Study on Statistical Method for Segmentation of Multi-object on Abdominal CT Images

year	2022-02
	CT
	3
	17104 546
URL	http://hdl.handle.net/10228/00008906

Doctoral Thesis

Study on Statistical Method for Segmentation of Multi-object on Abdominal CT Images

By

Jiaqi Wu

195F2002

A thesis submitted in partial

fulfillment of the requirements for the degree of

Doctor of Engineering

Professor Tohru Kamiya

Chairperson of Supervisory Committee

Department of Engineering

Kyushu Institute of Technology, Japan

February, 2022

Abstract

Computer aided diagnosis (CAD) is the use of a computer-generated output as an auxiliary tool for the assistance of efficient interpretation and accurate diagnosis. Medical image segmentation has an essential role in CAD in clinical applications. Generally, the task of medical image segmentation involves multiple objects, such as organs or diffused tumor regions. Moreover, it is very unfavorable to segment these regions from abdominal Computed Tomography (CT) images because of the overlap in intensity and variability in position and shape of soft tissues. In this thesis, a progressive segmentation framework is proposed to extract liver and tumor regions from CT images more efficiently, which includes the steps of multiple organs coarse segmentation, fine segmentation, and liver tumors segmentation.

Benefit from the previous knowledge of the shape and its deformation, the Statistical shape model (SSM) method is firstly utilized to segment multiple organs regions robustly. In the process of building an SSM, the correspondence of landmarks is crucial to the quality of the model. To generate a more representative prototype of organ surface, a k-mean clustering method is proposed. The quality of the SSMs, which is measured by generalization ability, specificity, and compactness, was improved. We furtherly extend the shapes correspondence to multiple objects. A non-rigid iterative closest point surface registration process is proposed to seek more properly corresponded landmarks across the multi-organ surfaces. The accuracy of surface registration was improved as well as the model quality. Moreover, to localize the abdominal organs simultaneously, we proposed a random forest regressor cooperating intensity features to predict the position of multiple organs in the CT image. The regions of the organs are substantially restrained using the trained shape models. The accuracy of coarse segmentation using SSMs was increased by the initial information of organ positions.

Consequently, a pixel-wise segmentation using the classification of supervoxels is applied for the fine segmentation of multiple organs. The intensity and spatial features are extracted from each supervoxels and classified by a trained random forest. The boundary of the supervoxels is closer to the real organs than the previous coarse segmentation.

Finally, we developed a hybrid framework for liver tumor segmentation in multiphase images. To deal with these issues of distinguishing and delineating tumor regions and peripheral tissues, this task is accomplished in two steps: a cascade region-based convolutional neural network (R-CNN) with a refined head is trained to locate the bounding boxes that contain tumors, and a phase-sensitive noise filtering is introduced to refine the following segmentation of tumor regions conducted by a level-set-based framework. The results of tumor detection show the adjacent tumors are successfully separated by the improved cascaded R-CNN. The accuracy of tumor segmentation is also improved by our proposed method.

26 cases of multi-phase CT images were used to validate our proposed method for the segmentation of liver tumors. The average precision and recall rates for tumor detection are 76.8% and 84.4%, respectively. The intersection over union, true positive rate, and false positive rate for tumor segmentation are 72.7%, 76.2%, and 4.75%, respectively.

Acknowledgments

First of all, I would like to thank my supervisor, Prof. Tohru Kamiya, for his invaluable advice and continued guidance throughout this research with my deepest respect. Your profound knowledge, experiences, and ideas have benefited me a lot in my path to scientific research. Your patience and encouragement have always inspired my confidence and strength to continue. I would like to thank Prof. Shuichi Kurogi, Prof. Joo Kooi Tan, Prof. Lifeng Zhang, and Prof. Huimin Lu, who have provided valuable advice and suggestions for my research.

My endless appreciation and gratitude go to my supervisor at Tiangong University, Dr. Guanxu Li, who has led me on the path of academic research. Thank you for your guidance and support and for all of the opportunities I was given to further my research. I would especially like to thank Prof. Philip O. Ogunbona at the University of Wollongong for his advice and guidance on academic research methodology.

I am grateful to the old boys in the Kamiya laboratory and friends in Kyutech, for their kind help and advice during my time here. I would also thank the old friends in my homeland, for their warm encouragement and care from thousands of miles away.

Finally, I would like to thank my family, for always being my home, and for their selfless love and unwavering faith in me, which taught me to be optimistic and gave me the strength and support to pursue my dreams.

