training and testing.

2.5 Summary

In this chapter, we present a two-stage multi-instance deep learning strategy to identify anatomical body parts by using discriminative local information. The method RPCNN2 does not require annotations of the discriminative local patches. Instead, it automatically *discovers* these local patches through multi-instance deep learning for the image classification task. We validate our method on a synthetic dataset and a CT dataset. It shows clear improvements compared with other state-of-the-art methods. Specifically, the proposed patch-based CNN (with 938,893 parameters) outperforms the standard CNN (with 6,218,292 parameters).

This slice based body part recognition algorithm can be trivially applied in 3D image data by labeling each slice one by one. Considering that no more than 7% error locating between continuous sections is acceptable in practice, the only gross error (less than 0.7%) could be easily eliminated by a smoothing filter on the predicted label distribution. For example, given a CT scan with 200 slices for a lung study, we could classify each 2D slice and get a 200-long label vector. After label smoothing, the slices containing the lung (of class 5-9) should be the ROI to initialize the lung-oriented work flow.

Chapter 3

Adaptive Atlas based Medical Image Segmentation

3.1 Introduction

In clinical researches and trials, quantitative assessment of organs or tissues is required for large data analysis. After proper localization of ROI from the image, robust and accurate image segmentation is the key step to obtain quantitative measurements. One of the important applications is the brain structure segmentation. Human brain is one of the most interested anatomies in clinical studies because any diseases or disfunction related to the brain could significantly affect people's life. For example, Alzhemimer's disease (AD) is the sixth leading cause of death in the United States. It is the only disease among the top 10 causes of death in America that cannot be prevented, cured or even slowed. Although deaths from other major causes have decreased significantly in the last decade, deaths from AD have increased significantly - 71% [22]. Besides of Alzheimer's, a host of neurological disorders such as Parkinson's, Huntington's, schizophrenia and ventriculomegaly manifest as changes in the size and shape of multiple structures throughout the brain, such as the thickness of the cortex [89]. Such changes can be quantified after delineating the structures in images captured with magnetic resonance imaging (MRI). Due to the intense labor and low reproducibility of manual labeling by expert anatomists, automated delineation methods are desired, particularly for large clinical studies and clinical trials [21].

Brain structures have complicated geometry and overlapping intensity distributions in MRI, therefore automated approaches employ priors to regularize the ill-posed task of structure delineation. Since human brains have high inter-subject variability in healthy cohorts, and even more so when cohorts with pathologies are studied, the priors should be appropriately adapted in order to avoid excessive bias in image interpretation. As

we show in this chapter, even the state-of-the-art methods do not possess sufficient ability to adapt their priors for cases in which the unseen target has variability not well represented by the training data. This has motivated our research to develop adaptive methods that quantify the suitability of the training data for each target subject.

Atlas-based approaches have become the most commonly proposed automated brain segmentation methods. In medical image segmentation, an atlas is a pair of image data and its corresponding label map, which is provided manually or semi-automatically by expert as ground truth. Atlas based methods are commonly used in brain segmentation because of the fact that despite the variances among different subjects the same anatomy still has essentially similar characteristics. There are two broad categories of these approaches. The first is the probabilistic atlas approach that aligns all training data into a common coordinate system, independent of the target [90, 91]. The second which is a multi-atlas approach, by contrast, registers the intensity image from each training subject directly to the intensity image of the target when it arrives [92, 93, 94, 95, 96, 97, 98, 99]. Once the images are aligned by this method, the training subject's manual labels can be propagated to the target. Due to errors in the inter-subject brain registration [100, 101], each voxel has typically more than one label, necessitating the use of a label fusion procedure to select the best label for each voxel.

Due to the need to resolve label ambiguity, label fusion has become an active area of research. A particularly vexing, open research problem is the construction of an adaptive label fusion method that generalizes well to minimize the discrepancy between training error and test error, when the morphological and intensity variability of targets are not well represented in the training data. This situation arises often because it is unrealistic to have a training set that includes images acquired with the latest technology for all types of pathological variations and even less realistic to have an expert manual labeling for them as well. An adaptive fusion method would not only utilize the priors from the training set but also estimate the extent to which those priors apply. Ideally, it would do so not at the whole atlas level but locally at the individual voxel level. Furthermore, it would be preferable if such a method would determine the local relaxation parameters automatically and outperform the leading methods in such

Healthy appearing brain
(psychiatric/healthy)

Degenerative brain
(Alzheimer's disease)
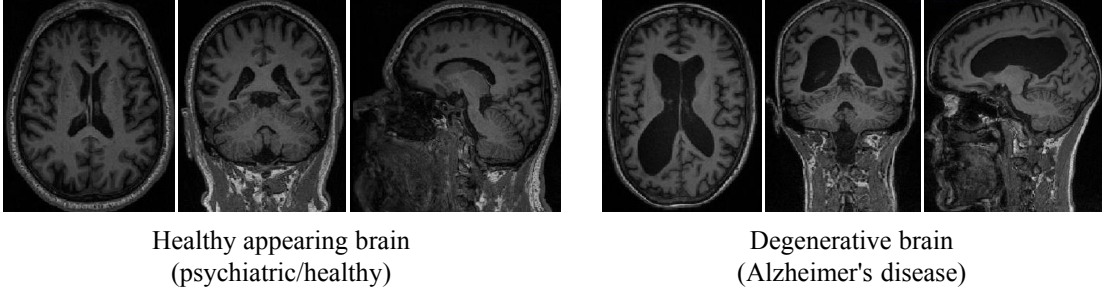
Figure 3.1: Our approach allows priors learned from healthy subjects (left) whose labels are available publicly to accurately parcellate subjects with Alzheimer's (right) despite large morphological differences.

situations. The method proposed in this chapter aims to fulfill these objectives.

Our research contributions are three-fold. *First* we automatically define a target-specific, spatially variant relaxation map which estimates for each individual voxel how well the target's anatomy is represented by the training multi-atlas data. This map determines where and by how much to relax the priors and is computed for each target subject based on multi-atlas label entropy[1] across the training atlases after high-dimensional, non-linear registration. To the best of our knowledge, this is the first development of such functionality. *Second,* to perform label fusion we use this relaxation map to guide an adaptive probabilistic atlas with a GMM of tissue intensities and spatially varying mixing coefficients. Using the multi-atlas fit, we derive target-specific intensity priors for the GMM (as in [102]) that do not necessitate the target be acquired with the same acquisition sequence as the training data, as is assumed in other methods [103]. To regularize the segmentation we add a Markov Random Field (MRF) label smoothness prior which also uses the preliminary multi-atlas fit to make the priors target specific. *Third*, we solve for the spatially varying mixing coefficients, GMM parameters, and tissue labels by embedding the complete model in an expectation-maximization (EM) framework. This is achieved by modeling the GMM mixing coefficients as a random variable drawn from a Dirichlet distribution whose parameters vary spatially according to the relaxation map and spatial priors. This enables the weighted regularization of an adaptive, *evolving* probabilistic atlas with spatially

---

[1]Entropy measures the lack of multi-atlas agreement voxel-by-voxel and our use of it is explicitly quantified in section 3.3.3.

variant GMM mixing coefficients. Through EM optimization, the target-specific adaptive probabilistic atlas evolves to fit the target subject.

We demonstrate, in direct head-to-head comparisons, that our approach outperforms current leading methods when labeling an Alzheimer's disease cohort using publicly available labeled healthy subjects as training data. Our tests include comparisons with the best multi-atlas algorithms using patch-based label fusion methods [100, 104, 105], commonly used probabilistic atlas methods such as FreeSurfer [106], and methods specifically targeting greater generalization [107, 89]. Additional tests show that high accuracy levels are maintained when our approach is used to parcellate healthy subjects.

3.2   Related Work

In recent decades, many approaches have been proposed to segment human organs or tissues in different modalities [9, 108, 109, 110]. The most interested ones are automatic techniques, including model based methods [111, 112, 113, 114, 115], learning based methods [16, 116, 117, 118, 119], atlas-guided approaches [91, 98, 90, 120, 121, 104], and so on. Usually, learning based approaches require large amount of consistent training data which could be label intensive in practice. Owing to the variety of data in clinical trials and limited amount of training samples, many learning based segmentation methods are hardly applied. Deformable model based methods [7, 110] are accurate but may need to tune parameters case by case. For some specific anatomies, prior knowledge like shape and topology [122, 123] could be incorporated to gain more robustness and accuracy. Atlas-guided approaches rely on image registration algorithms [124, 125, 126, 101] to transfer knowledge from training atlases to the target space and apply voxel classification to get final segmentation [127, 128].

Among these approaches, atlas based methods are the most commonly used approaches for brain image segmentation. Given several atlases, there are two ways to segment a new target image. The first is to learn a single statistical atlas that models the spatial priors for individual structures. A single probabilistic atlas is fit in a

Bayesian framework for voxel classification [90, 91]. The second approach is to register the set of atlases (multi-atlas) to the target image and then compute the final segmentation via a label fusion approach [99, 93, 96, 98]. Due to the errors of non-linear image registrations [129, 130], atlas-guided methods have limitation in segmenting objects with different pathological and geometric characteristics.

Generalization of brain segmentation methods has been the subject of recent research. Several categories of methods have been proposed to adapt priors learned during training to the target. In the probabilistic atlas genre, [97] build multiple probabilistic atlases *a priori* and then combine them to form a piecewise, regional probabilistic atlas. While this increases adaptivity across the brain as a whole, the variation found in each region must be represented by a subset of the training data and discontinuities between regional atlases must be smoothed.

In the multi-atlas genre, the original *majority voting* formulation of label fusion that applies equal weight to each atlas [92] has been made more adaptive. Several researchers have proposed *locally weighted label fusion* methods that derive weights from the local similarity between atlas and target. Such approaches allow the weights to vary spatially and have been optimized with expectation-maximization [131, 132, 133, 134, 135] and graph-cuts [102, 103]. *Non-local, patch-based label fusion* has also been proposed [136, 137, 105, 138, 100, 139, 104]. These methods extend locally weighted label fusion by considering multiple points within each atlas in a region around each voxel to be labeled. Such approaches have attained some of the highest published segmentation accuracies to date, including the generative model for joint label fusion [104] and the top two performing methods [100, 105] in the MICCAI 2012 atlas based brain parcellation challenge [140].

A *limitation of the patch-based approaches* is its reliance upon a fixed search volume size. This limits the generalization from the training set to only those target subjects whose anatomical correspondences differ by less than or equal to the search region radius. For example, several propsed methods [136, 105, 100, 104] require non-local patch matches to be between ±3mm to ±5mm from the voxel to be classified. This is because they use global fixed search windows of limited size, ranging from $7 \times 7 \times 7$ to

$11 \times 11 \times 11$ voxels centered on the voxel to be labeled.[2] Simply increasing the global fixed search volume size is not a viable solution because it increases computation time significantly and can reduce accuracy, particularly for smaller structures, by finding false matches that are not part of the structure to be segmented.

In practice, however, the morphological variation between conditions such as ventriculomegaly or Alzheimer's compared to healthy brains, can be much greater than 5mm. Such variability is illustrated in Fig. 3.1. Consequently new methods have been proposed for greater generalization while still obtaining comparable results when segmenting healthy brains. [107, 89] present two such approaches which assume the intensities of the target image voxels follow a GMM with spatially varying mixing coefficients and solve for the labels using expectation-maximization. Both methods also use a global fixed relaxation parameter (κ in [107] and R_f in [89]). A limitation in these approaches is that if the fixed relaxation constants are set too low insufficient adaptivity will result, because the priors are not adapted sufficiently to the variation present in the target. This is analogous to patch-based methods which have fixed global search window sizes that can also result in insufficient adaptivity. If the fixed constants are increased to allow for greater adaptivity, leakage can occur, particularly when attempting to simultaneously segment 30+ structures. This motivates our research to develop an adaptive method that handles the simultaneous segmentation of 30+ structures while spatially regulating the relaxation factor in a anatomically coherent way to reflect the local suitability of the training data to the target [141, 142].

3.3 Methodology

3.3.1 Algorithm Overview

Our proposed algorithm is illustrated in Fig. 3.2. It consists of four main steps: (1) Starting from the center right, we preprocess the target image to prepare the brain parenchyma for analysis. (2) The multiple atlases are non-rigidly registered to the target. (3) We define our Bayesian probabilistic model of brain parcellation and construct

---

[2]Assuming typical MRI voxels of size 1mm^3
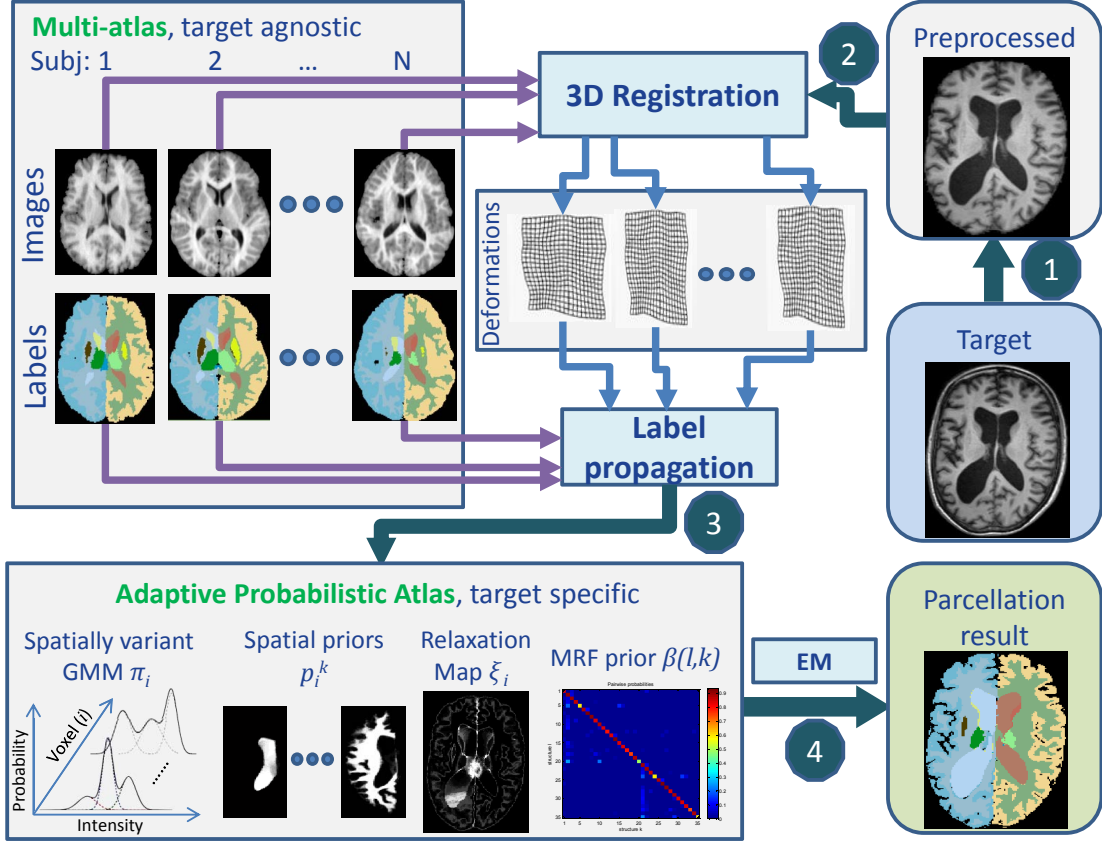
Figure 3.2: Workflow of our proposed method, *iMAEASA*, integrates a multi-atlas (top, green label) with our highly adaptive probabilistic-atlas (bottom, green label). The main steps include: (1) preprocessing, (2) target-specific multi-atlas fitting, (3) target specific adaptive probabilistic atlas construction, and (4) evolution and fitting of the probabilistic atlas.

the four primary components of our target-specific, adaptive probabilistic atlas that guides label fusion. (4) We label the target brain by evolving the target-specific adaptive probabilistic atlas to the target using an expectation maximization framework. We call our method *iMAEASA* because it combines an Intensity-driven Multi-Atlas with an Extended Adaptive Statistical Atlas.

In the first step, all MRI volumes are pre-processed. We extract the brain parenchyma using ROBEX [143], correct intensity inhomogeneity using N3 bias correction method [144], and align all brains to approximate Talairach space by applying a rigid-body transform from each volume to the MNI305 template [145, 146].

In the second step, we fit our training data in the form of a multi-atlas to the

target. This entails non-rigidly registering every training image to the target and warping the corresponding training labels using the same transformation. This adapts the multi-atlas to the target. Due to limitations in the registration algorithm and differences between the training and target subjects, some voxels have high multi-atlas label agreement, while others have low agreement. Agreement is inversely proportional to label uncertainty. This spatially varying label uncertainty is a key ingredient in our construction of a target-specific adaptive probabilistic atlas.

In the next two sections, we describe step three of our approach which is to define our Bayesian probabilistic model of brain parcellation and to construct the primary components of our target-specific probabilistic atlas to guide label propagation. In Sec. 3.3.4,

3.3.2   Model of Adaptive Probabilistic Atlas

Let N denote the number of voxels in an observed image, K be the number of neuroanatomical structures in the image, $\mathbf{x}_i \in \mathbb{R}^D$ denote the observed feature vector for voxel $i \in 1, \ldots, N$. Since the MR images we use are T1-weighted images, the feature of voxel i is its T1-weighted intensity in this chapter. We model each $\mathbf{x}_i$ as an observable random variable by a K-component spatially variant gaussian mixture model [147]. We assume the density function of each tissue type follows a Gaussian distribution:

$$G(x_i|\theta_k) = \frac{1}{(2\pi)^{D/2}\sigma_k} exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}\right).\tag{3.1}$$

The tissue intensity model is parameterized by $\theta_k \equiv (\mu_k, \sigma_k)$, i.e it denotes the parameters for the kth Gaussian component where μ_k is the mean intensity of kth neuroanatomical structure and σ_k is the standard deviation of intensity. Although we model the intensity of each voxel as a mixture of all tissue types, we need to assign a discrete label to each voxel. We let $\mathbf{z}_i$ be the label vector for voxel i corresponding to one of the standard basis vectors ($\mathbf{z}_i \in \mathbf{e}_k; e_k^{j=k} = 1, e_k^{j\neq k} = 0, 1 \leq j, k \leq K$). $z_i^k = 1$ means that the estimated label at voxel i is k. We assumed x_i and $\mathbf{z}_i$ are statistically independent and model the conditional distribution of complete data $\mathbf{y} \equiv (\mathbf{x}, \mathbf{z})$ as:

$$P(\mathbf{y}|\mathbf{\Psi}) = \prod_{i=1}^{N}\prod_{k=1}^{K} [\pi_i^k G(x_i|\theta_k)]^{z_i^k}\tag{3.2}$$

where $\pi_i{}^k$ denotes the mixing coefficient of kth class at the ith voxel ($\sum_{k=1}^{K} \pi_i{}^k = 1$ and $\pi_i{}^k \geq 0$). The unknown parameters, thus, are $\mathbf{\Psi} \equiv (\pi_1, \cdots, \pi_N, \theta_1, \cdots, \theta_K)$, where $\pi_i \equiv (\pi_i{}^1, \cdots, \pi_i{}^K)$. Note that $\mathbf{z}_i$ is a discrete random variable with probability function $P(\mathbf{z}_i = \mathbf{e}_k) = \pi_i{}^k, \forall i, k$.

Since brain structure labels are piecewise constant in the spatial space, we impose label smoothness using a MRF model for the random variable $\mathbf{z}_i$ [91, 127, 90]. Our MRF uses a Gibbs distribution:

$$P(\mathbf{z}_i) = \frac{1}{Z_{norm}} exp\left\{-U(\mathbf{z}_i)\right\}, \qquad (3.3)$$

where Z_{norm} is a constant normalizer and $U(\mathbf{z}_i)$ is the prior energy function:

$$U(\mathbf{z}_i) = \sum_{j \in N_i} \mathbf{z}_i^T \hat{V} \mathbf{z}_j. \qquad (3.4)$$

where N_i is the 6-connected neighborhood of voxel i. The matrix $\hat{V} \in R^{K \times K}$ has diagonal values close to 0, and $\hat{V}(l, k) = \beta(l, k)$, for all $1 \leq l, k \leq K$ and $l \neq k$. Here $\beta(l, k)$ encodes the probability that any two types of tissues appear at adjacent voxels.

Combining the gaussian mixture model and MRF regularization we have:

$$P(\mathbf{y}|\mathbf{\Psi}) = \frac{1}{Z_{norm}} \prod_{i=1}^{N} \prod_{k=1}^{K} \left(\pi_i{}^k G(x_i; \theta_k)\right)^{z_i{}^k} \exp\left\{-\sum_{j \in N_i} \mathbf{z}_i^T \hat{V} \mathbf{z}_j\right\}. \qquad (3.5)$$

3.3.3   Construction of Target-specific Priors

To construct an adaptive probabilistic atlas, we compute several target-specific components, including a probabilistic map for individual structures, a relaxation map, and a MRF prior. The probabilistic maps are used to initialize the EM algorithm by estimating the θ_k and $\pi_i{}^k$. The relaxation map then relaxes the prior learned during the EM fitting of the target-specific atlas, while the MRF prior ensures label smoothness.

Given the label images non-linearly registered to the target, we estimate a prior probability map for each structure k at each voxel i using:

$$p_i{}^k = \frac{1}{M} \sum_{m=1}^{M} \delta_k(\hat{L}_m^i) \qquad (3.6)$$

where function $\delta_k(l) = 1$ if $l = k$, otherwise $\delta_k(l) = 0$.

Spatially Varied Adaptivity

Next, we devise a measure of the suitability of the training data for the target at every voxel. We do this so that we may vary the level of atlas adaptivity locally. Concretely, we define the relaxation map based on the label entropy at each voxel:

$$H_i = \sum_{k=1}^{K} -p_i{}^k \log \left(p_i{}^k \right). \tag{3.7}$$

A voxel with larger entropy of multi-atlas labels has greater label uncertainty. Intuitively the greater the label uncertainty the more relaxation we should apply. We employ a sigmoid transfer function to define a smooth mapping from entropy to the relaxation map ξ_i :

$$\xi_i = \frac{b_2 - b_1}{1 + exp\left(-E_i\right)} + b_1, \tag{3.8}$$

where

$$E_i = s_1(H_i - (s_2 \cdot \min\left(H_i\right) + (1 - s_2) \cdot \max(H_i))), \tag{3.9}$$

b_1 and b_2 control the lower bound and upper bound, $s_1 \geq 1$ is a parameter for scaling, and $s_2 \in (0, 1)$ is a parameter for shifting the inflection point of the sigmoid curve. Fig. 3.3 illustrates our process and the plot in the center shows the transfer function's parameterization where $b_1 = 0.0, b_2 = 0.5, s_1 = 3.0, s_2 = 0.1$, and $0 \leq H_i \leq 20$. Since adaptivity should be mainly applied along the tissue class boundaries, we set these parameters such that the relaxation is much lower near the center of structures and higher along their boundaries. Small subcortical structures may have larger uncertainty compared with larger structures. To overcome such bias, we use a local mapping scheme. Let $\mathbf{X}_k = \{i|p_i{}^k > 0\}$ denote the subset of voxels that could belong to the structure k. We apply the mapping function on each structure individually to compute structure-based $\xi_i{}^k = \{\xi_i | i \in \mathbf{X}_k\}$. Then we compute the whole relaxation map $\xi_i = \sum_{k=1}^{K} \xi_i{}^k/n_i$ where n_i is the number of subsets $\mathbf{X}_k$ that the ith voxel is contained in.