Dedication

In this thesis, we propose several statistical methods for the problem of multi-object segmentation in medical images. The dedication is summarized below:

1. We present an architecture of statistical shape models (SSMs) building and its application to automatic organ segmentation from medical images. By using the deformation of shapes learned from labeled data of human organs, the model can delineate the contours of object organs from new images. The fundamental shape, size, and position information of organs, which is necessary to the computer-aided-diagnosis systems, can be provided by our proposed methods.

2. In the process of constructing single organ SSMs, we introduce a k-means clustering method to build a standard reference surface which is more representative. The quality of the SSMs built from the landmarks which are corresponded by the reference surface are increased.

3. Simultaneously landmarks correspondence of surfaces containing multiple organs is more complex than single organ, as the spherical-mapping-based methods are invalid. For this purpose, we introduce a non-rigid iterative closest point surface registration process to seek more properly corresponded landmarks across the multi-organ surface meshes.

4. In the process of organ segmentation by using SSMs, an initial position of the mean shape is required. We propose a random forest (RF) regression model to find the candidate position to initialize the SSMs, which increase the accuracy of the segmentation results.

5. For the multi-organ segmentation problem, we also propose a novel method based on supervoxel. The spatial information is introduced in the process of supervoxel classification by the RF classifier. This method provides smoother labeling of multiple organs, which can be furtherly introduced and fused to SSMs.

6. In liver tumor detection and segmentation, adjacent tumors are easily misrecognized as one tumor, which decreases the accuracy of the diagnosis. In this thesis, a region-based convolutional neural network is introduced to detect and separate the tumors in the liver regions from multi-phase computed tomography images with better performance.

Table of Contents

List of tables

List of figures

Chapter 1

Introduction

1.1 Computer Aided Diagnosis Technology

Since the invention of the X-ray by Wilhelm Röntgen in 1895, medical imaging techniques have been greatly developed and favored many medical applications, including improved screening, diagnosis, monitoring of disease [1]. Different modalities of medical images are generated when the human body is examined by various forms of imaging techniques. According to the procedure of energy that produced or detected from the human body, the medical images can be categorized into many different modalities, including endoscopy, microscopy, X-ray, computed tomography (CT), magnetic resonance imaging, positron emission tomography, amplification of infrared, ultrasound and so on. Different structures of the human body can be presented differently in different modalities, and thus the physicians tend to choose suitable modalities so that the organs and tissues in the image can be easily distinguished, for specific patients and diseases. Among the varies modalities of medical images, the CT images, which provide three-dimensional (3-D) information of high quality in a short time, are of important reference for the medical diagnosis of many diseases.

However, some of the tissues and lesions cannot be easily distinguished due to their imaging features being similar to ordinary CT images. To solve this problem, radiocontrast agents are introduced in the imaging examination. These substances can absorb external radiation and enhance the tissues (vessels) where the agents exist. Multi-phase CT angiography provided the physicians with multi-degree information as well as tissues difference from the temporal resolutions [2]. A series of phase signals are
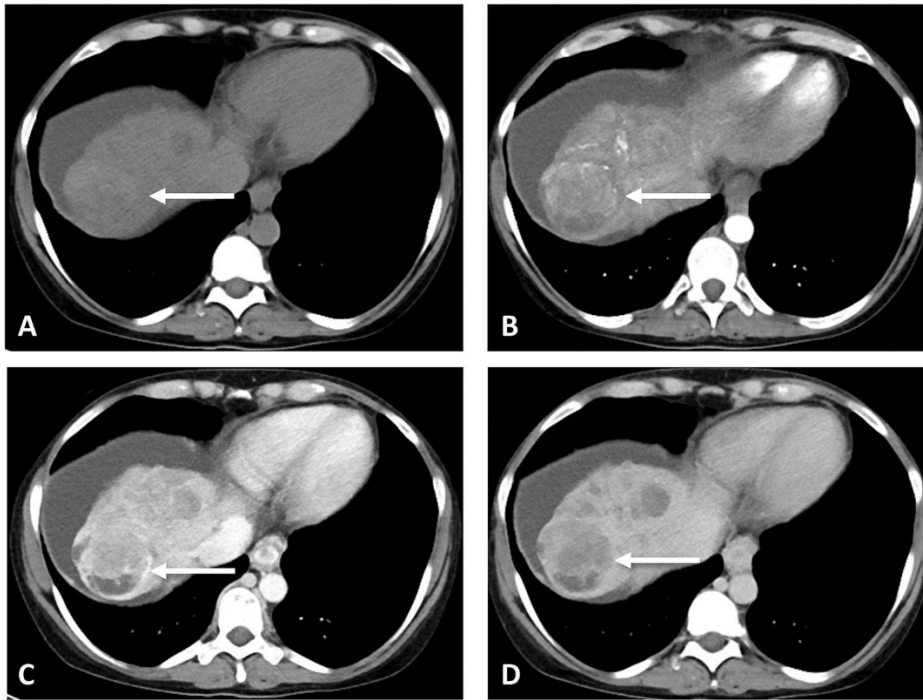
Figure 1.1 Appearance of masses in different phases of CT images. The masses in the images of four phases, i.e., the (A) NC, (B) ART, (C) PV and (D) DL, show different appearance [7].