Structure-level Relaxation

Several structures should receive greater adaptivity due to their geometric complexity. Thus, our relaxation map is a function of both spatial location *and* structure k. We
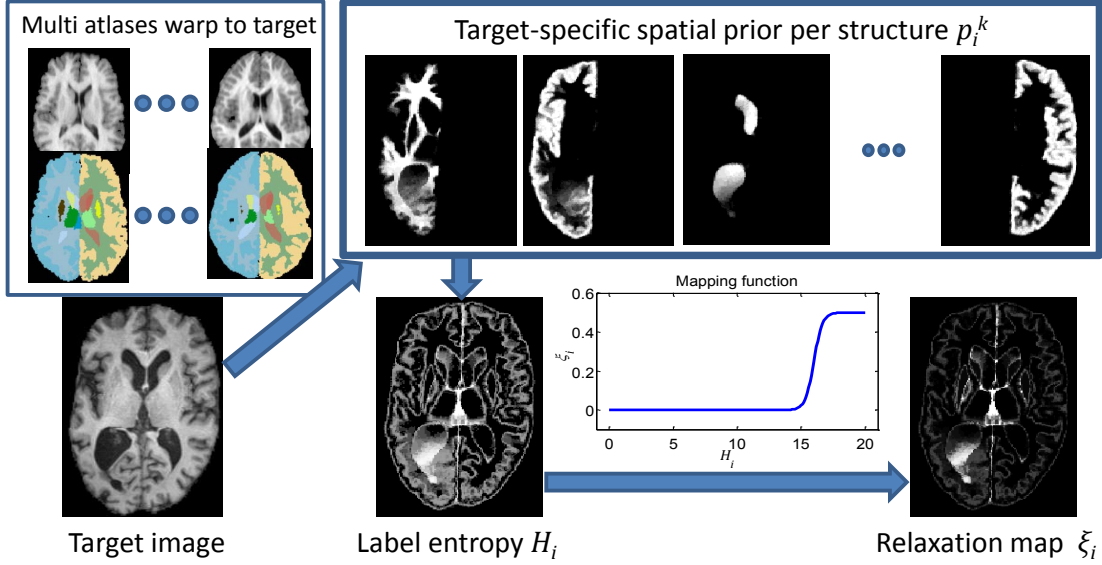
Figure 3.3: This figure illustrates the computation of relaxation map ξ_i from multi-atlas label entropy H_i.

define $\xi_i(k)$ as:

$$\xi_i(k) = \begin{cases} 1 - \alpha_1(1 - \xi_i) & if(i \in \{\mathbf{X}_k | \forall k \in \{k_{GM's}\}\}) \\ 1 - \alpha_2(1 - \xi_i) & if(i \in \{\mathbf{X}_k | k = k_{ECSF}\}) \\ 1 - \alpha_3(1 - \xi_i) & if(i \in \{\mathbf{X}_k | \forall k \in \{k_{WM's}\}\}) \\ \xi_i & otherwise \end{cases} \tag{3.10}$$

where $k_{GM's}$ include left and right cerebral and cerebellar cortex, $k_{WM's}$ include left and right cerebral and cerebellar WM, and $\alpha_1, \alpha_2, \alpha_3 \in (0..1)$ are empirically determined, structure-specific relaxation coefficients. For GM, $\alpha_1 = 1$ corresponds to no structure specific relaxation, while $\alpha_1 = 0$ corresponds to full relaxation.

Pairwise Label Smoothness Prior

To achieve spatially smooth label maps, our atlas includes a MRF model of the probability of structures l and k appearing at adjacent voxels. We estimate our MRF model parameters $\beta(l, k)$ in 3 steps:

1. Compute the class pairwise neighboring probabilities through a whole atlas. Let $P_{L_m}(l, k)$ denote the probability that a voxel with label k appears next to the
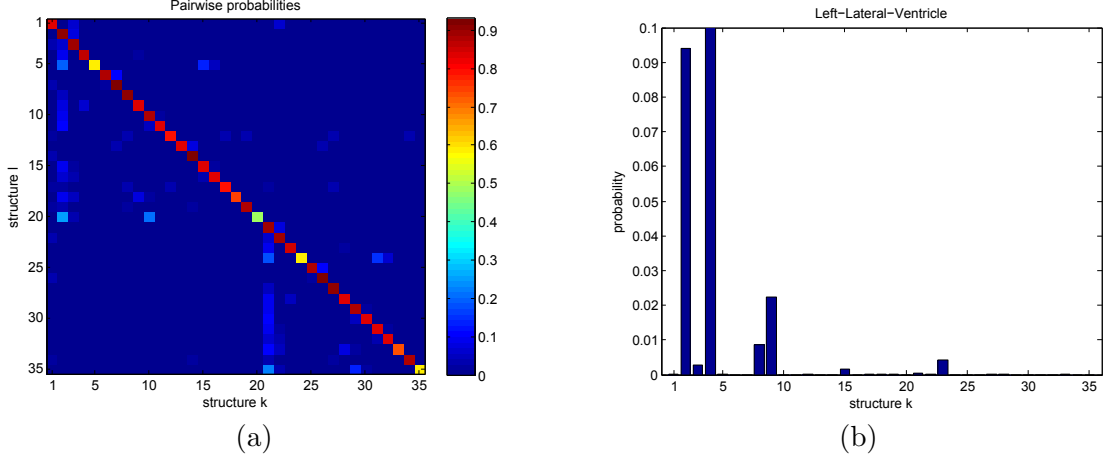
Figure 3.4: Learned neighborhood properties: (a) pairwise structure probabilities $P_{L_m}(l, k)$, note that the matrix is not symmetric because the value at (l, k) is normalized by the size of structure l as in Eq. 3.11; (b) $P_{L_m}(l = 4, k)$ for left lateral ventricle. The detailed label definition can be found in Table 3.1.

voxel with label l in label image L_m. Similar to [106], we define

$$P_{L_m}(l, k) = \frac{\left| \{ \langle i, j \rangle | j \in N_i, z_i{}^l = 1, z_j{}^k = 1, \forall 1 \leq i \leq N \} \right|}{\left| \{ \langle i \rangle | z_i{}^l = 1, \forall 1 \leq i \leq N \} \right|}. \qquad (3.11)$$

The pairwise probabilities from our atlas are plotted in Fig. 3.4.

2. Average P_{L_m} over all training atlases $m \in \{1, \ldots, M\}$.

3. Map the $P_{L_m}(l, k)$ to value $\beta(l, k) \in [b_3, b_4]$. A larger $P_{L_m}(l, k)$ should be mapped to a smaller $\beta(l, k)$, so the corresponding MRF model has a larger probability to connect the class pair $\langle l, k \rangle$. We form the matrix $\hat{V}$ for our complete brain model via the mapping function:

$$\beta(l, k) = -log(P_{L_m}(l, k)) \frac{b_4 - b_3}{max_{\forall 0 \leq l, k \leq K}(-log(P_{L_m}(l, k)))} + b_3, \qquad (3.12)$$

where

$$P_{L_m}(l, k) = \begin{cases} \gamma & if(P_{L_m}(l, k) < \gamma) \\ P_{L_m}(l, k) & otherwise \end{cases}. \qquad (3.13)$$

To keep P_{L_m} finite, we impose a minimum, γ. In our training data the number of voxels per volume is close to 10^7 therefore we choose $\gamma = 10^{-7}$.

3.3.4 Adaptive Label Propagation

Step four of our approach is to evolve the target-specific, adaptive probabilistic atlas until it fits the target. The complete data is $\mathbf{y} \equiv (\mathbf{x}, \mathbf{z})$ however we have observed only the target intensity volume, $\mathbf{x}$. In this case the MAP estimate of $\boldsymbol{\Psi}$ with latent labeling $\mathbf{z}$ is:

$$\hat{\boldsymbol{\Psi}}_{MAP} = \underset{\psi \in \Psi}{\operatorname{argmax}} \ln p(\boldsymbol{\Psi}|\mathbf{y}) = \underset{\psi \in \Psi}{\operatorname{argmax}} \{\ln p(\mathbf{y}|\boldsymbol{\Psi}) + \ln p(\boldsymbol{\Psi})\} \tag{3.14}$$

Since it is difficult to solve this directly, we use the Expectation-Maximization (EM) algorithm, which makes an initial guess about the complete data including the unknown parameters in $\boldsymbol{\Psi}$ and solves for the $\boldsymbol{\Psi}$ that maximizes the posterior of $\boldsymbol{\Psi}$ given $\mathbf{y}$. Let t denote the EM iteration counter. We initialize $\boldsymbol{\pi}^{(t=0)}$ using the learned prior information and $\boldsymbol{\theta}^{(t=0)}$ using the statistical information in the target image. Then iterate to update the parameters $\pi_i{}^k$ and θ_k and in the process find the labeling $\mathbf{z}$.

E-step: Given our observed data, $\mathbf{x}$, and our current estimate of the distribution parameters, $\boldsymbol{\Psi}^{(t)}$, we compute how likely the complete data is exactly $\mathbf{y}$ by computing the conditional expectation:

$$Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^{(t)}) = E_{\mathbf{y}|\mathbf{x},\boldsymbol{\Psi}^{(t)}}\{\ln P(\mathbf{y}|\boldsymbol{\Psi})\} \tag{3.15}$$

Assuming voxels are independent, we model the log function in Eq. 3.15 as follows:

$$\ln P(\mathbf{y}|\boldsymbol{\Psi}) = \ln p(\mathbf{x}, \mathbf{z}|\boldsymbol{\Psi}) = \ln p(\mathbf{x}|\mathbf{z}, \boldsymbol{\Psi}) + \ln p(\mathbf{z}|\boldsymbol{\Psi})$$
$$= \sum_{i=1}^{N} \ln p(x_i|\mathbf{z}_i, \boldsymbol{\Psi}) + [-U(\mathbf{z}|\boldsymbol{\Psi}) - \ln Z_{norm}(\boldsymbol{\Psi})] \tag{3.16}$$

We model the first term in Eq. 3.16 as:

$$\ln p(x_i|\mathbf{z}_i, \boldsymbol{\Psi}) = T(x_i|\boldsymbol{\Psi})\mathbf{z}_i \tag{3.17}$$

where $T(x_i|\boldsymbol{\Psi}) = (\ln(\pi_i{}^1 G(x_i; \theta_1)), \ldots, \ln(\pi_i{}^K G(x_i; \theta_K)))$. Thus,

$$Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^{(t)}) = \sum_{i=1}^{N} E\{\mathbf{z}_i|x_i, \boldsymbol{\Psi}^{(t)}\}T(x_i|\boldsymbol{\Psi}) + E\{-U(\mathbf{z}) - \ln Z_{norm}\}. \tag{3.18}$$

We estimate the main component in Eq. 3.18 is $E\{\mathbf{z}_i|x_i, \boldsymbol{\Psi}^{(t)}\}$ and this posterior probability using mean field approximation [148]:

$$E\{z_i{}^k|x_i, \theta_k^{(t)}\} \equiv \omega_{ik}^{(t)} \approx \frac{p\left(x_i|z_i{}^k = 1, \theta_k^{(t)}\right) p\left(z_i{}^k = 1|E\left\{z_{j \in N_i}\right\}^{(t-1)}\right)}{\sum_{k=1}^{K} p(x_i|z_i{}^k = 1, \theta_k^{(t)}) p(z_i{}^k = 1|E\left\{z_{j \in N_i}\right\}^{(t-1)})} \tag{3.19}$$

where

$$p(x_i|z_i{}^k = 1, \theta_k^{(t)}) = G(x_i; \mu_k^{(t)}, \Sigma_k^{(t)}), \tag{3.20}$$

$$p(z_i{}^k = 1|E\{z_{j \in N_i}\}^{(t-1)}) = \frac{exp\{-\sum_{j \in N_i} \sum_{l=1, l \neq k}^{K} \beta(k, l)\omega_{jl}^{(t-1)}\}}{\sum_{k=1}^{K} exp\{-\sum_{j \in N_i} \sum_{l=1, l \neq k}^{K} \beta(k, l)\omega_{jl}^{(t-1)}\}}. \tag{3.21}$$

M-step: We optimize the unknown parameters, $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$, of our probability distributions. Specifically the M-step maximizes $Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^{(t)}) + \ln p(\boldsymbol{\Psi})$ over $\psi \in \Psi$ to find:

$$\boldsymbol{\Psi}^{(t)} = \underset{\psi \in \Psi}{\operatorname{argmax}}(Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^{(t)}) + \ln p(\boldsymbol{\Psi})) \tag{3.22}$$

Assuming uniform priors on θ_k, we maximize the Gaussian parameters as:

$$\mu_k^{(t+1)} = \frac{\sum_{i=1}^{N} \omega_{ik}^{(t)} x_i}{\sum_{i=1}^{N} \omega_{ik}^{(t)}} \tag{3.23}$$

$$\sigma_k^{(t+1)} = \sqrt{\frac{\sum_{i=1}^{N} \omega_{ik}^{(t)}(x_i - \mu_k^{(t+1)})(x_i - \mu_k^{(t+1)})^T}{\sum_{i=1}^{N} \omega_{ik}^{(t)}}} \tag{3.24}$$

Maximizing Eq. 3.22 is achieved by taking the partial derivative with respect to π_i^k:

$$\frac{\partial}{\partial \pi_i^k}\{Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^{(t)}) + \ln p(\boldsymbol{\Psi}) + \lambda(1 - \sum_{k=1}^{K} \pi_i^k)\}|_{\pi_i^{k(t+1)}} = 0, \tag{3.25}$$

where λ is a Lagrange multiplier. We assume π_i follows a Dirichlet distribution (inspired by [107]). In a Dirichlet distribution, the key parameter is the concentration parameter, which represents the but our model varies spatically. with the relaxation map, ξ_i:

$$Dir(\pi_i^1, \cdots, \pi_i^K; \xi_i, p_i^1, \cdots, p_i^K) = \frac{1}{B(1 + (1/\xi_i - 1)p_i)} \prod_{k=1}^{K} (\pi_i^k)^{(1/\xi_i - 1)p_i^k}, \tag{3.26}$$

where B is a Beta function, p_i^k is the prior probability that voxel i is labeled as class k and ξ_i is the relaxation value at voxel i. After substituting Eq. 3.18 and Eq. 3.26 into Eq. 3.25, we have

$$\frac{\partial}{\partial \pi_i^k}\{(\omega_{ik}^{(t)} + (1/\xi_i - 1)p_i^k)\ln \pi_i^k - \lambda\pi_i^k\}|_{\pi_i^{k(t+1)}} = 0, \tag{3.27}$$

and thus:

$$\pi_i^{k(t+1)} = \frac{\omega_{ik}^{(t)} + (1/\xi_i - 1)p_i^k}{\lambda}. \tag{3.28}$$

Since $\sum_{k=1}^{K} \pi_i^{k(t+1)} = 1$, we have $\lambda = \sum_{k=1}^{K} (\omega_{ik}^{(t)} + (1/\xi_i - 1)p_i^k) = 1/\xi_i$. Thus, the mixing coefficients are updated by:

$$\pi_i^{k(t+1)} \approx \xi_i(G * \omega_{ik}^{(t)}) + (1 - \xi_i)p_i{}^k. \tag{3.29}$$

Here G is a Gaussian kernel to smooth the parameter distribution.

3.4   Materials

3.4.1   Datasets

While our method is largely disease agnostic, to make this study concrete we evaluate the method using four public brain datasets covering subjects with healthy and Alzheimer's diseased brains. Three of these datasets contain images of healthy brains: **IBSR**, **BrainWeb** and **MICCAI2012**. The **IBSR** dataset[3] contains 18 subjects with bias corrected, T1-weighted intensity volumes, and label volumes delineated by expert neuroanatomists. Each volume has <1mm voxel resolution in the coronal view and 1.5mm spacing between coronal slices. Each segmentation has labels for 34 structures (Table 3.1) encompassing multiple cortical and subcortical structures from each hemisphere. The **BrainWeb** dataset [149, 150] contains 12 anatomical labels on 20 T1 intensity images from a simulated spoiled FLASH pulse sequence with TR=22ms, TE=9.2ms, flip angle=30 degree, and 1mm isotropic voxel size. The discrete labels, representing the tissue which contributes the most to that voxel, include background, CSF, gray matter, white matter, fat, muscle, muscle/skin, skull, vessels, dura matter, and bone marrow. The **MICCAI2012** dataset from MICCAI 2012 Grand Challenge at the Workshop on Multi-Atlas Labeling [140] has 15 T1-weighted structural MRI training brains and 20 testing brains.[4] Each image data has 1mm isotropic voxel size. The training subjects come from 10 female and 5 male, whose age range from 19 to 34. The testing subjects come from 10 female and 5 male, whose age range from 18 to 90. Every image (train and test) has cortical and sub-cortical labels annotated by medical experts.

The diseased brains are from the **AIBL** database [151], which contains patients ranging from mild cognitive impairment (MCI) through fully developed Alzheimer's

---

[3]The IBSR (Internet Brain Segmentation Repository) MR brain datasets and their manual segmentations were provided by the Center for Morphometric Analysis, http://www.cma.mgh.harvard.edu/ibsr/.

[4]The MRI scans were obtained from the Open Access Series of Imaging Studies (OASIS) project (http://www.oasis-brains.org/), and the labels were provided by Neuromorphometrics, Inc. (http://Neuromorphometrics.com/).

Table 3.1: The structures our algorithm automatically labels.

ID	Structure name	ID	Structure name
1	Background (ECSF)		
2	L-Cerebral-WM	21	R-Cerebral-WM
3	L-Cerebral-Cortex	22	R-Cerebral-Cortex
4	L-Lateral-Ventricle	23	R-Lateral-Ventricle
5	L-Inf-Lat-Vent	24	R-Inf-Lat-Vent
6	L-Cerebellum-WM	25	R-Cerebellum-WM
7	L-Cerebellum-Cortex	26	R-Cerebellum-Cortex
8	L-Thalamus-Proper	27	R-Thalamus-Proper
9	L-Caudate	28	R-Caudate
10	L-Putamen	29	R-Putamen
11	L-Pallidum	30	R-Pallidum
12	3rd-Ventricle		
13	4th-Ventricle		
14	Brain-Stem		
15	L-Hippocampus	31	R-Hippocampus
16	L-Amygdala	32	R-Amygdala
17	CSF		
18	L-Accumbens-area	33	R-Accumbens-area
19	L-VentralDC	34	R-VentralDC
20	L-vessel	35	R-vessel

(AD). 28 subjects with AD were randomly chosen. A subset of nine (denoted as SevereVent_AD) were subsequently found to have severely enlarged lateral ventricles, the remaining 19 patients (denoted as ModerateVent_AD) have moderately enlarged lateral ventricles.[5] An expert anatomist manually labeled the brains in SevereVent_AD using 3D-Slicer [152].

3.4.2 Method Variants

In order to test the additive value of each aspect of *iMAEASA*, we constructed variations of the proposed approach that largely consist of subsets of the full method. These variants are summarized here.

ASA is the adaptive statistical atlas and it is a naive extension of the method in [107] that segments additional structures. This is achieved by simply adding more components to the Gaussian mixture model. Compared to *iMAEASA* this method retains the spatially varying mixing coefficients, but lacks the spatially

---

[5]Our designations correspond to visible anatomical size of the ventricles which is related to but not the same as disease severity.

varying relaxation map, structure specific relaxation and MRF smoothness priors, and lacks the target-specific spatial priors.

EASA adds the spatially varying relaxation map to ASA based on voxel label uncertainty; however the uncertainty is computed among the training volumes themselves. This map is then non-linearly registered to the target. Compared to iMAEASA, EASA [141] lacks structure specific relaxation and MRF smoothness priors, and lacks the target-specific spatial priors.

iEASA adds the structure-specific relaxation and pairwise neuroanatomical structure-specific MRF priors to EASA. Compared to iMAEASA, iEASA lacks the target specific spatial priors and the MRF priors are computed without regard to the target. This method is illustrated in (Fig. 3.5). We call this method iEASA because it is an Intensity-driven Extended Adaptive Statistical Atlas approach. It has two parts: a target-independent model construction (top of Fig. 3.5) and an application of model to the target (bottom of Fig. 3.5). In the training step, we choose one atlas at random as the frame of reference and non-rigidly register all of the training brains to it. We use these warped atlases to construct the adaptive probabilistic atlas which includes neuroanatomical structure-specific relaxation and pairwise structure-specific MRF prior. Note that an average intensity image needs to be computed from training intensity images, e.g. through a groupwise registration method [153]. To segment a novel target, we non-rigidly register the atlas to the target and use EM to fit the atlas to the target.

iMAEASA is our proposed method. It utilizes aspects from both the multi-atlas genre and the adaptive probabilistic atlas genre. The training atlases are non-linearly registered to the target and all components of the probability model are target specific at construction and further evolved to fit the target using EM optimization.

Among these variants, the first three (ASA, EASA and iEASA) integrate training information prior to observing the target subject and are hence probabilistic atlas
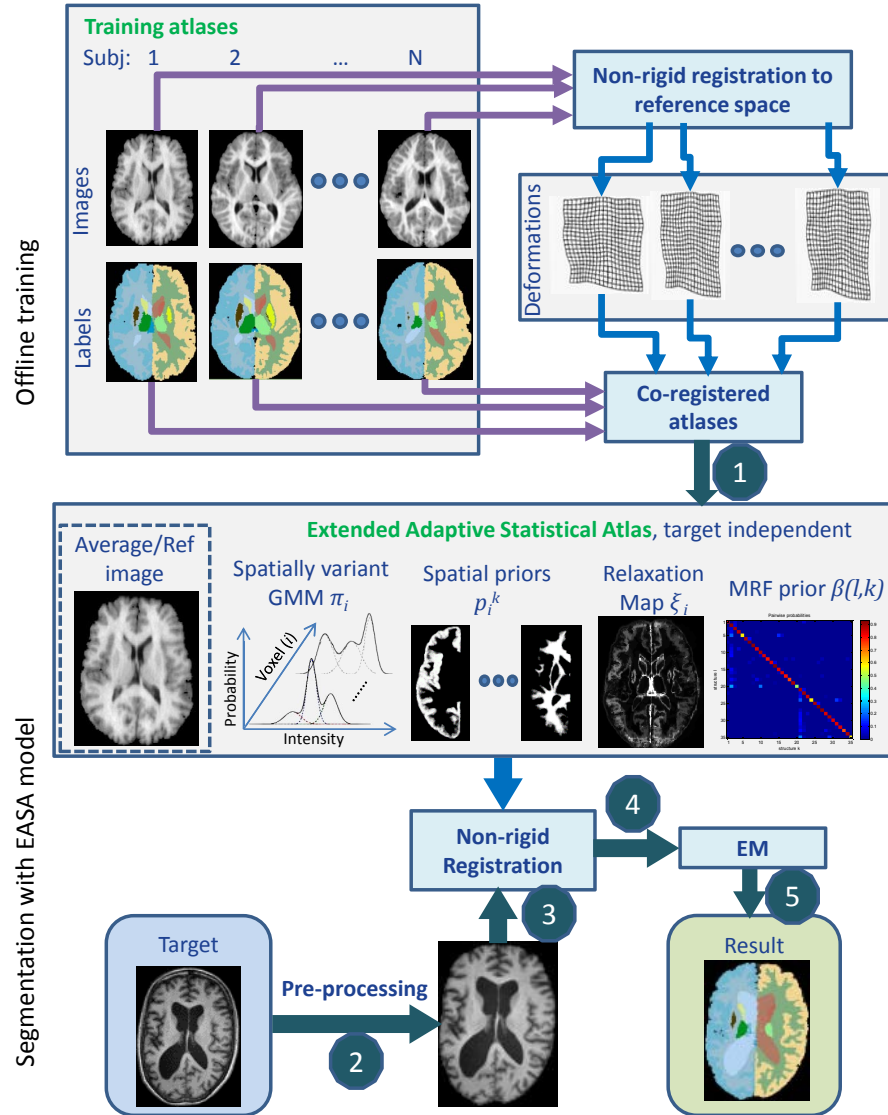
Figure 3.5: Workflow of a variant method called iEASA for fast labeling. Significant speed up is achieved by training offline, before the target arrives.

based approaches, while *iMAEASA*, which registers training subjects individually to the target, is amulti-atlas approach.

3.5 Experiments

3.5.1 Experimental Settings

We conduct several experiments to measure how well our method generates accurate parcellations in different scenarios. The first scenario tests our method on challenges faced in the clinic in which the target population contains diseased brains while the labeled training data does not. In the second scenario, our goal is to measure our method's performance when the target population is fairly well represented by the training data. Though this situation may be less realistic clinically and does not tend to tax our method's ability to generalize, it provides a reference point for comparison with many published results which only explore this situation. In the last experimental scenario, we explore the effect of the representativeness of training data to the target population by using different combinations of diseased and healthy brains in the training data.

We compare our method to *seven* (7) previously published methods by running them directly on the same data using the parameter settings recommended by the authors. (1) We compare against the widely used probabilistic atlas method FreeSurfer [106, 154], which is freely available. Due to its widespread adoption, it has also become a benchmark method for parcellation algorithm development. (2) We compare against a locally weighted label fusion method, iSTAPLE, [99], which improves upon the STAPLE [98], by adding target intensities as an additional guidance term. (3-5) We compare against three patch-based label fusion methods with best reported results. These include the top two performing methods from the 2012 MICCAI grand challenge: the first place, joint label fusion method (JLF) [100] and the second place, non-local STAPLE approach (NLS) [105]. We also compare to the recent generalization of joint label fusion (gJLF) method by [104]. (6-7) Lastly, we compare against two methods specifically developed to have good generalization. These include the locally adaptive

brain segmentation algorithm (LoAd) [89] and a naive extension to the adaptive probabilistic atlas method (ASA) [107] to handle 30+ structures rather than the original 4 structures.

We evaluate the accuracy of the parcellation of the brain structures listed in Table 3.1. To measure labeling accuracy we compute the overlapping scores, namely Dice Similarity Coefficient (DSC), between true and computed label maps A and B (respectively) as:

$$DSC(A, B) = \frac{2\,|A \cap B|}{|A| + |B|} \tag{3.30}$$

where $|.|$ denotes set cardinality. For *iMAEASA*, dice scores are measured when the EM algorithm converges, which we detect when the relative difference of log-likelihoods between two successive iterations is less than 0.0005.

3.5.2 From Healthy Brains to Diseased Brains

The first set of experiments tests our method's ability to handle difficult cases that can be commonly found in the clinic. These include cases in which the target contains a brain with visible abnormalities not found in the healthy training data. These tests measure how well a method can generalize learned priors. We use the 18 healthy brain **IBSR** dataset as training and the 28 AD diseased subjects from the AIBL dataset for testing. We separate the diseased subjects into two subsets: the 9 subject SevereVent_AD and 19 subject ModerateVent_AD based on the severity of the ventricular enlargement.

Figure 3.6: Direct comparison of the performance our method against many methods with recently reported high accuracies. Target subjects have AD and severely enlarged ventricles. Cortical, subcortical, and ventricular regions with visually superior parcellation using *iMAEASA* are circled (yellow).

Representative results on subjects with severely enlarged ventricles are shown in Fig. 3.6. From left to right, the columns show: an axial image[6] from each subject, expert's manually labeled ground truth, and automatically generated labels from our proposed method, *iMAEASA*, and from the comparison methods. Each row compares the results from one subject. There is good agreement between *iMAEASA* and the true parcellation. For many regions (circled), *iMAEASA* performs visually better than the comparison methods. *iMAEASA* is the only method able to accurately segment both the ventricles and cortex of these diseased brains. In particular FreeSurfer and the multi-atlas methods (iSTAPLE, JLF, NLS, gJLF) are unable to label the diseased ventricles well. For the cortex, both *iMAEASA* and LoAd produce visually accurate cortical segmentations. FreeSurfer tends to under-segment it, and the remaining multi-atlas methods tend to over-segment the cortex. Note that all the methods use the same warped multi-atlas[7] and ASA method in this section uses the warped multi-atlas to compute its target-specific probabilistic atlas.

A quantitative comparison is shown in Fig. 3.7. To increase readability, representative methods from Fig. 3.6 are selected for the plot. For the first 18 bar groups, each bar shows the DSC statistics for one structure across all subjects in SevereVent_AD for a method. In the last group, each bar shows the average for all structures except vessels for each method. In terms of the overall mean DSC, our method is better than all other comparison methods and has the smallest variance. Compared to FreeSurfer, *iMAEASA* performs significantly better ($p < 0.05$) on most structures. Compared to NLS and JLF, *iMAEASA* performs better on multiple structures with higher mean and lower variance, especially on cerebral WM, GM and ventricles. To boost the performance of JLF, we tried adding label correction [139] to JLF, however this decreased its accuracy and suggests that correctors learned from normal brains may not be suitable for improving diseased brain accuracy.

We further explored the use of machine learning based correction, this time assuming

---

[6]Intensity data and parcellations are volumetric

[7]Since the software for gJLF method requires ANALYZE image format, all input volumes are converted from NIFTI format to ANALYZE format by a public matlab tools for NIfTI and ANALYZE image.
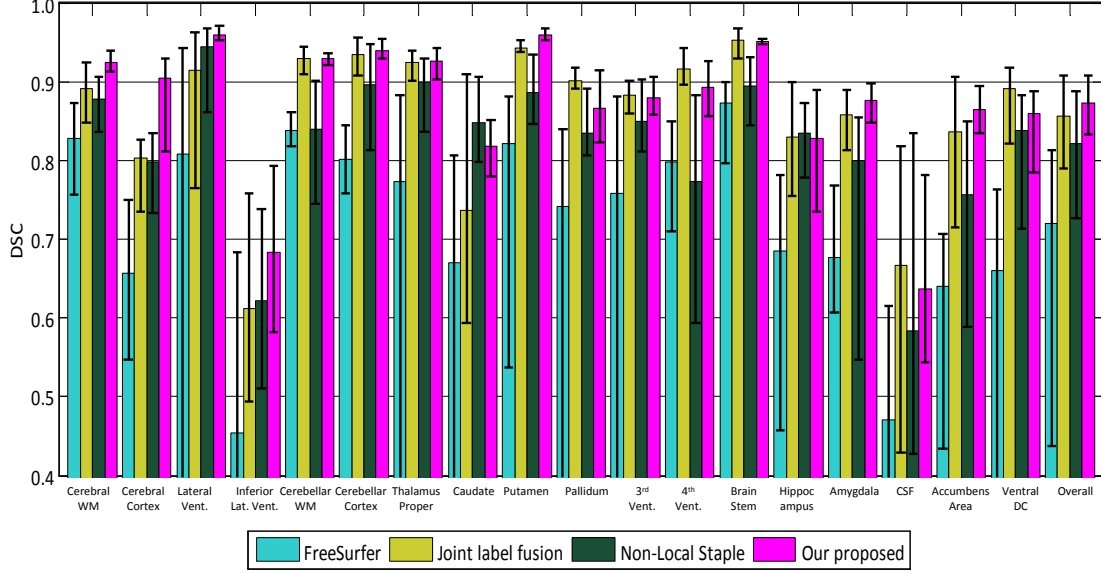
Figure 3.7: Quantitative comparison of dice scores on SevereVent_AD dataset from different approaches. We combine the left and right structures and report the mean score for each structure across all subjects. The average score of all structures except vessels are also reported.

the training data is representative of the target population (similar to the the MICCAI 2012 grand challenge). Thus, we ran experiments with 3-fold cross validation using, in each fold, 6 of the SevereVent_AD diseased brains for training and the remaining 3 diseased brains for testing. This increased the overall mean DSC (over all folds) to 0.89 for JLF. We trained learning based correctors using the results from *iMAEASA*, and this also increased the overall mean DSC to 0.89. This suggests that when training data is representative of the target population, the methods (JLF or *iMAEASA* with correction) perform similarly well.

We also ran tests on the 19 AD diseased brains in ModerateVent_AD which have moderately enlarged ventricles. We do not have ground truth for this particular dataset so we conduct a qualitative comparison of the methods by rendering the parcellations from *iMAEASA* and the compared methods side-by-side with the input image. Fig. 3.8 shows 5 subjects that are representative of the results. We observe that overall, the results from *iMAEASA* appear to have better agreement with the structures visible in the images than the compared methods.
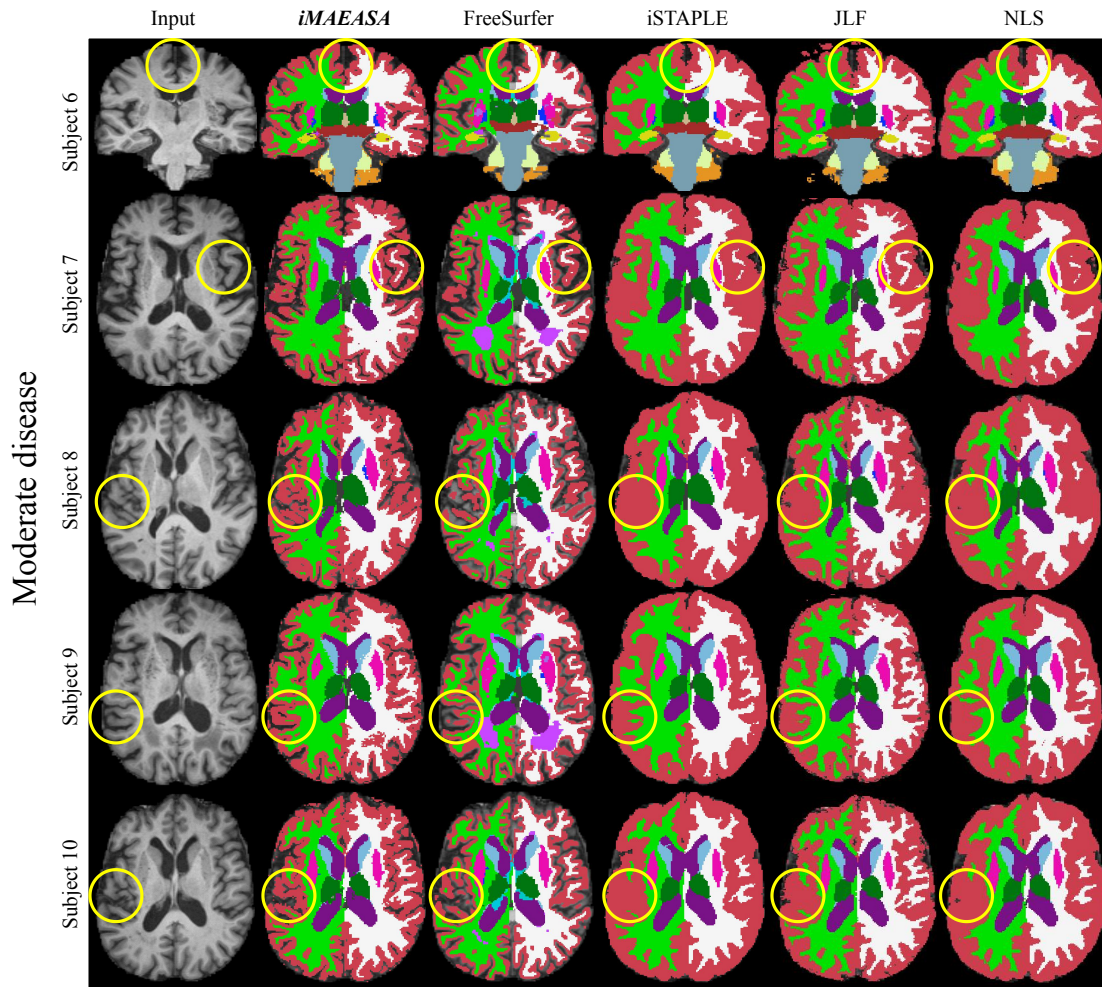
Figure 3.8: Visual comparison of results on moderately enlarged ventricles from FreeSurfer and several multi-atlas based methods. *iMAEASA* produces reliable results for most brain regions, and better results for cortex.

3.5.3 Target Brains Well Represented by Training Data

Traditionally, researchers have conducted experiments with target brains containing variation (anatomical and/or intensity) that is spanned well by the training dataset. So that our results are comparable with the literature, we also conduct experiments that test this situation. Many published methods implicitly assume that the target population is well represented by the training data and work reasonably well in this scenario. Since little adaptivity is required in this situation, these experiments actually test our methods ability to tone down adaptivity automatically.

We extensively evaluate our method on 3 datasets in 3 experiments. First we conduct an experiment on the **IBSR** dataset to quantify how well both our proposed method, *iMAEASA* and the variant, iEASA parcellate healthy brains using healthy brain training data. To achieve this we conduct a leave-one-out cross validation experiment. Table 3.2 lists the results of our proposed *iMAEASA* and its variant iEASA, as well as comparison methods including ASA (the naive extension to the method proposed in [107]), FreeSurfer and iSTAPLE [99].

We observe that *iMAEASA* works quite well and is comparable to the best methods. This is in part because the entropy calculation automatically measures the representativeness of the training data to the target and adjusts the adaptivity accordingly. In this experiment, the low label uncertainty causes our method to tone down its adaptiveness and use the training data more than the adaptivity. Additionally in the table we observe that (a) *iMAEASA* and iSTAPLE produce best mean DSC results for individual structures (**boldface**) across the database as well as overall mean accuracy across all structures in all subjects, (b) iEASA performs better than ASA and FreeSurfer; (c) multi-atlas based methods (iSTAPLE and *iMAEASA*) tend to have better accuracy than probabilistic-atlas based methods. Lastly, the table also shows the computation time. We observe a 100x computation time difference between multi-atlas methods and probabilistic-atlas methods. This is due to two factors. (1) ASA and iEASA require only one online registration of a single atlas to the target image rather than the multiple registrations in iSTAPLE and *iMAEASA*. (2) ASA and iEASA utilize a GPU

Table 3.2: Segmentation accuracy (mean DSC) across 18 healthy subjects. 'Overall' row shows average score of all structures except vessels. Bottom row shows the computation time per subject. Bold face indicates scores that are best or statistically indistinguishable from best.

Structure ID	ASA	iEASA	FreeSurfer	iSTAPLE	*iMAEASA*
L&R-Cerebral-WM	0.9	0.9	0.88	0.9	**0.91**
L&R-Cerebral-Cortex	0.88	0.88	0.8	**0.89**	**0.9**
L&R-Lateral-Ventricle	0.85	0.85	0.8	**0.87**	**0.87**
L&R-Inf-Lat-Vent	0.54	0.52	0.43	0.49	**0.61**
L&R-Cerebellum-WM	0.88	0.88	0.83	**0.89**	**0.89**
L&R-Cerebellum-Cortex	0.88	0.89	0.87	**0.91**	**0.91**
L&R-Thalamus-Proper	0.82	0.87	0.85	**0.9**	**0.89**
L&R-Caudate	0.80	0.84	0.82	**0.85**	**0.86**
L&R-Putamen	0.828	0.88	0.82	**0.9**	**0.9**
L&R-Pallidum	0.81	0.82	0.78	**0.84**	**0.84**
3rd-Ventricle	0.68	0.73	0.71	**0.78**	**0.76**
4th-Ventricle	0.73	0.80	0.77	**0.83**	**0.83**
Brain-Stem	0.84	0.88	0.85	**0.92**	0.9
L&R-Hippocampus	0.76	0.77	0.76	**0.82**	**0.82**
L&R-Amygdala	0.72	0.73	0.70	**0.76**	**0.77**
CSF	0.49	0.54	0.45	**0.63**	**0.57**
L&R-Accumbens-area	0.64	0.68	0.58	**0.75**	**0.75**
L&R-VentralDC	0.76	0.80	0.70	**0.85**	0.83
L&R-vessel	0.29	0.23	0.31	**0.38**	**0.38**
Overall	0.77	0.80	0.75	0.82	0.82
Time (minutes)	$\approx$6.5	$\approx$ 7	$\approx$ 720	$\approx$660	$\approx$667
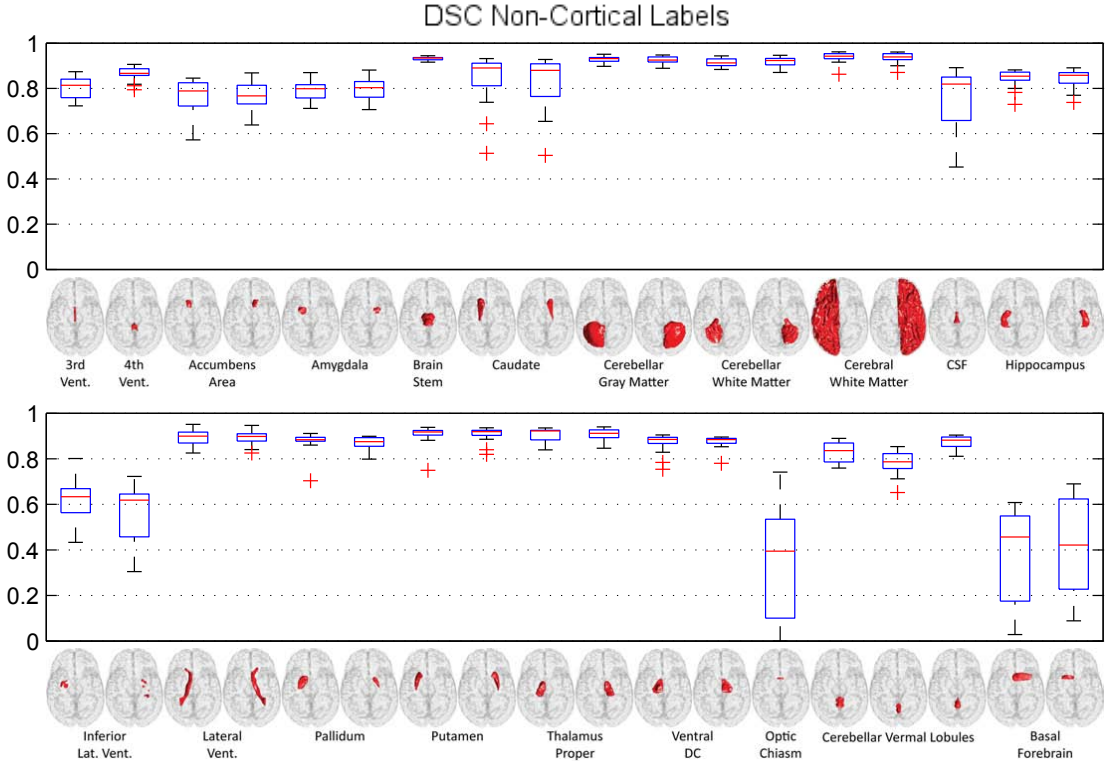
Figure 3.9: Dice scores from the proposed *iMAEASA* approach on the MICCAI12 grand challenge data for structures throughout the brain including non-cortical structures.
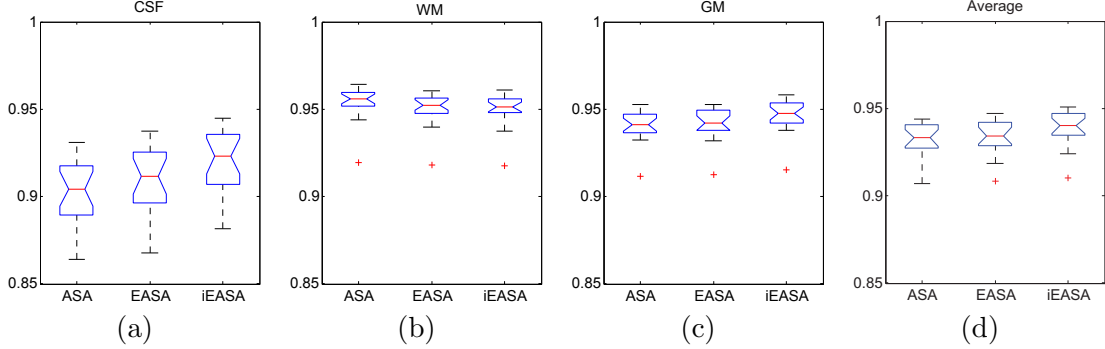
Figure 3.10: Statistical comparison using the **BrainWeb** dataset. Dice scores from three adaptive probabilistic atlas methods for (a) CSF, (b) white matter, (c) gray matter, and (d) averaged across these structures.

accelerated spline-based method by [155] which is much faster than the more accurate Symmetric Normalization method with cross-correlation [129] used in iSTAPLE and *iMAEASA*.

To compare our method to additional state-of-the-art approaches, we conduct another experiment using the **MICCAI2012** data [140]. Our algorithm does not currently parcellate the functional regions of the cortex whose boundaries are largely invisible in structural MRI; therefore we include dice scores for the other structures (e.g. non-cortical regions) as shown in Fig. 3.9. We use the same train/test split as the challenge and observe that *iMAEASA*'s overall mean dice score for non-cortical structures is 0.81 ± 0.026 over all subjects, while the top ranked method, joint label fusion with label correction, achieves 0.8377 ± 0.028.

We also use the **BrainWeb** dataset of 20 subjects to compare several of the variants of our approach, iEASA, ASA and EASA, to one another. For the **BrainWeb** dataset, we evaluate the accuracy of gray matter, white matter, and CSF. We merge its other labels to a background label. Fig. 3.10 shows box plots of the dice score distribution from these methods. On healthy brains, we observe that the iEASA produces the best results with higher mean accuracy per structure than the other methods. Using a paired T-test (significance level 0.05), we find that iEASA is statistically significantly better than ASA for both CSF and for the average accuracy over the three structures.

3.5.4 Impact of Representativeness of Training Data

So far, we have shown that the proposed *iMAEASA* outperforms other methods when parcellating diseased brains using priors learned from publicly available healthy brains. In the next suite of experiments, we study the impact of varying the training data to various combinations of healthy and diseased subjects. We do this by constructing three additional training sets:

H16 contains 16 healthy brains randomly chosen from the 18 brain **IBSR** set.

M16 contains a mixture of 8 healthy brains and 8 diseased from SevereVent_AD set.

D8 contains 8 brains randomly chosen from the 9 brain SevereVent_AD diseased brains.

We compare the performance of all 9 combinations of three segmentation approaches: ASA, iEASA and *iMAEASA* and these three atlases. For each ⟨method × atlas⟩ combination we apply leave-one-subject-out cross-validation to compute the test accuracy across all subjects in the SevereVent_AD dataset and then on the **IBSR** dataset. Our test results on diseased brains from SevereVent_AD are shown in Fig. 3.11. We can see that there is an upward slope on accuracies for almost all structures. Since the order of the algorithms corresponds to increasing addition of components (building up to *iMAEASA*), this provides evidence of the value of each of the components of *iMAEASA*. Specifically, probabilistic initialization (*iMAEASA*) outperforms deterministic (iEASA), and entropy-based, non-stationary relaxation (*iMAEASA*, iEASA) outperforms stationary relaxation (ASA). The figure also shows that to segment diseased brains, the inclusion of some diseased brains in the training set (e.g. M16 or D8) can slightly improve the accuracy. *iMAEASA* is better than other two methods. Using ⟨*iMAEASA* × H16⟩ as the reference, those combinations which vary significantly (at the 0.05 confidence level) are shown with an * above the bar. Notably we observe that for average dice score over all structures (the rightmost group of bars) no significant difference is found among the ⟨*iMAEASA* × training⟩ for the three training datasets H16, M16, and D8. This supports our previous experimental results showing that *iMAEASA*'s adaptivity enables diseased brains to be segmented whether or not the
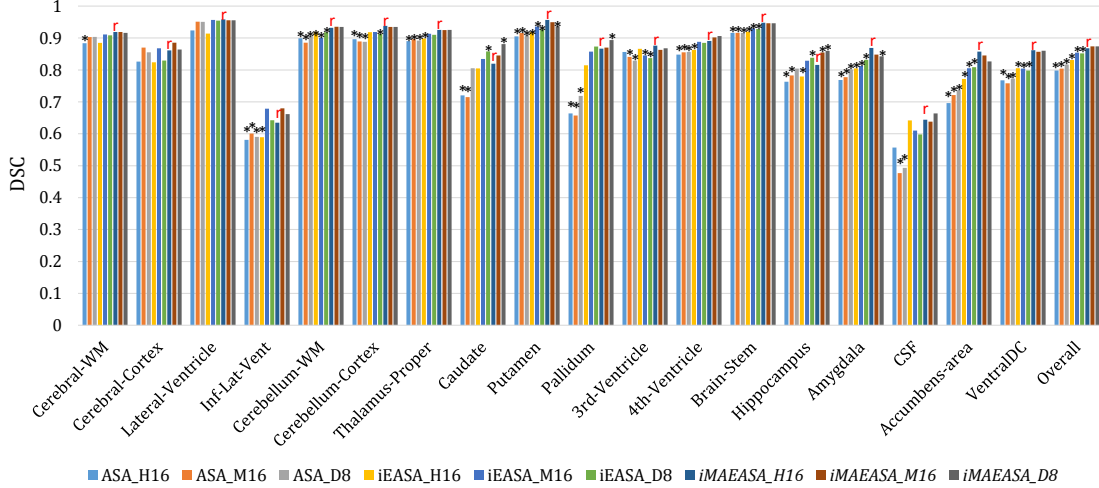
Figure 3.11: Quantitative comparison on SevereVent_AD dataset using ASA, iEASA and *iMAEASA* methods with different atlases. Bar annotation legend: **r** indicates the reference (our proposed method) while ***** indicates those methods with a statically significant difference ($p < 0.05$) from the reference.

training data contains many diseased brains. In summary, *iMAEASA* is better than other two methods in terms of the accuracy as well as the robustness to the training data.