obtained in turn after contrast injection, which include: non-contrast-enhanced (NC), arterial (ART) phase, portal venous (PV) phase, and delay (DL) phase [3]. These variations from multi-phase CT images provide a novel view for radiologists to distinguish liver tumors [4-6]. The Figure 1.1 illustrated the difference of appearance of several masses in difference phases of CT images.

The diagnosis, if simplified as a process to match the symptoms of the patients with the corresponding disease, has an expanding searching space. The symptoms, i.e., the current and historical information collected from the patients, are also boosting. The development of new medical examination means may create more data, including images of higher quality and quantity. Also, the knowledge of human anatomy and physiology has increased significantly as well as the discovered diseases. And the diagnostic rules of matching the symptom with the disease are more complicated than before.

To overcome these problems, the computer-aided diagnosis (CAD) systems, which aim at providing the physicians with a 'second opinion' before the final medical decisions [9], have developed into valuable tools in modern clinical medicine. The tedious examination tasks for the image data can be programmed and automated. The computer algorithms can be used to dig the deeper relevance between the complex process of diagnosis by introducing new techniques in computer science, like artificial intelligence, pattern recognition, data mining, computer vision, and machine learning. This can also eliminate the subjective factors of the diagnosis, as the clinical diagnosis is a practical science and relies on the individual experience of the physicians.

In the application of cancer diagnosis using medical images, like CT or MRI, the CAD can be grouped into two basic areas: computer-aided detection (CADe) and computer-aided diagnosis (CADx) [10]. The CADe focuses on finding the location of lesions from the images while the CADx performs the characterization of the lesions, for example, the distinction between benign and malignant tumors.

The concept of CAD, in its process of developing and applying, has derived a series of alternative terms or separate branches since the 1960s [11-13]. In 1998, the first FDA (The United States Food and Drug Administration) approved CAD product, the ImageCheckerM1000® by R2 Technology, Inc. (Los Altos, CA), was approved by premarket application (PMA) approval (P970058) [14]. The labeling of the product was initially for use on routine screening mammograms, but on May 29, 2001, the approval was granted for the expansion of the "Indications for Use" to cover diagnostic as well as screening mammograms. Since the success of ImageChecker1000®, CAD systems developed by the iCAD, Inc. (Nashua, NH) and the Eastman Kodak's Health Group (Care- stream Health Inc. since 2007) also obtained FDA approval for mammography in 2002 and 2004, respectively [9].

In this thesis, we introduced a CAD system for diagnosis of tumors in abdominal organs. The main aim of the system is to provide the physicians with regions of organs as

well as the regions of tumors which are precisely delineated in a given CT image of the patient. The intensity, shape, and position of the labeled organs and tumors can provide necessary information of the patient for the diagnosis. Precise analysis of lesions also relies on the region of interest (ROI) from the organ region segmented from the entire image. The information of tumors is further analyzed by physicians for final diagnosis and treatment. Besides, these fundamental data of the organ and tumors can also be used in the planning and navigation of surgeries.

1.2 Medical Images Segmentation

1.2.1 The Segmentation Problem

The image segmentation problem, also referred as delineation or extraction, is a fundamental task in image processing. The subjects of segmentation are generally two-dimensional (2-D) or 3-D digital medical images, which are composed of basic pixels or voxels, respectively. The intensity of the pixels or voxels reflect the physical features of corresponding human structures under certain medical imaging equipment. In the semantic segmentation, the pixels or voxels of the medical images are supposed to belongs to different segments, including background and different human structures. This process can also be regarded as assigning one of the known labels to each pixel in the digital image.