3.6 Discussions

3.6.1 Relationship with Patch-based Label Fusion

Aspects of recent patch-based approaches may be useful to improve our approach. For example, when constructing our relaxation map, we could derive a refined estimate of label uncertainty by weighting the votes of the posterior before computing its entropy. This is analogous to the way joint label fusion is computed [100] using an image similarity measurement after non-linear registration. Similarly, it is possible that aspects of our method may be useful to improve the patch-based methods. For example as shown in Fig. 3.12, when the training data consists of healthy brains and the target is diseased, the distance from where 50% of the training atlases agree (green pixels) and the true boundary (of the target's ventricle) can be 18-20 voxels. This far exceeds the search volume radius for matches employed by patch-based methods (3-5voxels).
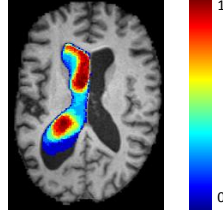
Figure 3.12: Map of the prior probability (p) for the right lateral ventricle of subject 2 in Fig. 3.6. Color bar shows the plotted color except for $p = 0$ which is rendered transparent for anatomical context.

Naively increasing the radius in a fixed global way yields lower performance for patch-based methods. However, our approach for measuring training suitability locally could be used to drive a *spatially varying* search volume radius for patch-based methods (i.e. increasing it *locally where* there is label uncertainty and increasing it *by how much* uncertainty there is between the multi-atlas training data and target).

3.6.2   Necessity of Components

Target-specific priors are necessary for maximum generalization of training data to parcellate an unseen target. This is illustrated in Fig. 3.13. The first column shows an axial image from a diseased target. The training data consists of healthy brains. If a probabilistic atlas prior is constructed using only the training data and without regard to the target, as is the case in iEASA, some speed up is afforded because the priors can be built offline. However, this comes at the price of reduced generalization. As shown in the third column, iEASA labels part of the ventricle as WM. On the other hand, *iMAEASA* delays atlas construction until the target arrives. It constructs priors only after non-linearly registering the healthy training data to the target. This allows for more accurate identification of the regions in which the priors need to be relaxed when parcellating the target. As shown in the fifth column *iMAEASA* properly adapts the priors from the same original healthy training data. When examining the relaxation maps ξ_i from the two approaches (last two columns), the bright region *iMAESA* has identified for further relaxation (in the yellow circle) is absent from ξ_i for iEASA.

While *iMAEASA* generates a more adaptive relaxation map because it is target specific, both structure-specific relaxation (section 3.3.3) and pairwise-structure specific
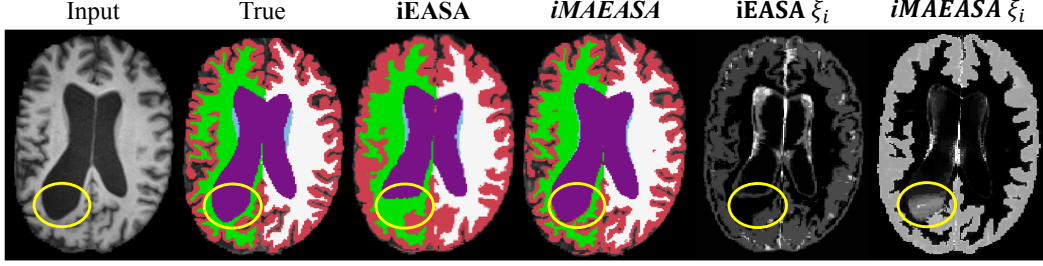
Figure 3.13: Advantage of target-specific atlas construction. When the atlas is constructed without regard to the target input, iEASA does not adapt sufficiently. The proposed *iMAEASA* produces better generalization and accurate parcellation result, because it constructs target-specific relaxation map ξ_i which identifies the region requiring additional adaptation (yellow circle).
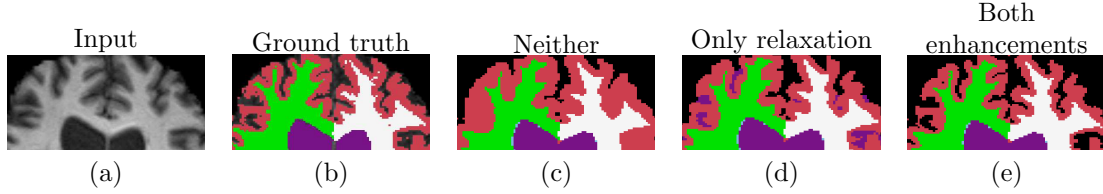
Figure 3.14: The need for structure-specific relaxation and pairwise-structure specific MRF priors. (a) shows the anterior region of mid-axial image, (b) is ground truth, (c) using neither, (d) only structure-specific relaxation, or (e) both enhancements.

MRF smoothness priors (section 3.3.3) further increase parcellation accuracy. This is demonstrated in Fig. 3.14. The example in (a) shows the anterior region of a mid-axial image and (b) is ground truth. Part (c) shows the result without structure-specific relaxation and MRF smoothness. Here we see the sulci are erroneously filled in. Part (d) shows the result using only structure-specific relaxation, and we observe erroneous ventricular CSF labels (purple) and over-segmented cortex. Part (e) shows the much-improved result using both structure-specific relaxation and pairwise-structure specific MRF smoothness priors.

3.6.3   Implementation Details

The primary parameters of *iMAEASA* control atlas construction. Using empirical data, we set these parameters once and kept them fixed for all our experiments. In particular, we set the parameters mapping entropy to the relaxation map, ξ_i, as: $b_1 = 0.0$, $b_2 = 0.5$, $s_1 = 3.0$, $s_2 = 0.1$, the structure-specific relaxation parameters as: $\alpha_1 = 0.9$, $\alpha_2 = 0.8$,

$\alpha_3 = 0.95$, and the parameters for MRF-based regularization as: $\gamma = 10^{-7}$, $b_3 = 0.0$, $b_4 = 1.0$. For non-linear SyN registration [129], we use a cross-correlation metric with window radius 4, weight 1 and gradient step length 0.25. The optimization is performed on 3 resolutions with a maximum of 70 iterations at the coarsest level, 50 at the next coarsest and 10 at the full resolution, after the affine registration (200, 100, 50 maximal iterations respectively). We use a Gaussian regularizer with sigma of 3 that operates on the similarity gradient and 0.5 on the deformation field. For the comparison methods, we used the implementation from the authors unless stated otherwise and optimized the parameters according to the author's recommendations in the articles, e.g. [139].

3.7 Summary

In the clinic, representative training data for the target population may not be available in many studies, and plenty of golden standard labeling is usually not realistic. In this work, we simulated one common situation in which diseased brains need to be segmented while there is only training data of healthy brains available. We propose a novel brain parcellation algorithm (*iMAEASA*) that integrates aspects from the multi-atlas genre with those from the adaptive probabilistic atlas (*ASA, EASA, iEASA*) genre. The training atlases are non-linearly registered to the target and a non-stationary relaxation map is automatically computed to tune the adaptivity of the multi-atlas. This is done on a voxel-by-voxel basis, according to the local suitability of the training data to the target. Then, all of the additional regularization components of its probability model are constructed in a target-specific fashion and then the complete model is further evolved to fit the target using EM optimization. Our approach simultaneously and accurately segments over 30 individual anatomical structures throughout the brain. For healthy brains, our method reveals comparable accuracy to the state-of-the-art parcellation approaches. For diseases with brain shape deformation such as Alzheimer's (with moderate to severe structural malformations), our method yields higher parcellation accuracy in terms of dice score than other leading methods. Other important diseases such as Huntington's, Parkinson's, Schizophrenia, and ventriculomegaly also have morphological shape changes and may be well served by our approach.

Chapter 4

Combining Atlas with Deformable Model

4.1 Introduction

In the last chapter, we discuss the adaptive atlas based multi-object segmentation algorithm. In this chapter, we propose a single-object segmentation method by integrating atlas-based prior information in an improved deformable model, and evaluate it with an application to liver segmentation in hepatic adipose assessment. The proposed adaptive atlas based multi-object segmentation method in the last chapter does not suit in this application, because there is no ground truth for each anatomy except the liver and so the background contains too complicated intensity distribution to be modeled by a simple Gaussian.

Fatty liver, also known as hepatic steatosis, is a worldwide common condition characterized by fat accumulation in liver cells. If more than 5% to 10% of the liver's total weight is fat, it is called a fatty liver. It is associated with many fatty liver diseases (FLD), like alcoholic liver disease, non-alcoholic fatty liver disease (NAFLD) and steato-hepatitis [156, 157]. NAFLD may affect people of any age and progress to end-stage liver diseases [158]. The prevalence of NAFLD may be as high as 30% in the United States [159] and 10% to 24% of general population in various countries [158]. Therefore, the clinical diagnosis of fatty liver disease is very important. The diagnosis procedure often requires imaging studies. It has been shown that MRI and magnetic resonance spectroscopy (MRS) [160] are better choices for accurate detection and quantification of the hepatic fat than ultrasonography and CT. Particularly, MRS is regarded as the most direct MR-based method to quantify water and fat components in liver. However, it is not widely applicable across standard clinical imaging centers due to the technical complexity. Instead, MRI based hepatic fat-fraction measurement is widely used. It

employs multi-echo chemical shift based methods and computes the fat-fraction image based on the separation of fat and water [160]. So far, most studies focus on improving the separation and quantification of fat and water components, while the estimated fat-fraction images are for the whole abdomen, not liver specific. Thus, automatic liver segmentation is necessary in generating liver-specific fat-fraction assessment for large data analysis in clinical trials.

However, there are some challenges in liver segmentation. First, image noises and artifacts can cause ambiguous object boundaries, which may affect edge-detection based methods. Second, nearby organs with similar intensity levels may cause misleading boundaries. Third, the liver morphology and structure may also change due to liver affecting disorders and diseases. Again, in clinical trials, the collected data have large variance due to different patients, hardware and protocols. Fig. 4.1 shows several T1-weighted MRI samples in a FLD clinical trial. They have significantly different appearances.

Many studies have shown different fully automatic liver segmentation methods in recent years. Most of these segment liver in CT images [161]. Ling et al. [16] proposed a hierarchical shape representation and learning based boundary localization technique. Kainmüller et al. [162] used a statistical shape constrained free form deformable segmentation method. Rikxoort et al. [128] and Linguraru et al. [163] incorporated atlas-based method into their segmentation for contrast enhanced CT images. However, CT images, especially those with contrast enhancement, have different appearance patterns with MR images which makes most algorithms designed for CT images not suitable for MRI. For liver segmentation in MR images, Massoptier et al. [164] used graph cut initialized from several carefully designed preprocessing steps. Platero et.al. [165] applied level set [114] to segment liver after manual selection of ROI. Logeswaran et.al. [166] proposed a watershed [167] based algorithm to segment liver in 2D MR images, which may loss 3D spatial constraints. Siewert et.al. [168] proposed an automated method for contrast enhanced MRI based on edge detection in 2D slices. Considering the large variety of data set in clinical trials, improving the robustness of automatic 3D segmentation is still under research.
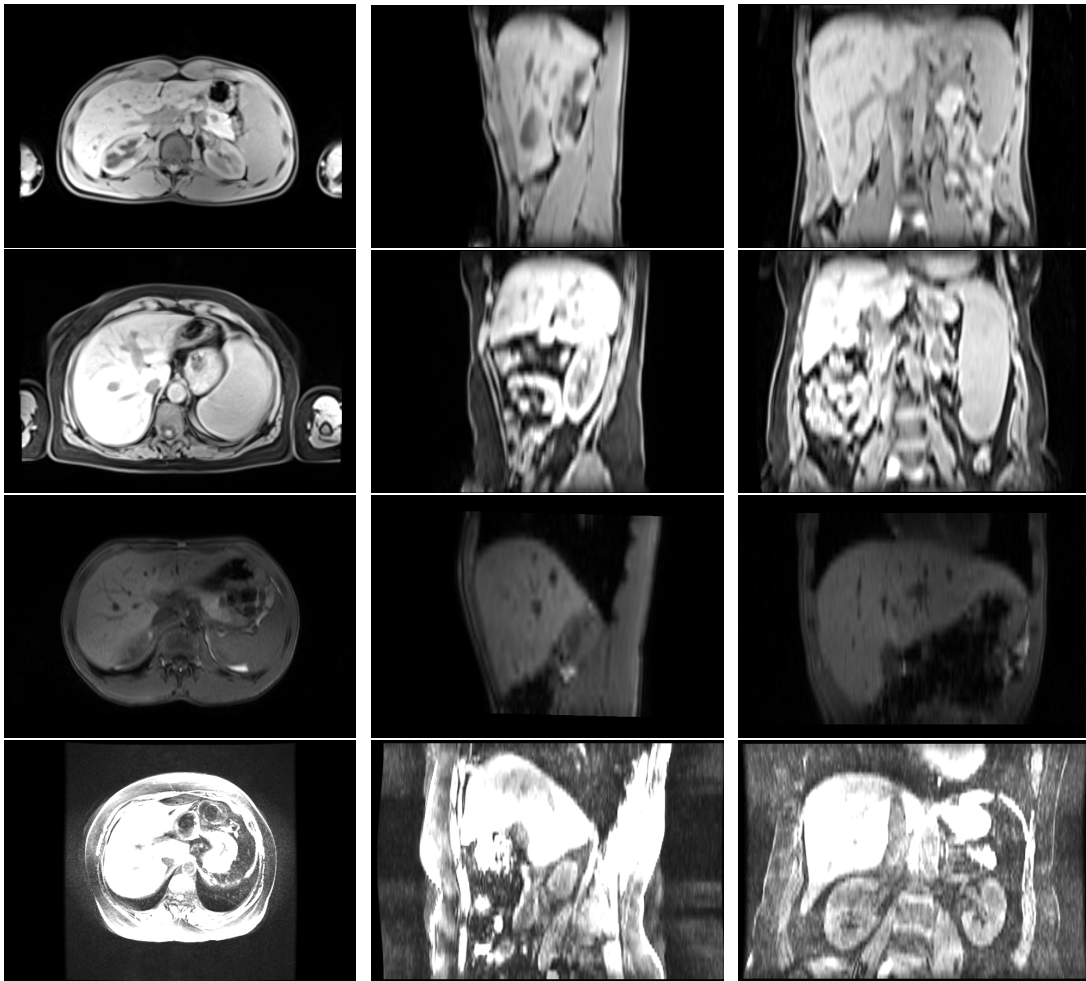
Figure 4.1: T1-weighted MR images in a FLD clinical trial. Each column in one row shows one subject in axial, sagittal and coronal views, respectively.

Owing to the variety of data in clinical trials and limited amount of training samples, most learning based segmentations [16, 116, 117, 118, 119] are hardly applied. Deformable model based methods [7, 110] are accurate but sensitive to initialization and misleading appearance clues. Huang et al. proposed a Metamorphs model [115] which combines the edge and region energy terms to make the deformable model robust to initialization. Some other studies use the deformable models with shape priors, e.g. Active Shape Model (ASM) [169] and sparse shape composite [122], are shown promising in cardiac segmentation [116], brain structure segmentation [170] as well as lung segmentation [171]. Atlas-guided approaches [98, 90, 120, 121, 104] relying on image registration may not work as well as in the brain since the abdominal section has much higher variety than the brain. Because the image registration algorithms are solved to obtain a global matching of target and reference images, there is no guarantee that the liver area is aligned properly. Despite the limitation, Slagmolen et al. [172] and Okada et al. [173] demonstrated the atlas-based method can provide a reliable initial segmentation for refinement procedure.

Based on these previous studies, we propose an robust and automated liver segmentation in clinical environment by incorporating the atlas based prior information with a shape constrained deformable model. Specifically, a statistical image atlas is constructed and employed to obtain a rough estimation of liver ROI. Then, a robust deformable model with shape prior is initialized from this estimation. Energy terms from edge and region information as well as the atlas prior are combined in this model for accurate and robust segmentation. Besides, Laplacian mesh representation [174, 175] is used to preserve surface details. The proposed segmentation method combines the robust of atlas-based approach with the accuracy of deformable models.

Moreover, the proposed framework could be easily extended to refine segmentation results interactively. From our observations, the fully automatic framework is not perfect in terms of segmentation accuracy at some local regions. In practice, clinical experts would like to have an intuitive way to correct the local mistakes efficiently. This motivates us to develop an interactive mesh editing tool to refine the segmentation results. Interactive segmentation strategies usually take different types of user

input to initialize the "editing" process, and then generate intermediate segmentation results based on automated algorithms. Users can provide or update inputs to correct the segmentation mistakes iteratively. For example, interactive graph cut [176] utilizes a user's input of foreground and background seeds to segment the foreground objects, based on the intensity statistics and spatial connectivity. The additional user inputs may affect the global statistics and cause some other places to be mislabeled. Active learning based interactive segmentation [177] allows a user to delineate a set of 2D contours of the 3D object in an active learning framework and reconstruct a 3D surface based on the labeled 2D contours. In general, these existing interactive approaches apply user inputs implicitly to the segmentation results.

Different from previous approaches, we extend our deformable model to enable 3D shape editing directly. In fact, 3D shape editing is a classical topic in computer graphics and is used for animation and visual reality [174, 175]. However, these approaches only rely on the geometric information. For medical image segmentation, we also need to incorporate the image context into the interactive mesh editing. Therefore, the parametric deformable model is a natural choice, since they represent the object boundary as a smooth surface model by simplex mesh. Therefore, we extend our deformable model and design a novel interactive framework to refine the segmentation results by mesh editing. This framework models the mesh editing process by user-provided control points as well as the image context. The control points not only specify the locations where the segmentation errors should be corrected, but also provide reliable landmarks where the local image information should be emphasized to extract more accurate object boundaries. With the aid of geometric optimization and image guidance, the proposed framework allows the doctors or researchers to explicitly edit the segmented shapes in an intuitive fashion. The refined results could be utilized to do clinical analysis with high confidence.

We first evaluate this automatic method on a dataset of 14 volumetric MR images from a FLD clinical trial. Then, we evaluate the editing framework on the same MRI liver data set and another CT lung data set.

4.2   Methodology

In this section, we first introduce our framework for liver segmentation and hepatic fat-fraction assessment. In the second part, the energy terms contributing in the deformable model are discussed in detail. Then we describe the construction of statistical image atlas and shape atlas, and the sparse shape prior representation.

4.2.1   Algorithm Framework

The demonstration of our framework is shown in Fig. 4.2. Firstly, through atlas construction, we build a statistical image atlas containing one reference intensity image and a corresponding spatial probability map (SPM) image for liver. Meanwhile, one shape atlas is constructed for shape representation model. The SPM and shape model are in the same reference image domain. After nonrigidly registering the reference image to target image, we deform the SPM and shape model to the target image space using the same transformation. Then, the deformable model is initialized by the SPM and evolves to fit the boundary of the target liver. The deformation is based on the image edge and region information and regulated by the warped SPM as well as the shape constraint. Finally, a chemical shift method can be applied to assess fat-fraction distribution in the liver from dual-phase MR images of the same target patient.

The key component of this framework is the robust deformable model, which is detailed in section 4.2.2. The atlas construction and shape regularization are discussed in section 4.2.3.

4.2.2   Deformable Model

There are two main categories of deformable model in literatures. One is explicit parametric models that use parametric representation for deformable curves or surfaces during the segmentation, such as Snakes or Active Contour Models [111, 113]. The other one is implicit geometric models [112, 114] that use the level set of a higher-dimensional scalar function to represent curves or surfaces implicitly. Although differ in implementation, both model the segmentation as an energy minimization problem
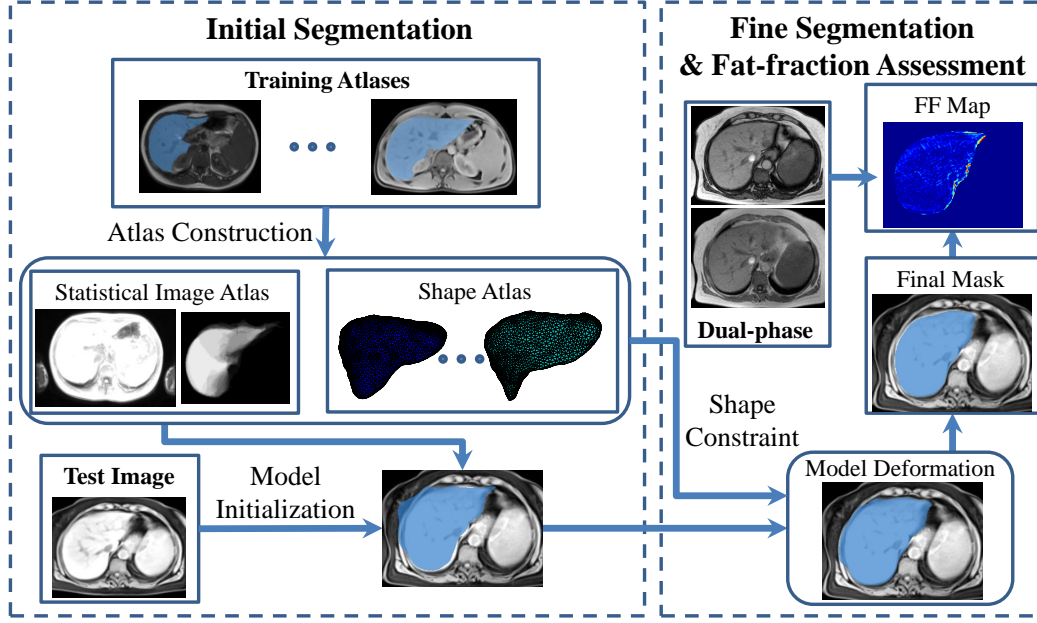
Figure 4.2: Segmentation Framework.

and rely on edge information to derive external forces. Here, we choose the explicit model because it is more computationally efficient than the implicit model.

Let $\mathbf{x} : \Lambda \subset R^2 \rightarrow R^3$ denote a surface representing object boundary, $\mathcal{I} \in R^3$ denote a 3D image. The surface model $\mathbf{x}$ can be discretized as a simplex triangle mesh represented by a graph $G = (V, E)$ with vertex V connected by edges E. $V = [v_1^T, v_2^T, \ldots, v_n^T]^T$, $v_i = [v_{ix}, v_{iy}, v_{iz}]^T \in R^3$. Let $G_0 = (V_0, E_0)$, $G_t = (V_t, E_t)$ denote the initial and target shape, respectively. We aim to evolve G_0 to $G_d = (V_d, E_d)$ which can best approximate G_t. The segmentation is formulated as an energy minimization problem:

$$E_{total} = E_{model}(G_0, G_d) + E_{shape}(G_0, G_d) + E_{target}(G_d, G_t). \tag{4.1}$$

Here, model energy E_{model} measures the similarity between G_0 and G_d. It preserves the initial geometric characteristics during the deformation of G_d. Shape energy E_{shape} is designed for shape optimization. It constrains the shape's tension and rigidity to ensure the global smoothness, while preserving local details as much as possible. Target energy E_{target} reflects the distance between G_d and G_t. It introduces an external force driving the shape model evolving towards the target shape G_t. In image segmentation, the G_t represents the boundary of the object of interest. By minimizing these three energies,

the model will deform to the target boundary while preserving its geometric properties. At the meantime, it avoids mesh degeneration during the deformation and produces an optimized one.

All the energy terms above are designed differentiable with respect to the model parameter (coordinates V). Thus, according to the Euler-Lagrange Theory the mesh deformation can be solved as a dynamic system using gradient descent algorithm. E_{model} and E_{shape} derive internal forces to preserve and optimize the shape properties. E_{target} derive external (image) forces to attract the model towards the target object boundary. The minimization is computed in small time steps. In our study, we keep the same topology of the model and only change the coordinates of V during the deformation. Now, we discuss all the energy terms in detail.

Internal Forces with Detail-preserved Smoothing

By representing as finite differences in time when G_0 deforms to G_d, the model energy E_{model} is defined as:

$$E_{model} = \gamma \|V^{(t)} - V^{(t-1)}\|^2 \tag{4.2}$$

The shape energy E_{shape} which controls the smoothness of deformable model is defined as:

$$E_{shape} = \alpha \|L_u V_p^{(t)} - W L_c V_p^{(t-1)}\|^2 + \beta \|L_u^2 V_p^{(t)}\|^2 \tag{4.3}$$

where $V^{(t)}$ are the vertices of the intermediate shape at time t, $V_p = [v_{1p}, v_{2p}, \ldots, v_{np}]^T$, $p \in \{x, y, z\}$ are coordinates along p direction. $L_u V_p$ and $L_c V_p$ are uniform and cotangent Laplacian used to approximate the discretization of Laplace operator to V [175, 178]. The uniform Laplacian of a vertex is a vector pointing to the centroid of its neighboring vertices, while cotangent Laplacian is a good approximation of the surface normal. The first term of E_{shape} controls the tension (for smoothing) by removing tangential components while keeping the surface details in normal direction. The diagonal weight matrix W for the cotangent Laplacian can be assigned as in [178] to segment objects with complex boundary shapes. The second term controls the rigidity of the shape (avoids oscillations by penalizing high curvatures). α and β are the balancing

weights.

External Forces with Atlas-based Potential

The external forces are derived from E_{Target}:

$$E_{Target} = \kappa_1 \int_\Lambda P_{Ext}(\mathbf{x})d\Lambda = \kappa_1 \int_\Lambda \mathcal{P}_{edge} + \kappa_2 \mathcal{P}_{region} + \kappa_3 \mathcal{P}_{atlas} d\Lambda \qquad (4.4)$$

Here, $P_{Ext} : R^3 \to R$ is the potential field induced by three parts which are detailed in the following paragraphs. κ_1, κ_2 and κ_3 are all scalar parameters used as balances of these potential terms.

The edge potential term $\mathcal{P}_{edge}$ can be defined using gradient magnitude map or edge distance map or GVF [111, 113]. The region potential term $\mathcal{P}_{region}$ is defined similar to the ROI-based balloon term in [115] that allows flexible model initialization:

$$\mathcal{P}_{region}(\mathbf{x}) = \Phi_\Lambda(\mathbf{x})\Phi_{ROI}(\mathbf{x}) \qquad (4.5)$$

Here, $\Phi_\Lambda(\mathbf{x})$ is the signed distance transform of the current model's surface, $\Phi_{ROI}(\mathbf{x})$ is the signed distance transform of the predicted liver boundary. Given the transformed SPM, both foreground (object) and background statistics are estimated. By applying Bayesian rule, we can compute a binary image representing the most likely object regions. Basically, this term computes current model-interior intensity statistics to predict objet boundary, which can overcome some misleading boundaries from the edge detection. Here, we only compute it once based on SPM before the model evolution.

Atlas-based energy. To better suppress possible side effects caused by image artifacts and noises, which may down-grade the edge- and region-based energy potentials, we introduce the third potential term $\mathcal{P}_{atlas}$ into the external force field. This potential function comes directly from the transformed SPM learnt from atlas construction:

$$\mathcal{P}_{atlas}(\mathbf{x}) = \Phi_{SPM}(\mathbf{x})^2 \qquad (4.6)$$

where $\Phi_{SPM}(\mathbf{x})$ is a signed distance map to the boundaries of $SPM = \theta$ ($\theta \in (0,1)$ is determined empirically). This term keeps the model from shrinking too small to nothing and dilating too big to leak to background. The motivation of such a term is

from the observations that the transformed SPM always overlaps nicely with the major part of the target liver even the nonrigid registration is not that accurate.

4.2.3 Construction of Image and Shape Atlases

In the atlas construction, we build one statistical image atlas and one shape atlas. The statistical image atlas contains one mean intensity image representing a group of training images, and one probabilistic liver map encoding the spatial distribution and variance of liver in the training samples. This statistical image atlas is used as initialization as well as a constraining energy term of the deformable model. The shape atlas is a set of surface meshes built from training data which is used to represent the shape prior of liver.

Image Atlas

Park et al. [179] built a liver probabilistic atlas based on landmarks. Observing the atlas construction is sensitive to the choice and accurate localization of landmarks, Xiong et al. [180] used a landmark-free method based on dense volumes to construct a linear unbiased liver atlas from CT images. Here we propose a strategy to build the probabilistic liver atlas from MR images similar to [180] by using a state-of-the-art symmetric diffeomorphic normalization (SyN) method [129]. SyN is shown to be one of the top-ranking methods with the best and most consistent accuracy in MRI brain registrations [101]. Let $\omega\left(\mathcal{I}, \mathcal{D}(\mathbf{x}), \mathcal{J}\right) \rightarrow \hat{\mathcal{I}}$ denote a nonrigid transformation ω from source image $\mathcal{I}$ to target image $\mathcal{J}$ by deformation field $\mathcal{D}(\mathbf{x})$, resulting in a warped image $\hat{\mathcal{I}}$ in the image space of $\mathcal{J}$. SyN method solves the registration problem symmetrically to get the deformation field $\mathcal{D}(\mathbf{x})$ and its inverse $\mathcal{D}^{-1}(\mathbf{z})$ at the same time, where $\mathbf{x}, \mathbf{z}$ are spatial coordinates in images $\mathcal{I}$ and $\mathcal{J}$.

Given M training simple atlases, our statistical atlas is built following the algorithm 2. The $\mathcal{I}_i$ and $\mathcal{S}_i$ denote the ith image and its golden standard label image, respectively. As the output, $\bar{\mathcal{I}}$ is the mean image and $\bar{\mathcal{S}}$ is the liver probabilistic map. $\bar{\mathcal{I}}$ is the unbiased representative of the training images. $\bar{\mathcal{S}}(\mathbf{x})$ represents the probability of liver tissue appearing at $\mathbf{x}$. Note that we do not preprocess the training dataset to normalize

them into a common size and spacing. Instead, the normalization is implicitly done after the first iteration since all the images are transformed to the normalized template $\bar{\mathcal{I}}^{(0)}$. The result of our atlas construction is shown in Fig. 4.3.

Algorithm 2 Build statistical atlas

Input:
  The training simple atlases $\{\mathcal{I}_i, \mathcal{S}_i\}, i \in \{1, \ldots, M\}$;
Output:
  The statistical atlas $\{\bar{\mathcal{I}}, \bar{\mathcal{S}}\}$;
 1: $k = 0$
 2: pick an arbitrary m, where $m \in \{1, \ldots, M\}$
 3: re-sample $\mathcal{I}_m$ to get isotropic voxel size $1mm \times 1mm \times 1mm$
 4: initialize template $\bar{\mathcal{I}}^{(0)}$ by normalized $\mathcal{I}_m$
 5: $\mathcal{I}_i^{(0)} = \mathcal{I}_i$, $\mathcal{S}_i^{(0)} = \mathcal{S}_i$
 6: **while** $k < MAX\_ITERATION$ **do**
 7:   **for** $i = 1$ to M **do**
 8:     compute $\mathcal{D}_i^{(k)}$ to get warped image: $\omega\left(\mathcal{I}_i^{(k)}, \mathcal{D}_i^{(k)}, \bar{\mathcal{I}}^{(k)}\right) \rightarrow \mathcal{I}_i^{(k+1)}$
 9:   **end for**
10:   update template: $\bar{\mathcal{I}}^{(k+1)} = Average_{i=1}^M \mathcal{I}_i^{(k+1)}$ [1]
11:   $k = k + 1$
12:   if $\bar{\mathcal{I}}^{(k+1)}$ converge, break
13: **end while**
14: **for** $i = 1$ to M **do**
15:   compute $\mathcal{D}_i^{(k)}$: $\omega\left(\mathcal{I}_i^{(k)}, \mathcal{D}_i^{(k)}, \bar{\mathcal{I}}^{(k)}\right) \rightarrow \hat{\mathcal{I}}_i$
16:   warp label image: $\omega\left(\mathcal{S}_i^{(k)}, \mathcal{D}_i^{(k)}, \bar{\mathcal{I}}^{(k)}\right) \rightarrow \hat{\mathcal{S}}_i$
17: **end for**
18: $\bar{\mathcal{I}} = Average_{i=1}^M \hat{\mathcal{I}}_i$
19: $\bar{\mathcal{S}} = \frac{1}{M} \sum_{i=1}^M \hat{\mathcal{S}}_i$

During the construction, each nonrigid (deformable) registration is performed after an affine registration. We use mutual information (MI) similarity metric in affine registration, and cross-correlation (CC) as the similarity metric for nonrigid registration.

Shape Atlas

Traditional shape models represent shape distributions by the mean and major variations, for example Active Shape Models (ASM) [169]. These statistical shape models do not handle gross and sparse errors well. In many cases, the initial segmentation is

---

[1]The *Average* of images are calculated using function *AverageImages* in Advanced Normalization Tools (ANTs) software [181].
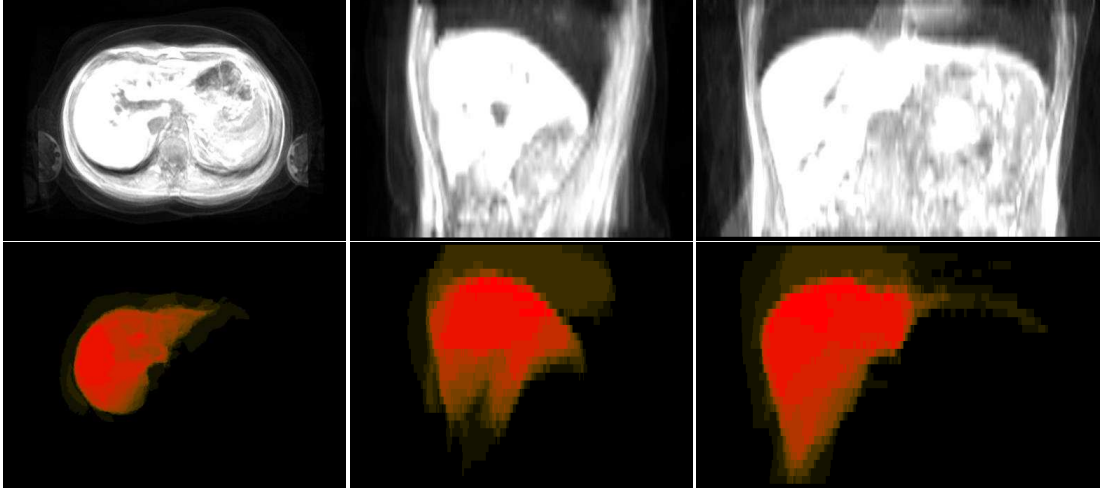
Figure 4.3: The mean image and liver probabilistic atlas are shown in axial, sagittal and coronal views. The brightness of red color is proportional to the probability of a voxel belonging to liver tissue.

relatively good and deformable model is reliable at most positions, only sparse gross errors instead of Gaussian errors may be observed during the model evolution. Thus, we incorporate a sparse shape composite model [122, 171] into our method for more accurate and effective results.

Having the statistical image atlas, a binary template label image can be estimated by thresholding $\bar{\mathcal{S}}$ after smoothing and morphological operations. Then a high-quality surface mesh [182] is built in the template domain. Similarly, all the training label images are converted to high-quality mesh models. To build the shape prior model, we first pre-aligned all the training shapes to the reference shape based on the generalized Procrustes analysis [183] to make sure they are all in the same coordinate space. Then the one-to-one correspondence of vertices on different meshes is obtained by registering the reference mesh to all the training shapes. Here, we employ the mesh quality-preserved deformable models to create the one-to-one correspondence [182]. This method is adapted from the adaptive focus deformable model (AFDM) [184] by registering shapes instead of images.

The sparse shape composite model used to incorporate with deformable model is defined as:

$$\operatorname*{argmin}_{\mathbf{c},\mathbf{e}} \|T(\mathbf{y}) - A\mathbf{c} - \mathbf{e}\|_2^2 + \lambda_1\|\mathbf{c}\|_1 + \lambda_2\|\mathbf{e}\|_1 \qquad (4.7)$$

where $\mathbf{y} \in R^{DN}$ is the vector representing all the vertices V_d in the deforming mesh, D is the degree of freedom, N is the number of vertexes in the mesh model, $T(\mathbf{y})$ is the transformation (i.e. Procrustes transformation) from target domain to training template domain, $\mathbf{e} \in R^{DN}$ is the sparse error vector, $A \in R^{DN \times M}$ is the matrix of training shapes, $\mathbf{c} \in R^M$ is the sparse coefficient vector. The optimized shape representation is $A\mathbf{c} - \mathbf{e}$, which can be transformed back to input space with the inverse transformation of T.

4.2.4 Implementation Details

In the construction of statistical image atlas, the nonrigid registration in each iteration requires about $M \times 40$ minutes. In [180], the authors showed that the convergence for a 15-sized dataset start after 3 iterations. Considering the high cost of computation, we set $MAX\_ITERATION = 3$ in algorithm 2.

The initialization of deformable model is based on one nonrigid registration from atlas mean image to target image. When a unseen target image comes, we compute the deformation field D_T from $\bar{\mathcal{I}}$ to target $\mathcal{I}_T$. Then target specific liver SPM is estimated by the warped probabilistic map $\omega\left(\bar{\mathcal{S}}, D_T, \mathcal{I}_T\right) \to \bar{\mathcal{S}}_T$. An initial liver segmentation $\mathcal{S}_{\text{Init}}$ is computed by thresholding $\mathcal{G}_\sigma * \bar{\mathcal{S}}_T$ with a scalar value θ_T, where $\mathcal{G}_\sigma$ is a Gaussian smoothing filter with scale size σ. This target specific $\mathcal{S}_{\text{Init}}$ is used to build the initial surface mesh for deformable model. The vertex correspondence between the target mesh and the reference shape is computed using the shape registration method in [182].

To improve the computational speed of the deformation D_T, an online atlas-to-target registration method proposed by [155] is used. This method is based on cubic B-spline interpolation and NMI similarity metric with GPU acceleration. It is much faster than the SyN method at the cost of losing some accuracy. Although the simple registration method produces less accurate transformation, the initialization is refined by deformable model afterwards.

To summarize, the evolution of deformable model with shape constraint is described in algorithm 3. After registering and warping the image atlas to the target domain, we construct the initial surface model by extracting an isosurface from the warped SPM.

Then the energy terms and force fields are computed for the deformable model before its evolving. Each energy term is described in detail in section 4.2.2. To keep up the efficiency of the evolution, the shape regularization is conducted every τ iterations. The evolution is stopped when the deformation is small enough. At last, the 3D surface model is converted to binary mask image as the segmentation result.

Algorithm 3 Segmentation framework

Input:
    The target image, trained statistical image atlas and shape atlas
Output:
    The binary mask of segmented liver in target image
 1: compute initial surface model $\mathbf{y}^{(0)}$ based on image atlas registration
 2: compute internal and external forces
 3: $t = 0$
 4: **repeat**
 5:     $t = t + 1$
 6:     deform model $\mathbf{y}^{(t-1)}$ to $\mathbf{y}^{(t)}$ using FEM
 7:     **if** $mod(t, \tau) == 0$ **then**
 8:         optimize $\mathbf{y}^{(t)}$ using Eq. 4.7 to get $\mathbf{y}^{(t)}_{\text{refined}}$
 9:         $\mathbf{y}^{(t)} = \mathbf{y}^{(t)}_{\text{refined}}$
10:     **end if**
11: **until** deformable model converges
12: convert the surface model $\mathbf{y}^{(t)}$ to mask image

---

4.2.5   Intuitive Segmentation Refinement

We keep the same notation for the deformable model as in section 4.2.2. Given a deformable mesh either from an initialization or a segmentation result with mistakes, we can conduct the interactive mesh editing (segmentation refinement) by setting control points in the local area. The control points introduce constraints into internal force and provide supplemental external force to ensure the correct estimate of the target model.

Internal Forces with Control-point Constraint

Let C be the set of control points (landmark points) defined by user. $c_j, j \in C$ denotes the coordinates of the points. $v_k, k \in V$ denotes the coordinates of those points selected

to respond the control of c_j. The shape energy E_{shape} is redefined as:

$$E_{shape} = \alpha\|L_u V_p^{(t)} - WL_c V_p^{(t-1)}\|^2 + \beta\|L_u^2 V_p^{(t)}\|^2 + \sum_{j,k} \omega_{kj}^2 |v_{kp}^{(t)} - c_{jp}|^2 \qquad (4.8)$$

The first two terms are the same as in Eq. 4.3, while the third term is the additional soft geometric constraint introduced by control points. ω_{kj} encodes the strength of the attraction from c_j to v_k regulated by the distances

$$\omega_{kj} = \frac{1 - \hat{dist}_{kj}}{\sum_k 1 - \hat{dist}_{kj}}, \qquad (4.9)$$

where $\hat{dist}_{kj} \in [0, 1]$ is the normalized distances from v_k to control point c_j.

External Forces with Control-point Enhancement

From Eq. 4.4 (may not have the atlas-based potential in practice), we can derive the image force F_{Ext}. However, due to ambiguous edges and low contrast in medical images, the traditional image forces may lead to inaccurate segmentation result. The intuitive goal of interactive aids is to emphasize the desired object boundaries to overcome misleading appearance cues in local image regions. The internal constraints (in Eq. 4.8) can pull vertices towards the desired position geometrically, but limited number of control points may cause sharp spines (Fig. 4.4 (b)). To overcome such limitation and improve the efficiency of user inputs, we utilize the local image information indicated by control points to form a new external force field F_C as a local enhancement to F_{Ext}:

$$\tilde{F}_{Ext} = F_{Ext} + \lambda F_C \qquad (4.10)$$

where λ is the weight of the force field generated based on control points. Let I_s be local image patch with control point as the center and window size s, we define F_C as:

$$F_C = \boldsymbol{K}(x, y, z) * |\nabla[G_\sigma * I_s]|^2 \qquad (4.11a)$$

$$\boldsymbol{K}(x, y, z) = [-x/(r+\epsilon)^3, -y/(r+\epsilon)^3, -z/(r+\epsilon)^3] \qquad (4.11b)$$

Here, G_σ is a 3D gaussian function with standard deviation σ. $\boldsymbol{K}(x, y, z)$ is a 3D Vector Field Convolution (VFC) kernel [185] in which all the vectors point to the kernel origin ($r = \sqrt{x^2 + y^2 + z^2}$ is the distance to the kernel origin). Comparing with other force

(a)   (b)   (c)   (d)
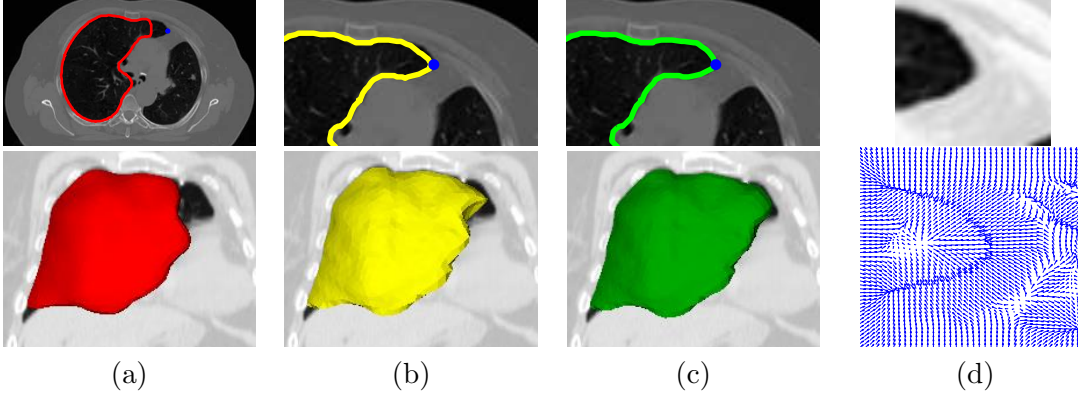
Figure 4.4: Example of interactive CT lung segmentation. (a) shows the initial segmentation (red) and user specified control point (blue); (b) shows the editing result (yellow) with only geometric constraint by control point; (c) shows the editing result (green) with both geometric constraint and image force F_C; (d) shows an axial image patch and the force field F_C.

field (e.g. gradient vector field), VFC has larger capture ranges, is more robust and computationally efficient. The radius of the kernel $\boldsymbol{K}$ is set to $s/4$.

Combining all the energy terms via shape optimization, image context and control points, the matrix representation of the overall minimization problem is formed as a very sparse and over-determined linear system:

$$\begin{bmatrix} A \\ diag(\sum_j \omega_{kj})I_k \end{bmatrix} V_p^{(t)} = \begin{bmatrix} \hat{f}_p^{(t-1)} \\ \Omega C_p \end{bmatrix} \tag{4.12}$$

$$A = \gamma I - \alpha L_u + \beta L_u^2 \tag{4.13}$$

$$\hat{f}_p^{(t-1)} = \gamma V_p^{(t-1)} + \kappa \tilde{F}_{Ext}(V_p^{(t-1)}) - \alpha W L_c V_p^{(t-1)} \tag{4.14}$$

Ω is a matrix formed by putting weight ω_{kj} at position (k,j); each row in I_k is the indicator vector e_k^T. $\tilde{F}_{Ext}(V_p^{(t-1)})$ is the column vector representing p-axis force values at vertices $V^{(t-1)}$ by trilinear interpolation. γ, α, β and κ are the balancing weights of each force. This linear system can be solved efficiently in a least squares fashion.

4.3   Experiments

In this section, we first evaluate the fully automatic liver segmentation, then conduct experiments for interactive segmentation refinement. At last, we discuss some issues

and limitations observed in the experiments. All the methods were implemented in Matlab, tested on 3.4 GHz Intel Core i7 computer with 8G RAM and 1GB GPU.