1.2.2 Challenges of Medical Image Segmentation

In the early stage, the segmentation of human organs or structures relies on the manual delineation of the images slice by slice. However, for each patient, a series of CT images contain hundreds of slices, which makes manual segmentation a tedious and time-consuming work. To reduce the burden of physicians, automatic segmentation of desired human organs and structures becomes a meaningful and essential issue.

Compared with general image segmentation problems, the task of medical image segmentation faces more difficulties and challenges due to the natural influence of

medical images. Among them, the characteristics of medical images that affect the results of medical image segmentation mainly come from two aspects, namely, the imaging process and the human body structure.

Limited by the characteristics of the current imaging technology, there is a loss of information between the collected medical images and the real conditions of the human body. Different organs and tissues may be indistinguishable by the imaging equipment. Also, different types of noises such as thermal noise, interference noise, quantization noise, and light are possibly introduced in the equipment imaging process. In addition, the distortion may also be caused by the movement of internal tissues and organs, such as human breathing, heartbeat, and blood flow during the acquisition process, and even the influence of external body position movement. These effects lead to the general vagueness and unevenness of medical images.

The structure and distribution of human tissues and organs are very complex. Most of the organs are variable in shape. The boundaries between different tissues or lesions are blurred and of low contrast. Moreover, the anatomy of different individuals is quite different. Even for the same individuals, their organs and tissues may behave differently at different imaging times.

These factors have made organ segmentation from medical images a difficult problem. So far, there is no recognized and universal method for organ segmentation. Therefore, the automatic segmentation method of organs in medical images is still a challenging but essential point for CAD systems as well as many other medical applications.

1.3 Aims and Objectives

The main objective of our work is to build a CAD system for tumor diagnosis, which are mainly realized by multi-organ segmentation using statistical shape models (SSMs) and tumor detection and segmentation using the region based convolutional neural network (R-CNN) and level set method. In details,

- Localize the position of multiple organs from an abdominal CT image by a random forest regressor which is trained using intensity features.
- Build SSMs with improved quality by introducing an optimized landmark corresponding process.
- Extend the object of the SSMs to multiple organs and introduce a non-rigid surface registration method to improve the accuracy of landmarks correspondence in the building process.
- A fast and coarse segmentation of multiple organs by using the SSMs.
- Accurate segmentation of multiple organs by supervoxel clustering and random forest classifier to refine the result from the SSMs.
- Detect and distinguish different tumors from the obtained regions of the organs by R-CNN.
- Precise segmentation of tumors from the candidate region detected.

1.4 Outline of the Article

This thesis is organized as follows (the main flowchart is shown in Figure 1.2.

In chapter 2, we make a series of literature surveys on our following research. We categorize the image segmentation into non-statistical approaches and statistical approaches.

In chapter 3, we describe a fundamental process of SSMs building problem, which is prepared for the construction of SMMs. A novel landmark corresponding method based on k-means clustering and Demons registration is proposed to train 3-D SSMs with higher quality. We also introduced a non-rigid iterative closest point (NICP) surface registration method to seek proper corresponded landmarks across the multi-organ surface meshes.

In chapter 4, we developed the multi-organ segmentation methods. In detail, a random forest (RF) regression model is trained to find the candidate position of organs. After

obtaining the location of the organs, an efficient and robust segmentation of multiple organs is realized by the searching and deformation of SSMs. To refine the coarse segmentation by the SSM, we also proposed a supervoxel-based segmentation method using spatial information in the CT images. The supervoxels with boundaries adjacent to anatomical edges are separated from the image by using the simple linear iterative clustering (SLIC) from the images. Then an RF classifier is built to predict the labels of the supervoxels according to their spatial and intensity features. The region of organs obtained from the segmentation provide the basic region for the following analysis of tumor detection and extraction.

In chapter 5, a hybrid framework is proposed for liver tumor segmentation in multiphase images. We first develop a cascade region-based convolutional neural network with a refined head to locate the tumors with higher accuracy, especially for the cases with adjacent tumors, which are difficult to distinguish by conventional tumors detection approaches. Meanwhile, a phase-sensitive noise filter is introduced to refine the segmentation conducted by a level-set-based framework. The final accuracy of tumor segmentation was improved.