The quantitative and statistical comparisons are conducted by leave-one-out cross validation. Each time we choose one sample as the testing and the remaining as the training set. We report the mean and standard deviation of overlapping score DSC (defined in Eq. 3.30), accuracy (ACC) and relative error (RE) referred to the ground truth. Let TP, TN, FP and FN denote the number of voxels correctly identified, correctly rejected, incorrectly identified, and incorrectly rejected as liver tissue, respectively. The DSC can be reformulated as $DSC = \frac{2TP}{2TP+FP+FN}$. The ACC and RE are defined as the following:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.15}$$

$$RE = \frac{FP + FN}{TP + FN} \tag{4.16}$$

We also measure the symmetrical surface distance error and Hausdorff distance [186] between the surfaces of segmentation results and those of the ground truth. Let X and Y denote point sets of ground truth surface mesh and segmentation result mesh, respectively; p_X be an arbitrary point in X and p_Y be an arbitrary point in Y. Eq. 4.17 represents p_X's closest point(s) in Y. On the contrary, Eq. 4.18 represents the point set in Y whose closest point in X is p_X. The symmetrical surface distance error at p_X is defined as Eq. 4.19.

$$\mathop{S}_{X \to Y}(p_X) = \{p_Y | \|p_X - p_Y\|_2 \le \|p_X - p_Y'\|_2, \forall p_Y, p_Y' \in Y\} \tag{4.17}$$

$$\mathop{S'}_{Y \to X}(p_X) = \{p_Y | \mathop{S}_{Y \to X}(p_Y) = p_X, \forall p_Y \in Y\} \tag{4.18}$$

$$Err_{surf}(p_X) = \max_{p_Y \in \{\mathop{S}_{X \to Y}(p_X), \mathop{S'}_{Y \to X}(p_X)\}} (\|p_X - p_Y\|_2) \tag{4.19}$$

The forward Hausdorff distance from X to Y is defined in Eq. 4.20, and we report the symmetrical Hausdorff distance as $max\{d(X,Y), d(Y,X)\}$.

$$d(X,Y) = \max_{p_X \in X} \min_{p_Y \in Y} \|p_X - p_Y\|_2 \tag{4.20}$$

4.3.1 Automatic MRI Liver Segmentation

An anonymized dataset of abdominal MR volumetric scans from 14 subjects is used for evaluation.[2] Each subject has one pair of OP/IP scans taken in a single breath-hold, and one T1-weighted MR scan acquired separately. We choose the T1-weighted MR images to segment liver, whose ground truths were manually labeled by experienced experts. Because the T1-weighted MRI may locate in different coordinate space from that of the corresponding OP/IP scans, we compute a rigid-only registration from T1 MRI to OP or IP MRI and apply the same transformation on the label image. Then hepatic fat-fraction distribution is assessed using IP and OP images with liver mask. Here, we only focus on the evaluation of liver segmentation from T1 MRI. In this dataset, image resolution ranges from 0.78 to 1.87mm in the axial slices with slice thickness from 3.5 to 7mm. The number of slices is from 30 to 104. The range in Z axis varies from 210mm to 364mm. The intensity level varies from 178 to 32767. Some of the samples are shown in Fig. 4.1.

We compare the proposed method with our implementations of two other methods. One is the atlas based registration method [172] (denoted as ATLAS), the other one is automatic graph cut method based on probabilistic atlas [187] (denoted as AGC). To make a fair comparison, all three methods use the same image atlas constructed in section 4.2.3. The atlas-to-target nonrigid registration is also same. The graph cut is based on the min-cut/max-flow algorithm [188] assuming 6-neighborhood connectivity.

Before comparing with different approaches, we first show the effect of the deformable model and the shape constraint in our proposed method. As shown in Fig. 4.5, red lines are ground truth boundaries, blue lines are the initial segmentations, and the green lines are the final results from our proposed approach. We observe that despite its inaccuracy, atlas-based registration provides reliable initialization for the finer segmentation. With the automatic initialization, the deformable model refines the segmentation to fit the real object boundary accurately. Since the right kidney is very close spatially and has similar intensities to the liver region, the deformable model can

---

(a) Subject 1   (b) Subject 2
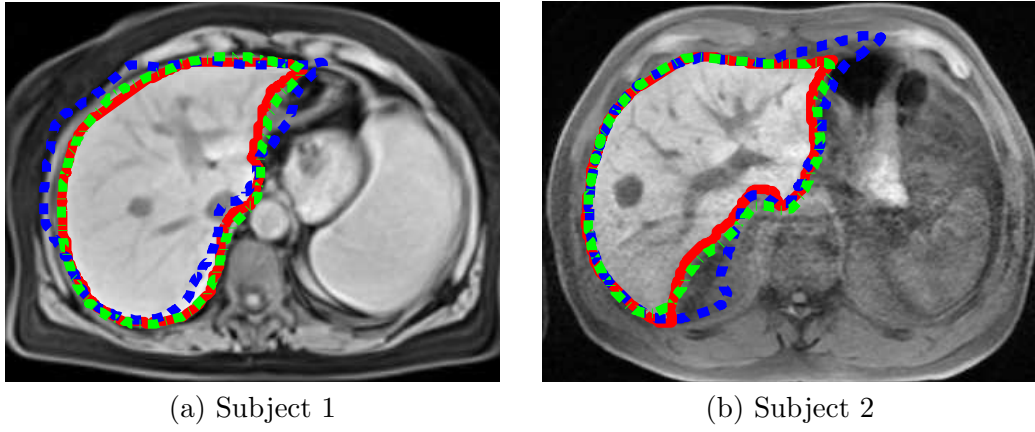
Figure 4.5: Visual comparison of initial and final segmentations in an axial slice from two samples. The blue dotted lines are the surfaces of initializations. The green dotted lines are the surfaces of final segmentation results. And ground truth delineations are represented by the red solid lines.

(a)   (b)

(c)   (d)
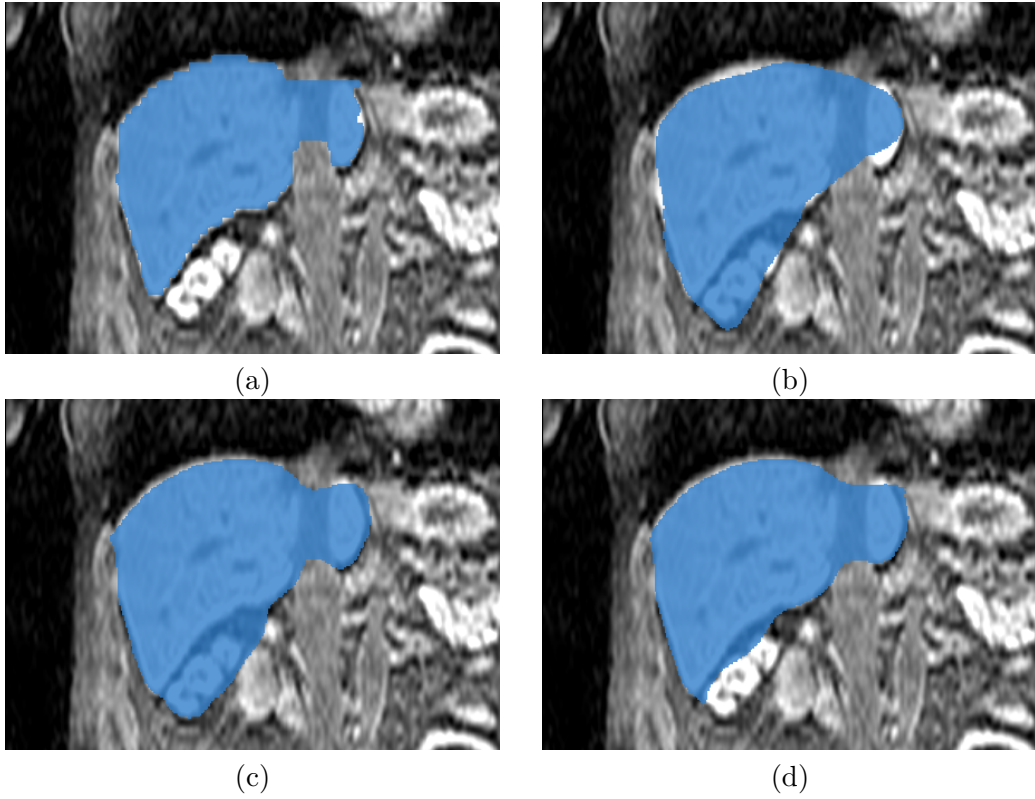
Figure 4.6: Visual comparison of proposed deformable model with or without the shape constraint. (a) is ground truth segmentation; (b) is the initial segmentation from atlas registration; (c) is the result from proposed deformable model without any shape prior; (d) is the result from proposed method with the sparse shape prior.

Table 4.1: Quantitative comparison of segmentations in overlapping measurement. Here, μ is the mean value and σ is the standard deviation of the measurements.

Method	Dice score		Accuracy		Relative error	
	μ	σ	μ	σ	μ	σ
ATLAS	0.70	0.13	0.82	0.07	0.63	0.39
AGC	0.71	0.30	0.84	0.14	0.52	0.49
Initial	0.76	0.13	0.84	0.08	0.58	0.47
Ours	**0.86**	**0.05**	**0.91**	**0.05**	**0.29**	**0.08**

easily leak to the kidney region. In Fig. 4.6, it demonstrates that the sparse shape representation can help to exclude the over-segmented region (i.e. the kidney or heart) effectively.

After showing the effectiveness of our proposed deformable model, we compare the quantitative performances of different approaches. Quantitative comparisons are shown in Table 4.1. This table shows that our proposed method produce the best performance with respect to the overlapping measurements. It has larger average dice score and accuracy, smaller relative error, as well as smaller variances (σ). The ATLAS approach has relatively small variance but the lowest average performance. The AGC approach is little better than ATLAS in the average performance, but the worst in the variance. Our atlas-based initialization is comparable to ATLAS method, which provides good initialization for our deformable model. The comparison results indicate that the proposed method is accurate and robust comparing with the other two approaches. Similar observations are shown in the Fig. 4.7. In this figure, we also plot the proposed method without shape constraint (denoted as OursN) besides the above three methods. Comparing the OursN and Ours method, we observe a slightly better average performance and less variance in the method with shape constraint. As shown in Fig. 4.6, the difference is not big because over-segmentation of deformable model only appear in small areas where nearby anatomies (e.g. kidney or heart) have similar intensities. With more shapes in the training dataset, the sparse shape composition should be more accurate. Due to the limited training samples in our application, the sparse shape composition may not be beneficial to every testing case.

Fig. 4.8 compares segmentation results in an axial slice for two subjects. The red
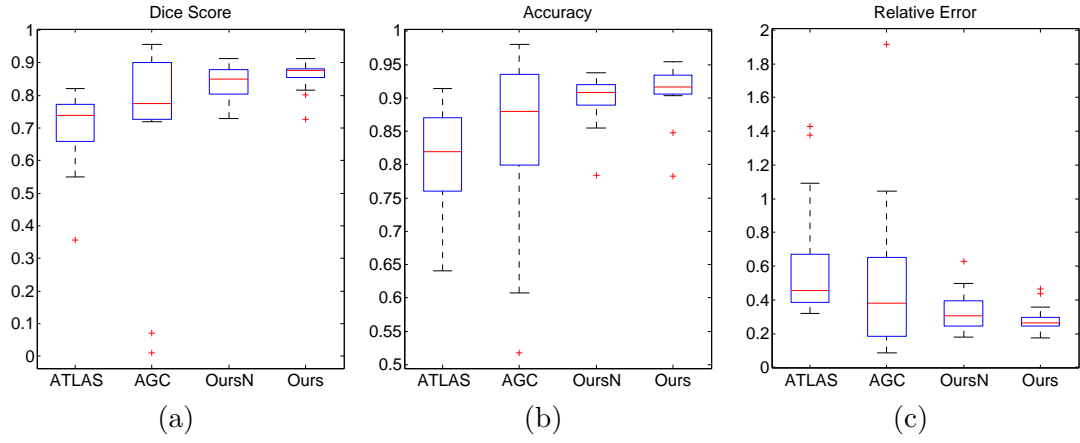
Figure 4.7: Statistical comparisons of dice scores (a), accuracies (b) and relative errors (c). In each plot, the four boxes are for ATLAS result, AGC result, OursN result (proposed method without shape constraint), and Ours result (proposed method with shape constraint), respectively.

(a) Subject 1   (b) Subject 2
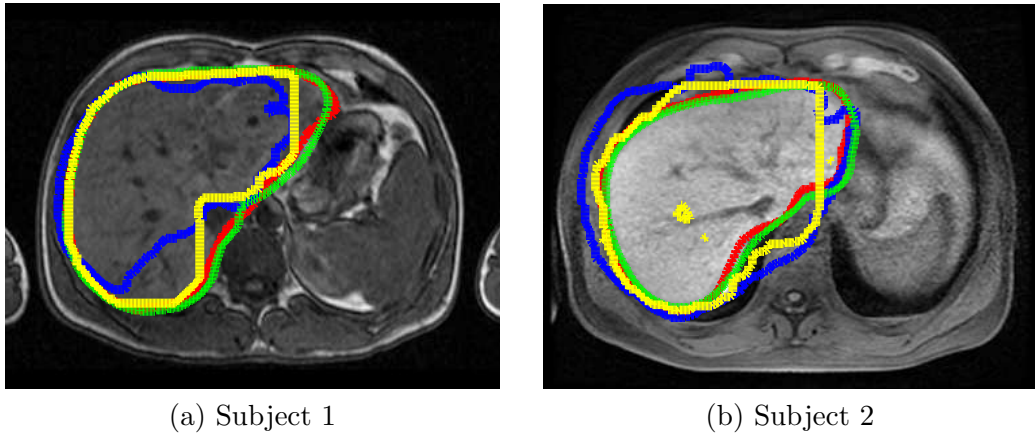
Figure 4.8: Visual comparison of ATLAS results (blue lines), AGC results (yellow lines) and our proposed results (green lines) against manual delineations (red lines) in an axial slice from two subjects.

lines are ground truth boundaries, and the green lines are the segmentation results from our proposed method. The ATLAS results (blue) are not accurate mainly due to the approximate registration we use for the atlas-to-target transformation. The inaccuracy of AGC results may due to the inaccurate initialization and the image inhomogeneities. Fig. 4.9 shows visual comparisons in 3D view. The surface distance errors are plotted for three subjects (three columns). Fig. 4.9(a)-(c) show the surface distances between ground truth and ATLAS method, while Fig. 4.9(d)-(f) show the error map for AGC approach. Results by our proposed method are shown in the last row (Fig. 4.9(g)-(i)), which demonstrate the least errors. The mean surface errors of the segmentations are $17.95mm$, $19.02mm$ and $7.07mm$ for ATLAS, AGE and Ours method, with standard deviations 16.14, 24.23 and 6.68, respectively. This figure shows that the proposed method has the best performance with regard to the surface distance errors. With robust liver segmentation results, we can further conduct the hepatic fat-fraction assessment in clinical trials by using a magnitude-based chemical shift method as shown in [189].

4.3.2 Intuitive Segmentation Refinement

We validate methods on two data sets. The first one is the data set of T1-weighted MR scans for liver described in the Sec. 4.3.1. The second one includes 10 3D CT chest scans from which we segment the right lungs. The CT chest images have 0.98mm in-plane resolution and 3mm between-slice resolution. The in-plance sizes are 512×512, the number of slices range from 97 to 148.

We first use TurtleSeg software [177] to obtain all the initial segmentations with no more than 7 labeled contours per subject. The initial segmented objects are exported as surface meshes and binary mask images for different interactive methods to refine. Since it is hard to measure the interaction efforts during the refinement, we limit different methods to only edit or correct labels at roughly same location in the same amount of slices. Then we compare the refined segmentation results. The graph cut is conducted with 26-neighborhood connectivity in down-sampled input images due to the large memory consumptions. Our model is initialized by decimated surface meshes. And the
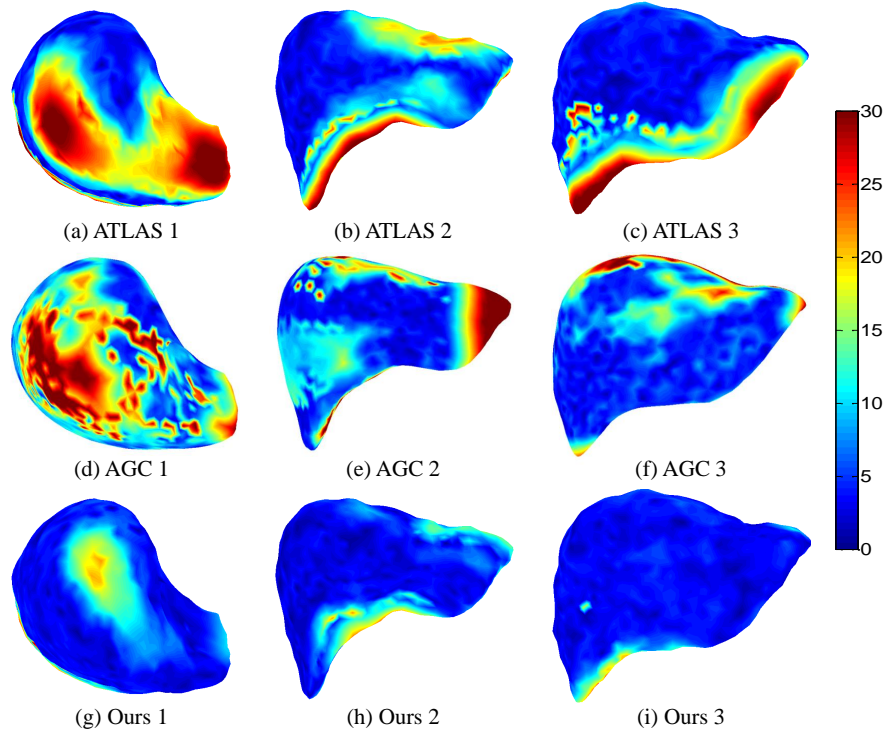
Figure 4.9: Visual comparison of surface distance errors for three subjects. Each column is for one subject. First row is for ATLAS segmentations; second row is for AGC results; the third row is for our proposed method.
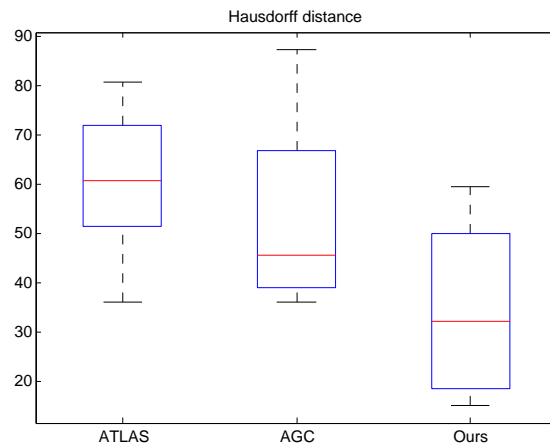
Figure 4.10: Statistical comparison of Hausdorff distance for segmentation results by ATLAS, AGC and our proposed method.

(a) Case 1: MRI liver   (b) Case 2: CT lung
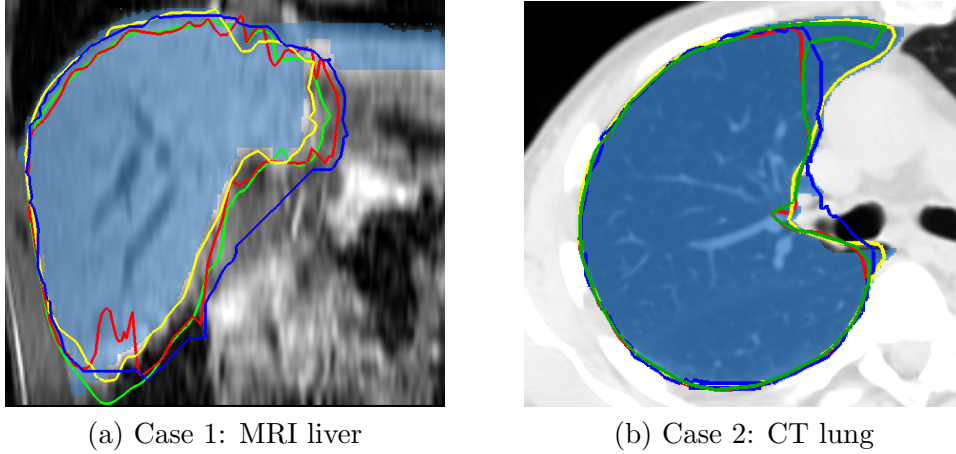
Figure 4.11: 2D visual comparison of refined results by different methods. The objects of interest are masked. Red lines are initial models; yellow, green and blue lines are our model, AL and GC refinement, respectively.

parameters are tuned for one data and applied to all the others in our data sets. $\gamma = 1$, $\alpha = 0.15$, $\beta = 0$, $\kappa = 1$ and $\lambda = 1$. We first analyze the improvements generated by control points in our framework. Then we qualitatively and quantitatively compare the refinements with two interactive methods in [176] and [177]. Dice Similarity Coefficient (DSC), surface distance error and Hausdorff distance are reported.

Since our interactive framework is integrated in a parametric deformable model approach, the traditional matrix A and the internal constraints by control points are assembled once before the iteration. In each iteration, only $\hat{f}^{(t-1)}$ is updated and new $V^{(t)}$ is solved efficiently. In our cases, the surface meshes have about 2000 vertices, and the online computation is fast, about 5 frames per second, which allows real-time feedbacks of the interactive controls.

Before comparing different approaches, we first show the difference between our method with or without local image context information. In Fig. 4.4, (b) shows the refined model by editing with only geometric constraint (3rd term in Eq. 4.8), but without F_C; (c) shows the result by editing with both geometric constraint and local image force F_C. In this case, only one control point easily correct the segmentation error in the initial model. And we can observe that only geometric editing is not enough for desired object boundary. By combining both geometric and image forces, it achieves better improvement.

Table 4.2: Quantitative comparison of segmentations in overlapping measurement and Hausdorff distance. Initial means the initial models. CG and AL stand for Graph Cut and Active Learning-based refinement, respectively.

Method	MRI Liver					
	Global DSC		Local DSC		Local Hausdorff	
	$\mu(\%)$	σ	$\mu(\%)$	σ	$\mu(\text{mm})$	σ
Initial	84.61	0.07	44.35	0.19	8.41	4.29
GC [176]	81.89	0.09	65.87	0.10	7.84	4.20
AL [177]	86.31	0.06	67.14	0.05	7.91	4.71
Ours1	86.69	0.06	65.84	0.16	5.45	3.50
Ours2	**87.27**	**0.06**	**71.79**	0.12	**5.12**	**2.24**

Table 4.3: Quantitative comparison of segmentations in overlapping measurement and Hausdorff distance. Initial means the initial models. CG and AL stand for Graph Cut and Active Learning-based refinement, respectively.

Method	CT Lung					
	Global DSC		Local DSC		Local Hausdorff	
	$\mu(\%)$	σ	$\mu(\%)$	σ	$\mu(\text{mm})$	σ
Initial	90.08	0.03	24.63	0.27	10.50	7.2
GC [176]	87.28	0.05	74.05	0.28	4.14	4.13
AL [177]	90.89	0.03	71.48	0.14	4.51	2.49
Ours1	92.21	0.02	83.20	0.03	4.08	2.28
Ours2	**92.43**	**0.02**	**85.00**	**0.03**	**3.96**	**2.26**

After showing the effectiveness of our proposed interactive framework, we compare the qualitative and quantitative performances of different approaches. Fig. 4.11 shows visual comparisons for two cases in 2D view. The yellow contour produced by the proposed method fit to the object boundary nicely. For the surface distance errors, the mean for initial model, Active Learning result, Graph Cut result, and our results in MRI liver data set are 8.30, $7.65mm$, $10.28mm$ and $7.42mm$, respectively. The mean distance errors in CT lung data set are $5.61mm$, $5.13mm$, $6.24mm$ and $4.21mm$, respectively. A detailed 3D visual comparisons can be found in [190].

Quantitative comparisons of overlapping accuracies and Hausdorff distances are shown in Table 4.2 and Table 4.3. Here we interactively refine initial segmentations in no more than three 2D slices using different methods and summarize all the refined results. **Ours1** are results with only geometric editing (without F_C). **Ours2** are results using all information from geometry and image. Since the overall overlapping measurement may not reflect the local refinement clearly, we measure local performances in a

small ROI containing the interaction site. It shows that all the methods except GC improve the global dice scores, because graph cut may need much more background labels to avoid mislabeling. In both global and local measurements, **Ours2** produces the best performance, demonstrating the efficacy of our method.

4.3.3 Discussions

1) For the online automatic segmentation, most of the time is spent on the nonrigid registration from trained statistical atlas to a testing sample (around 4 minutes). Besides of the registration, ATLAS method uses about 1 minutes per subject, AGC method and the proposed method need about 4 minutes. Note that the cost of memory and time increase dramatically as the image size increases in graph cut algorithm. If the testing subject is re-sampled to $1mm \times 1mm \times 1mm$ resolution (needed in some isotropic measurement), the memory cost of AGC could boost to around 6GB in some cases in our experiment, while the memory and time cost is relatively stable in the proposed method, since our deformable model has small number of vertexes independent to the image size.

2) All the parameters used in the automatic methods, including nonrigid registration, deformable model and sparse shape constraint, are tuned experimentally for one case and then apply for all the testing samples. Thus, the above comparison results indicate that our proposed approach has better and robust performance. As discussed before, the atlas based prior knowledge (including image and shape) can help the initialization and evolution of the deformable model. Thus, the proposed method does not require a large number of consistent training samples and ad-hoc designed features in learning based approaches. This is a very important feature which can benefit in other segmentation tasks in clinical researches.

3) For segmentation refinement, GC approach requires users to provide seeds to constrain the weak or misleading appearance cues and correct mislabeled regions. However, the additional user inputs also affect the global statistics and may cause some other places to be mislabeled. In the AL approach, the users can only edit the segmentation

result by adding or editing 2D contours and re-generating the 3D surface. On the contrary, our approach can intuitively and directly edit the shape at local region to refine the segmentation result.

4.4   Summary

In this chapter, we present an automatic 3D medical image segmentation method incorporating atlas prior with the deformable model. It can be easily extended for segmentation refinement by utilizing the Laplacian-coordinate-based geometric information and local image texture information indicated by the user specified control points. The constraints and supplemental force field introduced by control points are integrated to form a unified framework. The matrix representation is efficient to compute. From the experimental results, we observe accuracy improvement both qualitatively and quantitatively, compared to baseline methods. Our refinement approach requires fewer user interaction and provides real-time results, and thus has the potential to improve work efficiency in many clinical applications.

Chapter 5

Conclusions

In this thesis, we present developments of automatic and robust anatomy recognition and segmentation algorithms for medical image analysis. Our contributions are summarized as follows.

1. First, we present a two-stage multi-instance deep learning based algorithm to identify anatomical body parts in medical image slices. The proposed method does not require annotations of the discriminative local patches. Instead, it automatically *discovers* these local image patches through multi-instance deep learning for image classification task. The method is evaluated on a synthetic data set and CT data set, and shows $> 4\%$ improvements in terms of F_1 score compared with other methods, including conventional learning and standard CNN. Since no manual annotations are required to label the local patches, this weakly supervised discriminative patch discovery and classification method could be easily applied to other image classification tasks where local information should be leveraged but no prior knowledge provided. It may benefit radiological workflow in different aspects. For example, it could make the planning of the scanning range in topogram or scout scans be conducted on-the-fly to significantly save scanning time. It could also serve as an initialization module for other higher level medical image interpretation tasks, e.g. anatomy segmentation, to make the work flow more efficient in clinical studies.

2. Second, we propose an adaptive probabilistic atlas based algorithm for multiple brain structure segmentation. In medical image segmentation, an atlas is a pair of image data and its corresponding label map as ground truth. The probabilistic atlas approach aligns the training multi-atlas into a common coordinate system

to extract prior information to guide target segmentation. Our adaptive method utilizes the priors from the training set but also estimate the extent to which those priors apply on the target. The adaptivity is finely controlled by a non-stationary prior relaxation map, which determines the local relaxation parameters automatically in the individual voxel level and structure level. The adaptive atlas model and the label-smooth MRF are constructed in a target-specific fashion and the complete model is further evolved to fit the target by EM optimization. This method can finely adapt the learned priors from the training set to new testing cases, even if the testing case is not well represented by the training population. In a battery of tests, we find that our approach simultaneously and accurately segments over 30 individual anatomical structures throughout the AD brain using healthy training brains. Due to the finely controlled adaptivity, our method yields higher parcellation accuracy ($> 2\%$ in terms of overall dice score) than other leading methods, including FreeSurfer, iSTAPLE, non-local Staple and joint label fusion approaches. Therefore, it could be used in similar voxel labeling tasks for large-scale clinical researches where the available training samples are limited.

3. Besides multi-object segmentation, we develop a deformable model based algorithm by utilizing atlas-based priors for single-object segmentation. This method utilizes symmetric diffeomorphic image registration to learn a statistical image atlas, and uses mesh quality-preserved deformable model to build shape atlases from a limited number of training samples. The statistical image atlas provides a reliable initialization, and contributes in the external potential energy for the deformable model. The learnt shape atlas is used to generate a constraint for the deformation by sparse shape composition model. We validate this method in a dataset of 3D abdominal MR scans to segment liver. Compared with baseline methods, our approach is more robust to various anatomical shapes and diverse data sources coming from different scanners with inconsistent imaging protocols and scanning parameters. We anticipate that an automated robust algorithm can benefit quantitative analysis in clinical trials.

4. To better control segmentation accuracy, we extend our deformable model to enable intuitive result refinement by interactive mesh editing. This mesh editing method combines Laplacian detail-preserving geometric information and image context information for the local boundaries to correct local mistakes on the segmentation surface. It only requires few user interaction (sparse control points) and updates the surface mesh in real time. This tool is very useful for clinical experts who would like to correct the automatic results to obtain more accurate quantitative assessments.

Large-scale and fine-grained medical image analytics is important and attract more attentions these years [191, 192, 193, 194, 195]. Although there are more and more medical images collected in clinical researches, few methods are able to analyze large-scale medical image databases in real-time, not to mention incorporating user-provided domain knowledge interactively. Our approaches are developed towards bridging the gap between large medical image data and limited training samples.

In future, there are some opportunities revealed from our work to help towards large-scale medical studies and data analytics in healthcare. For instance, our deep learning based anatomy recognition method could enable scalable content-based image retrieval [38, 39] by learning a similarity measurement of the learned hierarchical features (from CNN classification) to improve the PACS systems. The retrieval results could help to annotate more medical images. If sufficient 3D training data available, it would be straight forward to extend this method to handle 3D cases directly by replacing the 2D filters with 3D filters in the convolutional layers of CNN. It is also possible to extend it to multi-modal deep learning [196, 197]. It is not limited in just multiple image modalities [198]. Since shape prior models have shown the benefits in many applications, the geometric information should be able to act as a complementary modality besides visual appearance to be incorporated in deep learning frameworks [199, 200].

In terms of the anatomy segmentation, the proposed atlas-based segmentation methods does not require expensive training or prior knowledge of pathological variance. This feature is very important in many clinical trials which may only have limited training

samples. We could apply the multi-object segmentation framework to other tasks. One example is muscle and fat segmentation in the thigh for knee osteoarthritis analysis [201]. Since different muscle blocks and types (subcutaneous, inter-muscular and intra-muscular) of adipose tissues may only be differentiated by the spatial prior knowledge, it is similar to the application of brain segmentation. One major limitation of this method is the pre-defined number of structures. It would be very interesting to extend the method to handle new or missing structures not explained by the model. The missing structures could be handled by introducing some latent variables to represent the absence of structures, while the new structure, such as lesions not present in our training data, could be detected and segmented separately [202, 203]. Our single-object segmentation approach could be easily extended to handle multi-modality images for more accurate segmentation by incorporating more comprehensive boundary features from different modalities into the energy term of the deformable model. With the help of the interactive result refinement, researchers can better control the quality of automatic segmentation results. Those accurate results are expected to benefit many higher-level medical data analytics for more reliable outcomes, including population health analytics, disease classification or disease stage prediction.

References

[1] F. Mazzocchi and L. A. Vignolo, "Localisation of lesions in aphasia: clinical-ct scan correlations in stroke patients," *Cortex*, vol. 15, no. 4, pp. 627–653, 1979.

[2] S. Ogawa, T.-M. Lee, A. R. Kay, and D. W. Tank, "Brain magnetic resonance imaging with contrast dependent on blood oxygenation," *Proceedings of the National Academy of Sciences*, vol. 87, no. 24, pp. 9868–9872, 1990.

[3] D. L. Bailey, D. W. Townsend, P. E. Valk, and M. N. Maisey, *Positron emission tomography*. Springer, 2005.

[4] D. W. Townsend and T. Beyer, "A combined pet/ct scanner: the path to true image fusion," *The British Journal of Radiology*, vol. 75, no. suppl_9, pp. S24–S30, 2002.

[5] M. S. Judenhofer, H. F. Wehrl, D. F. Newport, C. Catana, S. B. Siegel, M. Becker, A. Thielscher, M. Kneilling, M. P. Lichy, M. Eichner *et al.*, "Simultaneous pet-mri: a new approach for functional and morphological imaging," *Nature Medicine*, vol. 14, no. 4, pp. 459–465, 2008.

[6] G. Antoch and A. Bockisch, "Combined pet/mri: a new dimension in whole-body oncology imaging?" *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 36, no. 1, pp. 113–120, 2009.

[7] T. McInerney and D. Terzopoulos, "Deformable models in medical image analysis: a survey," *Medical Image Analysis*, vol. 1, no. 2, pp. 91–108, 1996.

[8] J. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical Image Analysis*, vol. 2, no. 1, pp. 1–36, 1998.

[9] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annual Review of Biomedical Engineering*, vol. 2, no. 1, pp. 315–337, 2000.

[10] D. L. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Physics in Medicine and Biology*, vol. 46, no. 3, p. R1, 2001.

[11] M. Betke, H. Hong, D. Thomas, C. Prince, and J. P. Ko, "Landmark detection in the chest and registration of lung surfaces with an application to nodule registration," *Medical Image Analysis*, vol. 7, no. 3, pp. 265–281, 2003.

[12] M. Dikmen, Y. Zhan, and X. S. Zhou, "Joint detection and localization of multiple anatomical landmarks through learning," in *Medical Imaging*. International Society for Optics and Photonics, 2008, pp. 691 538–691 538.

[13] Y. Zheng, D. Liu, B. Georgescu, H. Nguyen, and D. Comaniciu, "3D deep learning for efficient and robust landmark detection in volumetric data," in *Medical Image Computing and Computer Assisted Intervention*. Springer, 2015, pp. 565–572.

[14] S. Mitchell, J. Bosch, B. Lelieveldt, R. van der Geest, J. Reiber, and M. Sonka, "3-D active appearance models: segmentation of cardiac MR and ultrasound images," *IEEE Transactions on Medical Imaging*, vol. 21, no. 9, pp. 1167–1178, 2002.

[15] D. Shen, S. Moffat, S. M. Resnick, and C. Davatzikos, "Measuring size and shape of the hippocampus in MR images using a deformable shape model," *NeuroImage*, vol. 15, no. 2, pp. 422–434, 2002.

[16] H. Ling, S. Zhou, Y. Zheng, B. Georgescu, M. Suehling, and D. Comaniciu, "Hierarchical, learning-based automatic liver segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008, pp. 1–8.

[17] C. Li, R. Huang, Z. Ding, J. C. Gatenby, D. N. Metaxas, and J. C. Gore, "A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 2007–2016, 2011.

[18] R. Bellotti, F. De Carlo, G. Gargano, S. Tangaro, D. Cascio, E. Catanzariti, P. Cerello, S. C. Cheran, P. Delogu, I. De Mitri *et al.*, "A cad system for nodule detection in low-dose lung cts based on region growing and a new active contour model," *Medical Physics*, vol. 34, no. 12, pp. 4901–4910, 2007.

[19] L. A. Meinel, A. H. Stolpen, K. S. Berbaum, L. L. Fajardo, and J. M. Reinhardt, "Breast MRI lesion classification: Improved performance of human readers with a backpropagation neural network computer-aided diagnosis (CAD) system," *Journal of Magnetic Resonance Imaging*, vol. 25, no. 1, pp. 89–95, 2007.

[20] K. Doi, "Current status and future potential of computer-aided diagnosis in medical imaging," *The British Journal of Radiology*, 2014.

[21] A. P. Zijdenbos, R. Forghani, and A. C. Evans, "Automatic "pipeline" analysis of 3-D MRI data for clinical trials: application to multiple sclerosis," *IEEE Transactions on Medical Imaging*, vol. 21, no. 10, pp. 1280–1291, 2002.

[22] A. Association *et al.*, "2016 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 12, no. 4, pp. 459–509, 2016.

[23] http://orthoinfo.aaos.org/topic.cfm?topic=a00389.

[24] S. Kurtz, K. Ong, E. Lau, F. Mowat, and M. Halpern, "Projections of primary and revision hip and knee arthroplasty in the united states from 2005 to 2030," *J Bone Joint Surg Am*, vol. 89, no. 4, pp. 780–785, 2007.

[25] L. Ferrarini, W. M. Palm, H. Olofsen, M. A. van Buchem, J. H. Reiber, and F. Admiraal-Behloul, "Shape differences of the brain ventricles in Alzheimer's disease," *NeuroImage*, vol. 32, no. 3, pp. 1060 – 1069, 2006.

[26] S. M. Nestor, R. Rupsingh, M. Borrie, M. Smith, V. Accomazzi, J. L. Wells, J. Fogarty, R. Bartha *et al.*, "Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database," *Brain*, vol. 131, no. 9, pp. 2443–2454, 2008.

[27] B. R. Ott, R. A. Cohen, A. Gongvatana, O. C. Okonkwo, C. E. Johanson, E. G. Stopa, J. E. Donahue, and G. D. Silverberg, "Brain ventricular volume and cerebrospinal fluid biomarkers of Alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 20, no. 2, pp. 647–657, 2010.

[28] M. Lehmann, S. J. Crutch, G. R. Ridgway, B. H. Ridha, J. Barnes, E. K. Warrington, M. N. Rossor, and N. C. Fox, "Cortical thickness and voxel-based morphometry in posterior cortical atrophy and typical Alzheimer's disease," *Neurobiology of Aging*, vol. 32, no. 8, pp. 1466–1476, 2011.

[29] C. Peterfy, E. Schneider, and M. Nevitt, "The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee," *Osteoarthritis and Cartilage*, vol. 16, no. 12, pp. 1433–1441, 2008.

[30] G. Vincent, C. Wolstenholme, I. Scott, and M. Bowes, "Fully automatic segmentation of the knee joint using active appearance models," *Medical Image Analysis for the Clinic: A Grand Challenge*, pp. 224–230, 2010.

[31] R. Daffner, A. Lupetin, N. Dash, Z. Deeb, R. Sefczek, and R. Schapiro, "Mri in the detection of malignant infiltration of bone marrow," *American Journal of Roentgenology*, vol. 146, no. 2, pp. 353–358, 1986.

[32] J. Elliott, G. Jull, J. T. Noteboom, R. Darnell, G. Galloway, and W. W. Gibbon, "Fatty infiltration in the cervical extensor muscles in persistent whiplash-associated disorders: a magnetic resonance imaging analysis," *Spine*, vol. 31, no. 22, pp. E847–E855, 2006.

[33] M. W. Strudwick, S. E. Anderson, S. Dimmick, M. D. Saltzman, and W. K. Hsu, "The pearls and pitfalls of magnetic resonance imaging of the upper extremity," *Journal of Orthopaedic & Sports Physical Therapy*, vol. 41, no. 11, pp. 861–872, 2011.

[34] R. Park, T. Nyland, J. Lattimer, C. Miller, and J. Lebel, "B-mode gray-scale ultrasound: Imaging artifacts and interpretation principles," *Veterinary Radiology*, vol. 22, no. 5, pp. 204–210, 1981.

[35] T. B. Smith and K. S. Nayak, "MRI artifacts and correction strategies," *Imaging in Medicine*, vol. 2, no. 4, pp. 445–457, 2010.

[36] F. E. Boas and D. Fleischmann, "CT artifacts: causes and reduction techniques," *Imaging in Medicine*, vol. 4, no. 2, pp. 229–240, 2012.

[37] M. O. Gueld, M. Kohnen, D. Keysers, H. Schubert, B. B. Wein, J. Bredno, and T. M. Lehmann, "Quality of dicom header information for image categorization," in *Medical Imaging*. International Society for Optics and Photonics, 2002, pp. 280–287.

[38] M. Jiang, S. Zhang, H. Li, and D. N. Metaxas, "Computer-aided diagnosis of mammographic masses using scalable image retrieval," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 783–792, 2015.

[39] X. Zhang, W. Liu, M. Dundar, S. Badve, and S. Zhang, "Towards large-scale histopathological image analysis: Hashing-based image retrieval," *IEEE Transactions on Medical Imaging*, vol. 34, no. 2, pp. 496–506, 2015.

[40] Z. Yan, Y. Zhan, Z. Peng, S. Liao, Y. Shinagawa, D. N. Metaxas, and X. S. Zhou, "Bodypart recognition using multi-stage deep learning," in *International Conference on Information Processing in Medical Imaging*. Springer, 2015, pp. 449–461.

[41] Y. Rejani and S. T. Selvi, "Early detection of breast cancer using SVM classifier technique," *arXiv preprint arXiv:0912.2314*, 2009.

[42] V. F. Van Ravesteijn, C. Van Wijk, F. M. Vos, R. Truyen, J. F. Peters, J. Stoker, and L. J. Van Vliet, "Computer-aided detection of polyps in CT colonography using logistic regression," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 120–131, 2010.

[43] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[44] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[45] Y. Wei, W. Xia, J. Huang, B. Ni, J. Dong, Y. Zhao, and S. Yan, "Cnn: Single-label to multi-label," *arXiv preprint arXiv:1406.5726*, 2014.

[46] Z. Yan, Y. Zhan, Z. Peng, S. Liao, Y. Shinagawa, S. Zhang, D. Metaxas, and X. Zhou, "Multi-instance deep learning: Discover discriminative local anatomies for bodypart recognition," *IEEE Transactions on Medical Imaging*, pp. 1332–1343, 2016.

[47] J. Park, G. Kang, S. Pan, and P. Kim, "A novel algorithm for identification of body parts in medical images," in *Fuzzy Systems and Knowledge Discovery*, 2006, pp. 1148–1158.

[48] L. Hong and S. Hong, "Methods and apparatus for automatic body part identification and localization," May 15 2008, US Patent App. 11/933,518.

[49] Y. Zhan, X. S. Zhou, Z. Peng, and A. Krishnan, "Active scheduling of organ detection and segmentation in whole-body medical images," in *Medical Image Computing and Computer Assisted Intervention*. Springer, 2008, pp. 313–321.

[50] A. Criminisi, J. Shotton, D. Robertson, and E. Konukoglu, "Regression forests for efficient anatomy detection and localization in CT studies," in *International MICCAI Workshop on Medical Computer Vision*. Springer, 2010, pp. 106–117.

[51] A. Criminisi, J. Shotton, and S. Bucciarelli, "Decision forests with long-range spatial context for organ localization in CT volumes," in *Medical Image Computing and Computer Assisted Intervention*, 2009, pp. 69–80.

[52] R. Donner, B. H. Menze, H. Bischof, and G. Langs, "Global localization of 3D anatomical structures by pre-filtered Hough Forests and discrete optimization," *Medical Image Analysis*, vol. 17, no. 8, pp. 1304–1314, 2013.

[53] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[54] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1. IEEE, 2005, pp. 886–893.

[55] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[56] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[57] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[58] D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 3642–3649.

[59] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in Neural Information Processing Systems*, 2014, pp. 3104–3112.

[60] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[61] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," in *Advances in Neural Information Processing Systems*, 2013, pp. 2553–2561.

[62] D. Yang, S. Zhang, Z. Yan, C. Tan, K. Li, and D. Metaxas, "Automated anatomical landmark detection ondistal femur surface using convolutional neural network," in *IEEE International Symposium on Biomedical Imaging*. IEEE, 2015, pp. 17–21.

[63] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[64] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, and N. Navab, "Aggnet: Deep learning from crowds for mitosis detection in breast cancer histology images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1313–1321, 2016.

[65] Y. Guo, G. Wu, L. A. Commander, S. Szary, V. Jewells, W. Lin, and D. Shen, "Segmenting hippocampus from infant brains by sparse patch matching with deep-learned features," in *Medical Image Computing and Computer Assisted Intervention*. Springer, 2014, pp. 308–315.

[66] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer Assisted Intervention*. Springer, 2015, pp. 234–241.

[67] T. Brosch, L. Tang, Y. Yoo, D. Li, A. Traboulsee, and R. Tam, "Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1229–1239, 2016.

[68] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. Benders, and I. Isgum, "Automatic segmentation of MR brain images with a convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1252–1261, 2016.

[69] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.

[70] H. R. Roth, C. T. Lee, H.-C. Shin, A. Seff, L. Kim, J. Yao, L. Lu, and R. M. Summers, "Anatomy-specific classification of medical images using deep convolutional nets," in *IEEE International Symposium on Biomedical Imaging*, 2015, pp. 101–104.

[71] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014, pp. 1701–1708.

[72] R. Girshick, "Fast R-CNN," in *International Conference on Computer Vision*, 2015, pp. 1440–1448.

[73] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[74] S. Singh, A. Gupta, and A. A. Efros, "Unsupervised discovery of mid-level discriminative patches," in *European Conference on Computer Vision*. Springer, 2012, pp. 73–86.

[75] M.-M. Cheng, Z. Zhang, W.-Y. Lin, and P. Torr, "BING: Binarized normed gradients for objectness estimation at 300fps," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014, pp. 3286–3293.

[76] O. Maron and T. Lozano-Pérez, "A framework for multiple-instance learning," *Advances in Neural Information Processing Systems*, pp. 570–576, 1998.

[77] S. Andrews, I. Tsochantaridis, and T. Hofmann, "Support vector machines for multiple-instance learning," in *Advances in Neural Information Processing Systems*, 2002, pp. 561–568.

[78] Q. Zhang and S. A. Goldman, "EM-DD: An improved multiple-instance learning technique," in *Advances in Neural Information Processing Systems*, 2001, pp. 1073–1080.

[79] J. Wu, Y. Yu, C. Huang, and K. Yu, "Deep multiple instance learning for image classification and auto-annotation," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2015, pp. 3460–3469.

[80] P. O. Pinheiro and R. Collobert, "From image-level to pixel-level labeling with convolutional networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1713–1721.

[81] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *International Conference on Machine Learning*, 2010, pp. 807–814.

[82] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[83] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural Networks*, vol. 12, no. 1, pp. 145–151, 1999.

[84] J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley, and Y. Bengio, "Theano: a CPU and GPU math expression compiler," in *Proc. of the Python for Scientific Computing Conference (SciPy)*, Jun 2010.

[85] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

[86] P. Sermanet and Y. LeCun, "Traffic sign recognition with multi-scale convolutional networks," in *International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2011, pp. 2809–2813.

[87] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008, pp. 1–8.

[88] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.

[89] M. Cardoso, M. Clarkson, G. Ridgway, M. Modat, N. Fox, and S. Ourselin, "LoAd: A locally adaptive cortical segmentation algorithm," *NeuroImage*, vol. 56, no. 3, pp. 1386–97, 2011.

[90] J. Ashburner and K. Friston, "Unified segmentation," *NeuroImage*, vol. 26, pp. 839–851, 2005.

[91] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based tissue classification of MR images of the brain," *IEEE Transactions on Medical Imaging*, vol. 18, no. 10, pp. 897–908, 1999.

[92] T. Rohlfing, R. Brandt, R. Menzel, D. B. Russakoff, and C. R. Maurer Jr, "Quo vadis, atlas-based segmentation?" in *Handbook of Biomedical Image Analysis*. Springer, 2005, pp. 435–486.