In chapter 6, we conclude this thesis and make a planning for future works.
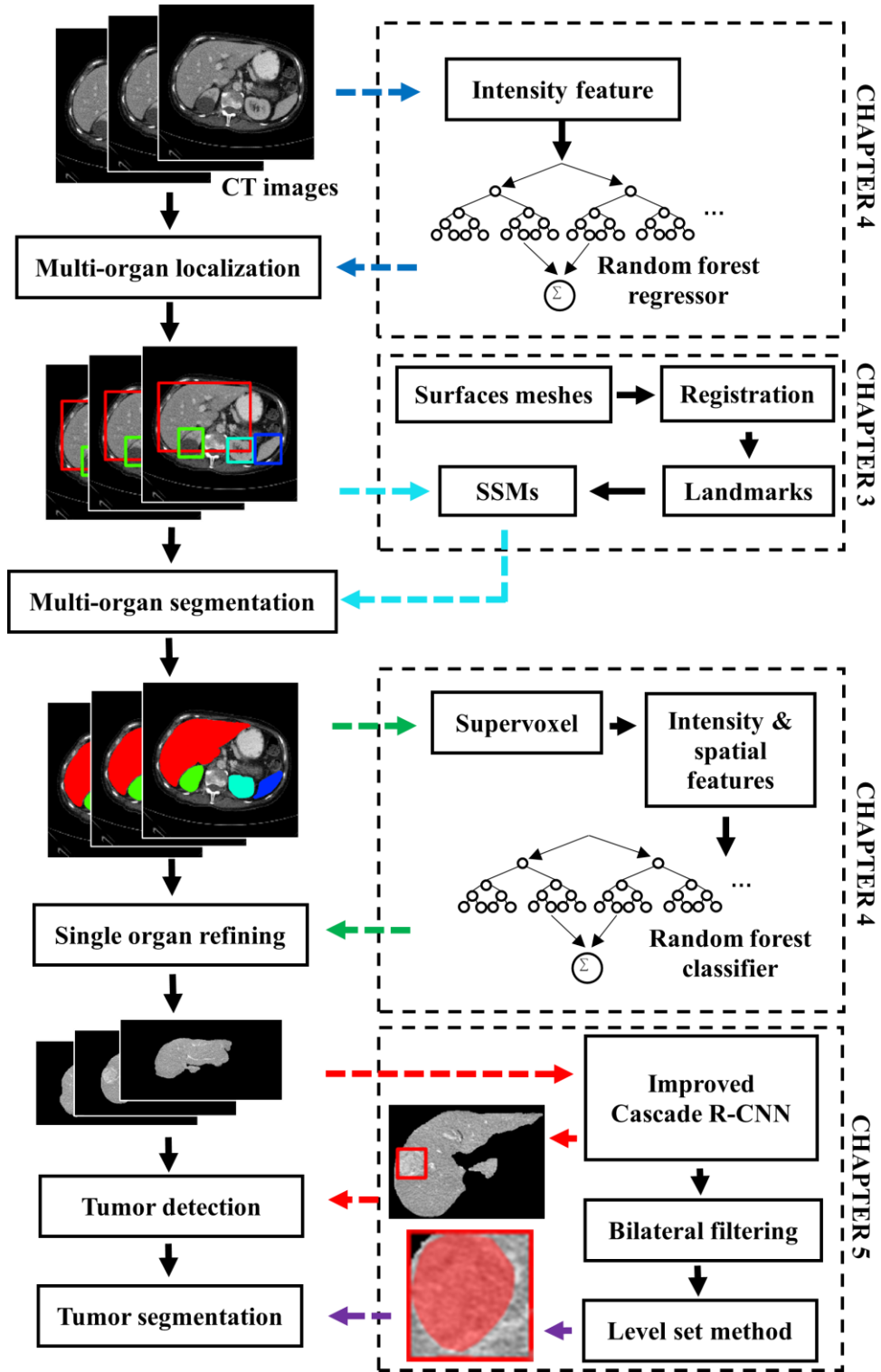
Figure 1.2 Flow chart depicting the overall workflow of the proposed CAD system.

Chapter 2

Literature Review

In this chapter, we review relevant methods and approaches of the multi-object segmentation tasks. The fundamental research of medical image segmentation problems is firstly provided in section 2.1. We discuss non-statistical approaches and statistics approaches in section 2.1.1 and section 2.1.2, respectively.

Among these methods, the approaches of SSMs are of high accuracy and robustness owing to the modeling of shape and modes of deformation for certain organs. Section 2.2 provides methods relevant to the problem of SSMs building. The basic representation methods are introduced in section 2.2.1. In section 2.2.2, approaches of landmarks correspondence, which is a key process in model building, are discussed. The localization problems for initializing the SSMs in medical images are also important are reviewed in section 2.2.3.

Furthermore, we focus on the specific problem of multi-organs segmentation in medical images. We introduced the conventional approaches in section 2.3.1 and the recently developed deep-learning-based methods in section 2.3.2.

2.1 Statistical Image Segmentation Methods

2.1.1 Non-statistical Approaches

(1) Thresholding

Threshold-based methods [15-16] usually use the gray histogram of the image to find the most suitable threshold and extract the region of interest. This method is suitable for images with similar gray levels of pixels in the target area but large differences from the

background gray level.

(2) Texture

The texture-based method [17-18] divides the image into regions with different texture properties, and the regions of interest can be discriminated from the background.

(3) Region growing

The method based on Region Growing [19-20] combines the local relative position information of the voxel. The seed point and the conditions which determine whether the seed point and the surrounding pixels belong to the same area are initially set. And in the following iterative process, the neighboring pixels that meet the conditions are successively added. It takes simple calculation for small structures, but it is not suitable for overlapping area, and manually selection of seed point is necessary.

(4) Watershed

The method based on Watershed [21-22] is a based on mathematical morphology. In the process of segmentation, the watershed basin can be found first, and then the watershed edge can be found through the complement operation. It is simple to implement, intuitive, and can be operated in parallel. However, since this method is also sensitive to noise, it is easy to produce over-segmentation results.

(5) Edge detection

The method based on Edge Detection [23-28] uses the feature of large changes in the gray level difference between adjacent tissue regions and solves the segmentation problem by calculating the gradient features of the entire image to detect the boundaries of different regions. Commonly used operators include gradient operator, Sobel operator, Roberts operator, Prewitt operator, Laplace operator, Kirsch operator, Canny operator, LOG operator, etc.

(6) Filters

Filter-based methods [29-30] mainly implement image denoising and target

enhancement through operators. Among them, the Hessian matrix based on the image scale-space theory is a classic filtering method. The detection result of the structure is often used for the segmentation of the blood vessel target in the image.

2.1.2 Statistical Approaches

The segmentation techniques which are based on statistics are classified into two main categories: the surface-oriented approaches and the region-oriented approaches. The surface-oriented approaches focus on finding a relatively smooth contour of the object and the pixels or voxels inside of the contours are regarded as the segmentation results. The region-oriented approaches tend to distinguish the region of interest directly.

2.1.2.1 Surface-oriented approaches

(1) Statistical models

Statistical methods [31-33] regard the gray level of pixels in the image as a statistical model that conforms to a certain probability distribution and regard the segmentation of the image as a problem of obtaining the combination of regions in the image with the largest probability. According to Bayes' theorem, it is to find the distribution problem with the largest posterior probability from the combination of adjacent voxels. In actual application, the mixed distribution method is often used to construct a statistical model containing multiple distribution functions. The gray distribution histogram curve of voxels or pixels is approximated through the combination of different types and different numbers of distribution functions in statistics. The neighborhood of the advanced random process of Maximum a posteriori estimation (MAP) Constraints can exclude the noise.

Wilson and Noble used statistical models to achieve blood vessel segmentation in angiography (magnetic resonance angiography, MRA) [31], but the segmentation results using the maximum likelihood method in the early days are often affected by factors such as noise. Hassouna et al. proposed a method of simultaneously using low-level random and high-level random processes to establish a gray distribution histogram curve [32]. Among them, the low-level random process includes a limited number and types of

distribution function combination models to express the gray distribution of voxels, and the high-level random process introduces Markov Random Field (MRF), using its neighborhood Constraints to suppress noise [33].

Since the statistical model relies on the feature distribution function and is affected by the nature of the gray distribution histogram of the pixel, the random field is very sensitive to the boundary. Therefore, the establishment of the distribution function, the automatic selection of the neighborhood structure, and the parameter estimation of the MRF-MAP model remain current problems. Segmentation methods based on statistical models only use the statistical information of image pixel gray levels, and it is still difficult to obtain ideal segmentation results for images where the gray levels of segmentation target and background have intersections.

 (2) Atlas

The atlas-based methods [34-36] are based on the registration between the general atlas and the target. The voxel atlas contains grayscale images of templates and segmented images with marks of regions and contours. After the registration of the template and the target image is completed, the target area can be found on the image through the marks on the atlas.