[93] X. Han, L. S. Hibbard, N. P. Ooconnell, and V. Willcut, "Automatic segmentation of parotids in head and neck CT images using multi-atlas fusion," *Medical Image Analysis for the Clinic: A Grand Challenge*, pp. 297–304, 2010.

[94] R. A. Heckemann, S. Keihaninejad, P. Aljabar, D. Rueckert, J. V. Hajnal, and A. Hammers, "Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation," *NeuroImage*, vol. 51, no. 1, pp. 221–227, 2010.

[95] T. Liu, H. Li, K. Wong, A. Tarokh, L. Guo, and S. T. Wong, "Brain tissue segmentation based on DTI data," *NeuroImage*, vol. 38, no. 1, pp. 114 – 123, 2007.

[96] F. Rousseau, P. A. Habas, and C. Studholme, "A supervised patch-based approach for human brain labeling," *IEEE Transactions on Medical Imaging*, vol. 30, no. 10, pp. 1852–1862, 2011.

[97] F. Shi, P.-T. Yap, Y. Fan, J. H. Gilmore, W. Lin, and D. Shen, "Construction of multi-region-multi-reference atlases for neonatal brain MRI segmentation," *NeuroImage*, vol. 51, no. 2, pp. 684–693, 2010.

[98] S. Warfield, K. Zou, and W. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *IEEE Transactions on Medical Imaging*, vol. 23, no. 7, pp. 903–921, 2004.

[99] X. Liu, A. Montillo, E. Tan, and J. Schenck, "iSTAPLE: improved label fusion for segmentation by combining STAPLE with image intensity," in *SPIE Medical Imaging*, vol. 8669, 2013, pp. 86 692O1–86 692O6.

[100] H. Wang, J. W. Suh, S. R. Das, J. B. Pluta, C. Craige, and P. A. Yushkevich, "Multi-atlas segmentation with joint label fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 611–623, 2013.

[101] A. Klein, J. Andersson, B. a. Ardekani, J. Ashburner, B. Avants, M.-C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, P. Hellier, J. H. Song, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R. P. Woods, J. J. Mann, and R. V. Parsey, "Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration," *NeuroImage*, vol. 46, no. 3, pp. 786–802, 2009.

[102] R. Wolz, P. Aljabar, D. Rueckert, R. A. Heckemann, and A. Hammers, "Segmentation of subcortical structures and the hippocampus in brain MRI using graph-cuts and subject-specific a-priori information," in *IEEE International Symposium on Biomedical Imaging*. IEEE, 2009, pp. 470–473.

[103] van der Lijn, Fedde, M. d. Bruijne, S. Klein, T. den Heijer, Y. Y. Hoogendam, van der Lugt, Aad, Breteler, Monique M B, and W. J. Niessen, "Automated brain structure segmentation based on atlas registration and appearance models," *IEEE Transactions on Medical Imaging*, vol. 31, no. 2, pp. 276–286, 2012.

[104] G. Wu, Q. Wang, D. Zhang, F. Nie, H. Huang, and D. Shen, "A generative probability model of joint label fusion for multi-atlas based brain segmentation," *Medical Image Analysis*, vol. 18, no. 6, pp. 881–890, 2014.

[105] A. J. Asman and B. A. Landman, "Non-local statistical label fusion for multi-atlas segmentation," *Medical Image Analysis*, vol. 17, no. 2, pp. 194–208, 2013.

[106] B. Fischl, D. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. Dale, "Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, no. 3, pp. 341–355, 2002.

[107] N. Shiee, P. Bazin, J. Cuzzocreo, A. Blitz, and D. Pham, "Segmentation of brain images using adaptive atlases with application to ventriculomegaly," in *International Conference on Information Processing in Medical Imaging*, 2011, pp. 1–12.

[108] T. Heimann and H.-P. Meinzer, "Statistical shape models for 3D medical image segmentation: a review," *Medical Image Analysis*, vol. 13, no. 4, pp. 543–563, 2009.

[109] X. Huang and G. Tsechpenakis, "Medical image segmentation," *Information Discovery on Electronic Health Records*, vol. 10, pp. 251–289, 2009.

[110] G. Tsechpenakis, "Deformable model-based medical image segmentation," in *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies.* Springer, 2011, pp. 33–67.

[111] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[112] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International Journal of Computer Vision*, vol. 22, no. 1, pp. 61–79, 1997.

[113] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 359–369, 1998.

[114] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.

[115] X. Huang and D. N. Metaxas, "Metamorphs: deformable shape and appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 8, pp. 1444–1459, 2008.

[116] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu, "Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features," *IEEE Transactions on Medical Imaging*, vol. 27, no. 11, pp. 1668–1681, 2008.

[117] Y. Zhan, M. Dewan, and X. S. Zhou, "Cross modality deformable segmentation using hierarchical clustering and learning," in *Medical Image Computing and Computer Assisted Intervention.* Springer, 2009, pp. 1033–1041.

[118] S. Liao, Y. Gao, Y. Shi, A. Yousuf, I. Karademir, A. Oto, and D. Shen, "Automatic prostate MR image segmentation with sparse label propagation and domain-specific manifold regularization," in *International Conference on Information Processing in Medical Imaging.* Springer, 2013, pp. 511–523.

[119] Y. Guo, Y. Zhan, Y. Gao, J. Jiang, and D. Shen, "MR prostate segmentation via distributed discriminative dictionary (DDD) learning," in *IEEE International Symposium on Biomedical Imaging*, 2013, pp. 868–871.

[120] F. Shi, P.-T. Yap, Y. Fan, J. H. Gilmore, W. Lin, and D. Shen, "Construction of multi-region-multi-reference atlases for neonatal brain MRI segmentation," *NeuroImage*, vol. 51, no. 2, pp. 684–693, 2010.

[121] X. Zhuang, K. S. Rhode, R. S. Razavi, D. J. Hawkes, and S. Ourselin, "A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI," *Medical Imaging, IEEE Transactions on*, vol. 29, no. 9, pp. 1612–1625, 2010.

[122] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. N. Metaxas, and X. S. Zhou, "Towards robust and effective shape modeling: Sparse shape composition," *Medical Image Analysis*, vol. 16, no. 1, pp. 265 – 277, 2012.

[123] M. Gao, C. Chen, S. Zhang, Z. Qian, D. Metaxas, and L. Axel, "Segmenting the papillary muscles and the trabeculae from high resolution cardiac CT through restoration of topological handles," in *International Conference on Information Processing in Medical Imaging.* Springer, 2013, pp. 184–195.

[124] D. Shen and C. Davatzikos, "Hammer: hierarchical attribute matching mechanism for elastic registration," *IEEE Transactions on Medical Imaging*, vol. 21, no. 11, pp. 1421–1439, 2002.

[125] S. Ourselin, R. Stefanescu, and X. Pennec, "Robust registration of multi-modal images: towards real-time clinical applications," in *Medical Image Computing and Computer Assisted Intervention.* Springer, 2002, pp. 140–147.

[126] G. Wu, F. Qi, and D. Shen, "Learning-based deformable registration of MR brain images," *IEEE Transactions on Medical Imaging*, vol. 25, no. 9, pp. 1145–1157, 2006.

[127] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 45–57, 2001.

[128] E. van Rikxoort, Y. Arzhaeva, and B. van Ginneken, "Automatic segmentation of the liver in computed tomography scans with voxel classification and atlas matching," in *Proc. MICCAI Workshop 3D Segmentation in the Clinic: A Grand Challenge.* Citeseer, 2007, pp. 101–108.

[129] B. Avants, C. Epstein, M. Grossman, and J. Gee, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," *Medical Image Analysis*, vol. 12, no. 1, pp. 26–41, 2008.

[130] G. Wu, M. Kim, Q. Wang, and D. Shen, "S-HAMMER: Hierarchical attribute-guided, symmetric diffeomorphic registration for MR brain images," *Human Brain Mapping*, 2012.

[131] C. Ledig, R. Wolz, P. Aljabar, J. Lotjonen, R. A. Heckemann, A. Hammers, and D. Rueckert, "Multi-class brain segmentation using atlas propagation and EM-based refinement," in *IEEE International Symposium on Biomedical Imaging*, 2012, pp. 896–899.

[132] C. Ledig, R. A. Heckemann, P. Aljabar, R. Wolz, J. V. Hajnal, A. Hammers, and D. Rueckert, "Segmentation of MRI brain scans using MALP-EM," in *MICCAI Workshop on Multi-Atlas Labeling*, 2012, pp. 79–82.

[133] C. Ledig, R. A. Heckemann, A. Hammers, J. C. Lopez, V. F. Newcombe, A. Makropoulos, J. Lötjönen, D. K. Menon, and D. Rueckert, "Robust whole-brain segmentation: Application to traumatic brain injury," *Medical Image Analysis*, vol. 21, no. 1, pp. 40–58, 2015.

[134] M. R. Sabuncu, Yeo, B T Thomas, K. van Leemput, B. Fischl, and P. Golland, "A generative model for image segmentation based on label fusion," *IEEE Transactions on Medical Imaging*, vol. 29, no. 10, pp. 1714–1729, 2010.

[135] X. Artaechevarria, A. Munoz-Barrutia, and C. Ortiz-de Solorzano, "Combination strategies in multi-atlas image segmentation: application to brain MR data," *IEEE Transactions on Medical Imaging*, vol. 28, no. 8, pp. 1266–1277, 2009.

[136] P. Coupé, J. V. Manjón, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins, "Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation," *NeuroImage*, vol. 54, no. 2, pp. 940–954, 2011.

[137] A. J. Asman and B. A. Landman, "Non-local STAPLE: An intensity-driven multi-atlas rater model," in *Medical Image Computing and Computer Assisted Intervention*. Springer, 2012, pp. 426–434.

[138] H. Wang, B. Avants, and P. A. Yushkevich, "Grand Challenge on Multi-Atlas Segmentation: A Combined Joint Label Fusion and Corrective Learning Approach," in *MICCAI Workshop on Multi-Atlas Labeling*, 2012, pp. 91–94.

[139] H. Wang and P. A. Yushkevich, "Multi-atlas segmentation with joint label fusion and corrective learningjan open source implementation," *Frontiers in Neuroinformatics*, vol. 7, 2013.

[140] B. A. Landman and S. K. Warfield, "Miccai 2012 workshop on multi-atlas labeling," in *MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling Challenge Results*, 2012.

[141] Z. Yan, S. Zhang, X. Liu, D. N. Metaxas, and A. Montillo, "Accurate segmentation of brain images into 34 structures combining a non-stationary adaptive statistical atlas and a multi-atlas with applications to Alzheimer's disease," in *IEEE International Symposium on Biomedical Imaging*. IEEE, 2013, pp. 1202–1205.

[142] Z. Yan, S. Zhang, X. Liu, D. N. Metaxas, A. Montillo *et al.*, "Accurate whole-brain segmentation for Alzheimer's disease combining an adaptive statistical atlas and multi-atlas," in *International MICCAI Workshop on Medical Computer Vision*. Springer, 2013, pp. 65–73.

[143] J. Iglesias, C. Liu, P. Thompson, and Z. Tu, "Robust brain extraction across datasets and comparison with publicly available methods," *IEEE Transactions on Medical Imaging*, vol. 30, no. 9, pp. 1617–1634, 2011.

[144] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, "N4ITK: improved N3 bias correction," *IEEE Transactions on Medical Imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.

[145] A. C. Evans, D. L. Collins, S. Mills, E. Brown, R. Kelly, and T. M. Peters, "3D statistical neuroanatomical models from 305 MRI volumes," in *Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record*. IEEE, 1993, pp. 1813–1817.

[146] D. L. Collins, P. Neelin, T. M. Peters, and A. C. Evans, "Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space," *Journal of Computer Assisted Tomography*, vol. 18, no. 2, pp. 192–205, 1994.

[147] S. Sanjay-Gopal and T. J. Hebert, "Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm," *IEEE Transactions on Image Processing*, vol. 7, no. 7, pp. 1014–1028, 1998.

[148] J. Zhang, "The mean field theory in EM procedures for Markov random fields," *IEEE Trans. Signal Process.*, vol. 40, no. 10, pp. 2570–2583, 1992.

[149] B. Aubert-Broche, A. C. Evans, and L. Collins, "A new improved version of the realistic digital brain phantom," *NeuroImage*, vol. 32, no. 1, pp. 138–145, 2006.

[150] B. Aubert-Broche, M. Griffin, G. B. Pike, A. C. Evans, and D. L. Collins, "Twenty new digital brain phantoms for creation of validation image data bases," *IEEE Transactions on Medical Imaging*, vol. 25, no. 11, pp. 1410–1416, 2006.

[151] K. Ellis *et al.*, "The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease," *International Psychogeriatrics*, vol. 21, no. 04, pp. 672–687, 2009.

[152] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka *et al.*, "3D Slicer as an image computing platform for the Quantitative Imaging Network," *Magnetic Resonance Imaging*, vol. 30, no. 9, pp. 1323–1341, 2012.

[153] G. Wu, H. Jia, Q. Wang, and D. Shen, "Sharpmean: groupwise registration guided by sharp mean image and tree-based registration," *NeuroImage*, vol. 56, no. 4, pp. 1968–1981, 2011.

[154] B. Fischl, D. H. Salat, A. J. van der Kouwe, N. Makris, F. Ségonne, B. T. Quinn, and A. M. Dale, "Sequence-independent segmentation of magnetic resonance images," *NeuroImage*, vol. 23, pp. S69–S84, 2004.

[155] M. Modat, G. R. Ridgway, Z. A. Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin, "Fast free-form deformation using graphics processing units," *Computer Methods and Programs in Biomedicine*, vol. 98, no. 3, pp. 278–284, 2010.

[156] J. M. Clark and A. M. Diehl, "Nonalcoholic fatty liver disease," *JAMA: the journal of the American Medical Association*, vol. 289, no. 22, pp. 3000–3004, 2003.

[157] A. J. Sanyal, C. Campbell-Sargent, F. Mirshahi, W. B. Rizzo, M. J. Contos, R. K. Sterling, V. A. Luketic, M. L. Shiffman, and J. N. Clore, "Nonalcoholic steatohepatitis: association of insulin resistance and mitochondrial abnormalities," *Gastroenterology*, vol. 120, no. 5, pp. 1183–1192, 2001.

[158] P. Angulo, "Nonalcoholic fatty liver disease," *New England Journal of Medicine*, vol. 346, no. 16, pp. 1221–1231, 2002.

[159] L. A. Adams, J. F. Lymp, J. St Sauver, S. O. Sanderson, K. D. Lindor, A. Feldstein, and P. Angulo, "The natural history of nonalcoholic fatty liver disease: a population-based cohort study," *Gastroenterology*, vol. 129, no. 1, pp. 113–121, 2005.

[160] S. Reeder, I. Cruite, G. Hamilton, and C. Sirlin, "Quantitative assessment of liver fat with magnetic resonance imaging and spectroscopy," *Journal of Magnetic Resonance Imaging*, vol. 34, no. 4, pp. 729–749, 2011.

[161] T. Heimann, B. van Ginneken, M. Styner, Y. Arzhaeva, V. Aurich, and et al., "Comparison and evaluation of methods for liver segmentation from CT datasets," *IEEE Transactions on Medical Imaging*, vol. 28, no. 8, pp. 1251–1265, 2009.

[162] D. Kainmüller, T. Lange, and H. Lamecker, "Shape constrained automatic segmentation of the liver based on a heuristic intensity model," in *Proc. MICCAI Workshop 3D Segmentation in the Clinic: A Grand Challenge*, 2007, pp. 109–116.

[163] M. Linguraru, J. Sandberg, Z. Li, F. Shah, and R. Summers, "Automated segmentation and quantification of liver and spleen from CT images using normalized probabilistic atlases and enhancement estimation," *Medical Physics*, vol. 37, pp. 771–783, 2010.

[164] L. Massoptier and S. Casciaro, "Fully automatic liver segmentation through graph-cut technique," in *IEEE Engineering in Medicine and Biology Society*. IEEE, 2007, pp. 5243–5246.

[165] C. Platero, J. Poncela, P. Gonzalez, M. C. Tobar, J. Sanguino, G. Asensio, and E. Santos, "Liver segmentation for hepatic lesions detection and characterisation," in *IEEE International Symposium on Biomedical Imaging*, 2008, pp. 13–16.

[166] R. Logeswaran, T. Haw, and S. Sarker, "Liver isolation in abdominal MRI," *Journal of Medical Systems*, vol. 32, no. 4, pp. 259–268, 2008.

[167] K. Haris, S. N. Efstratiadis, N. Maglaveras, and A. K. Katsaggelos, "Hybrid image segmentation using watersheds and fast region merging," *IEEE Transactions on Image Processing*, vol. 7, no. 12, pp. 1684–1699, 1998.

[168] R. Siewert, D. Schnapauff, T. Denecke, T. Tolxdorff, and D. Krefting, "Automatic liver segmentation in contrast-enhanced MRI," in *Bildverarbeitung für die Medizin*, 2010, pp. 405–409.

[169] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, Jan. 1995.

[170] D. Shen, E. H. Herskovits, and C. Davatzikos, "An adaptive-focus statistical shape model for segmentation and shape modeling of 3-D brain structures," *IEEE Transactions on Medical Imaging*, vol. 20, no. 4, pp. 257–270, 2001.

[171] S. Zhang, Y. Zhan, and D. N. Metaxas, "Deformable segmentation via sparse representation and dictionary learning," *Medical Image Analysis*, vol. 16, no. 7, pp. 1385–1396, 2012.

[172] P. Slagmolen, A. Elen, D. Seghers, D. Loeckx, F. Maes, and K. Haustermans, "Atlas based liver segmentation using nonrigid registration with a b-spline transformation model," in *Proc. MICCAI Workshop 3D Segmentation in the Clinic: A Grand Challenge.* Citeseer, 2007, pp. 197–206.

[173] T. Okada, R. Shimada, Y. Sato, M. Hori, K. Yokota, M. Nakamoto, Y.-W. Chen, H. Nakamura, and S. Tamura, "Automated segmentation of the liver from 3D CT images using probabilistic atlas and multi-level statistical shape model," in *Medical Image Computing and Computer Assisted Intervention.* Springer, 2007, pp. 86–93.

[174] O. Sorkine, D. Cohen-Or, Y. Lipman, M. Alexa, C. Rössl, and H.-P. Seidel, "Laplacian surface editing," in *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing.* ACM, 2004, pp. 175–184.

[175] A. Nealen, T. Igarashi, O. Sorkine, and M. Alexa, "Laplacian mesh optimization," in *Proceedings of the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia.* ACM, 2006, pp. 381–389.

[176] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in nd images," in *International Conference on Computer Vision*, vol. 1. IEEE, 2001, pp. 105–112.

[177] A. Top, G. Hamarneh, and R. Abugharbieh, "Active learning for interactive 3D image segmentation," in *Medical Image Computing and Computer Assisted Intervention.* Springer, 2011, pp. 603–610.

[178] T. Shen, X. Huang, H. Li, E. Kim, S. Zhang, and J. Huang, "A 3D laplacian-driven parametric deformable model," in *International Conference on Computer Vision.* IEEE, 2011, pp. 279–286.

[179] H. Park, P. H. Bland, and C. R. Meyer, "Construction of an abdominal probabilistic atlas and its application in segmentation," *IEEE Transactions on Medical Imaging*, vol. 22, no. 4, pp. 483–492, 2003.

[180] W. Xiong, S. Ong, Q. Tian, G. Xu, J. Zhou, J. Liu, and S. Venkatash, "Construction of a linear unbiased diffeomorphic probabilistic liver atlas from CT images," in *IEEE International Conference on Image Processing.* IEEE, 2009, pp. 1773–1776.

[181] B. B. Avants, N. Tustison, and G. Song, "Advanced Normalization Tools (ANTS)," *Insight Journal*, 2009.

[182] S. Zhang, Y. Zhan, X. Cui, M. Gao, J. Huang, and D. Metaxas, "3D anatomical shape atlas construction using mesh quality preserved deformable models," *Computer Vision and Image Understanding*, vol. 117, no. 9, pp. 1061–1071, 2013.

[183] C. Goodall, "Procrustes methods in the statistical analysis of shape," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 285–339, 1991.

[184] D. Shen and C. Davatzikos, "An adaptive-focus deformable model using statistical and geometric information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 906–913, 2000.

[185] B. Li and S. T. Acton, "Active contour external force using vector field convolution for image segmentation," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2096–2106, 2007.

[186] N. Aspert, D. Santa-Cruz, and T. Ebrahimi, "Mesh: Measuring errors between surfaces using the hausdorff distance," in *IEEE International Conference on Multimedia and Expo (ICME)*, vol. 1. IEEE, 2002, pp. 705–708.

[187] C. Platero, V. Rodrigo, M. C. Tobar, J. Sanguino, O. Velasco, and J. M. Poncela, "Probabilistic atlas based segmentation using affine moment descriptors and graph-cuts," in *Computer Analysis of Images and Patterns.* Springer, 2011, pp. 144–151.

[188] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient nd image segmentation," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109–131, 2006.

[189] Z. Yan, S. Zhang, C. Tan, H. Qin, B. Belaroussi, H. J. Yu, C. Miller, and D. N. Metaxas, "Atlas-based liver segmentation and hepatic fat-fraction assessment for clinical trials," *Computerized Medical Imaging and Graphics*, vol. 41, pp. 80–92, 2015.

[190] C. Tan, Z. Yan, K. Li, D. Metaxas, and S. Zhang, "Laplacian shape editing with local patch based force field for interactive segmentation," in *International Workshop on Patch-based Techniques in Medical Imaging.* Springer, 2015, pp. 95–103.

[191] G. Langs, A. Hanbury, B. Menze, and H. Müller, "Visceral: Towards large data in medical imaging – challenges and directions," in *MICCAI International Workshop on Medical Content-Based Retrieval for Clinical Decision Support.* Springer, 2012, pp. 92–98.

[192] X. Zhang, F. Xing, H. Su, L. Yang, and S. Zhang, "High-throughput histopathological image analysis via robust cell segmentation and hashing," *Medical Image Analysis*, vol. 26, no. 1, pp. 306–315, 2015.

[193] X. Zhang, H. Su, L. Yang, and S. Zhang, "Fine-grained histopathological image analysis via robust segmentation and large-scale retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5361–5368.

[194] S. Zhang and D. Metaxas, "Large-scale medical image analytics: Recent methodologies, applications and future directions," *Medical Image Analysis*, vol. 33, pp. 98–101, 2016.

[195] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.

[196] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *International Conference on Machine Learning*, 2011, pp. 689–696.

[197] A. Karpathy and F.-F. Li, "Deep visual-semantic alignments for generating image descriptions," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3128–3137.

[198] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, 2015.

[199] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3D shapenets: A deep representation for volumetric shapes," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1912–1920.

[200] Y. Fang, J. Xie, G. Dai, M. Wang, F. Zhu, T. Xu, and E. Wong, "3D Deep Shape Descriptor," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2319–2328.

[201] C. Tan, K. Li, Z. Yan, D. Yang, S. Zhang, H. J. Yu, K. Engelke, C. Miller, and D. Metaxas, "A detection-driven and sparsity-constrained deformable model for fascia lata labeling and thigh inter-muscular adipose quantification," *Computer Vision and Image Understanding*, 2016.

[202] N. Moon, E. Bullitt, K. Van Leemput, and G. Gerig, "Automatic brain and tumor segmentation," in *Medical Image Computing and Computer Assisted Intervention*. Springer, 2002, pp. 372–379.

[203] S. Doyle, F. Vasseur, M. Dojat, and F. Forbes, "Fully automatic brain tumor segmentation from multiple mr sequences using hidden markov fields and variational em," *Procs. NCI-MICCAI BraTS*, pp. 18–22, 2013.