The atlas-based method is usually affected by two factors, namely the method of constructing the atlas and the method of registration. First, the construction of the atlas needs to consider the application segmentation object structure, shape, appearance and other differences. If the difference between the objects is too large, different cases need to be grouped, and multiple atlases are required to adapt to different situations. Another factor is the registration of the atlas and the image. Affine registration methods such as rigid or elastic, demons, nonlinearity, deformation vector, B-spline are usually used [34]. Since the atlas-based method contains a large amount of global prior information, it can reduce the over-segmentation of the region, and directly map the segmentation result to the target region to obtain more (or even other organization) structural information.

12

However, due to the source of the voxel atlas and the specific case, the representation of the atlas is poor, and the segmentation effect depends heavily on the accuracy of the registration.

(3) Local prior shape

The segmentation method based on the local prior shape [37-39] uses the local prior features of the target area to realize the description of the smoothness of the surface and interior of the segmented target part.

In 1987, Kass [37] proposed the Snakes model. This method uses some marked points with the ability to represent shapes to form a continuous curve that is connected end to end as a template and designs an energy function that characterizes the elastic deformation of the contour of the curve. By controlling the local features of the image, the parameters of the curve are matched to reach a harmonious state, and the image segmentation process is transformed into a process of solving the minimum energy functional. The position of the corresponding contour when the final energy is at the minimum becomes the segmentation result of the target area. There are two forces involved in the iterative deformation process of the curve, one is derived from the internal force of the curve itself, which is used to constrain the contour of the model and keep it smooth, and the other is derived from image features (grayscale, texture, edge, etc.) The external force is used to adjust the relationship between the contour and the image to make it constantly approach the boundary of the target area. Under the action of the two forces, the curve contour keeps approaching the boundary of the target area, and the iteration stops when the energy functional reaches the minimum.

The advantage of this approach is that, by setting the internal energy parameters, the Snakes model can effectively avoid the influence of noise or weak edges. However, this method has higher requirements for the initial position. If the initial position is far away from the segmentation target, it is easy to fall into a local minimum and the segmentation fails. In addition, because the shape parameter involves all the marking points on the

contour, it may not be possible. When adapted to the area covered by a certain part of the mark on the contour, the segmentation may result in leakage to the adjacent area.

Because the Snakes model makes good use of the local features of the image, and at the same time uses the curve to approximate the target contour, it effectively integrates the upper-level knowledge to obtain the overall contour of the target. It has a far-reaching impact on computer vision, and not only inspired many improvements in the construction and solution methods based on energy functions, but it has also derived many new contour-based models.

(4) Level set

The level set-based method [40-43] can be regarded as a comprehensive application of the Snakes method in the field of 3-D image segmentation. The basic principle of the level set method is: for a closed curve on a plane, a higher-dimensional level point set including the same function value is used to express the curve, and a zero level set is embedded in this high-dimensional level set function. According to the constant iterative evolution of the closed surface equation, the equation corresponding to the ever-changing level set function is obtained. When the evolution becomes stable, the embedded closed surface is finally determined, that is, the final evolution result of the mobile interface is obtained through the determination of the zero level set.

The advantage of the level set method is that because it introduces an equation that is one dimension higher than the segmented image, it will not cause changes in the topological structure of the surface, so that it has strong low-dimensional topological variability, and it can be natural in the evolution process. The topological structure of the curve is changed to increase the expression ability of the segmentation result, and it is easy to expand to higher-dimensional applications. Similar to the shortcomings of the Snakes method, the segmentation result may have boundary leakage. In addition, the selection of the velocity field parameters used for the evolution of the segmentation of multiple types of images and the lower segmentation speed are also disadvantages.

(5) Shape model

Among the methods based on the shape models, the most typical is the SSM method [44-53], including the active shape model (ASM) and active appearance model (AAM). The basic idea of the method is to first establish a training set containing multiple different and similar target shapes for the target to be segmented and construct an average model that includes the variation range of each marked point that composes the shape through a statistical method. Then the average model is placed near the target area of the image, and the shape position and posture are adjusted according to the image features around the marker points, and the segmentation target area is found.

Other methods based on shape models are somewhat clouded by polygon surface method, single network method, B-spline expression method, geometric network method, and finite element method [54-56]. The basic idea is similar to the SSM method. All are achieved by the method of adaptively adapting the average shape extracted from the original training set and training in the segmented image. Because the shape model method contains prior knowledge of the shape of all structures, it can prevent the segmentation results from entering the adjacent tissue structure. It has a good segmentation effect for targets with complex shapes but limited overall contour changes; due to the existence of the shape change range Due to certain constraints, the segmentation effect for objects without specific contours is poor. These characteristics make the shape-based model method suitable for organ segmentation of medical images. The SSM method obtains accurate and stable segmentation results through statistical learning of the shape and surface features of the organ samples in the training set, which is suitable for medical image segmentation tasks.

2.1.2.2 Region-oriented approaches

Methods based on Artificial Neural Network (ANN) [57-63] usually regard the image segmentation as a problem of classifying image voxels, and each voxel in the segmented image is judged as a target or background through a classifier. In the field of medical image segmentation, deep learning methods such as Back Propagation Neural Network,

Convolutional Neural Network (CNN), and Fully Convolutional Networks (FCN) are commonly used.

The so-called neural network, imitating the principle of neuron signal transmission in the human brain, is a network composed of multi-layer neuron models. For the input signal, the neurons on each layer will be derived from the previous layer of neurons according to the weight parameters obtained by training. The output signal of the element makes two responses of 'excitement' and 'inhibition' and is transmitted to the next layer of the network until the final output result is obtained.

In the convolutional network [59] that uses a deep network structure, the characteristics of the target are learned by extracting more advanced features layer by layer through pixel convolution and pooling of the input image. In the network training process, it is necessary to input the manually marked images into the network, and design the error function, use the error between the current network result and the marked correct result to adjust the network parameters, so that the network automatically learns the features of the segmentation target in the image. The quality of the network segmentation effect is affected by the network structure (neuron connection method, network layer number, etc.), network parameter training method, and training set data.

In practical applications, the neural network is an effective image segmentation method. Hinton et al. used the CNN method [59] to win the championship in ImageNet 2012 with an accuracy rate exceeding second place by 10%. Because the neural network has a good adaptive ability, it has good stability against interference factors such as noise. Although the deep learning method has a good segmentation effect, it lacks a complete theoretical system and cannot give a complete and self-consistent explanation from a mathematical perspective. The selection, parameter adjustment, and optimization of the network model rely more on manual experience and skills. In addition, the training cost of the deep learning model is high, a large amount of manually labeled training data is required, and the training is time-consuming.

2.2 Statistical Shape Model

2.2.1 Representation of Shape

The original training data set used for the SSM is usually the extracted voxel image. For different SSMs, they can be expressed by binary voxel data, fuzzy voxel data, or shape surface nets. Generally, all model expression methods can be transformed into each other [42]. The choice of the shape model expression method has an important influence on the subsequent model construction and the model-based segmentation process. Many of these methods will be limited by the model expression method.

The voxel-based expression method is a method that uses a regular grid to represent the volume. McInerney and Terzopoulos proposed a physics-based anatomical surface segmentation and reconstruction method [64]. Although medical image data sets are composed of anisotropic shapes, in traditional medical image analysis and diagnosis, doctors usually need to check 2-D image slices in sequence. To obtain an intuitive 3-D image, the process of model building and image reconstruction usually uses image isotropic means to down-sample the high-resolution original image and super-sample the inter-layer resolution to overcome the regular rectangular pixels. Shape and unify different data sets.

The method based on the point distribution model (PDM) is a simple and commonly used method of model expression. This method uses the statistical expression of the spatial distribution of the landmarks that constitute the surface of the shape to express the changing relationship of each shape in the training set [44]. First, it is necessary to randomly select a certain number of different similar target shapes to form a training set. In the PDM method, the points that can reflect the shape and surface characteristics of the training set are called marked points. Usually, statistical analysis is used to extract the changing modes from the marked points of multiple shapes in the training set, to obtain a compact model that can change within a certain constraint range.

17

The method based on Spherical Harmonics (SPHARM) uses a system of equations to express the topological structure of a closed spherical surface. Through a set of description functions defined in the spherical coordinate system, a set of orthonormal basis is constructed in the spherical domain. The harmonic function can express shape data with few coefficients and has a wide range of applications in computer graphics. It was introduced by Szekely and others in the image segmentation method based on the deformation model [65].

The method based on the centerline (m-rep) or skeleton line representation [66] uses the centerline of the shape and the boundary vector pointing to the object to express the shape model. This method is usually more concise than the marked points and was first used to describe the shape of organisms.

Compared with the other three expression methods, the PDM has a simple structure, better shape description ability, and versatility. Therefore, the PDM method is used to express the SSMs in this thesis.

2.2.2 Correspondence of Marker Points

In the PDM, the marked points form the basis of the shape model and represent the characteristics of the shape and contour of the object. The corresponding problem of marked points, i.e., finding a set of marked points with the same characteristics from the surface of each training sample shape included in the training set is a key step in the construction of the shape model, which has an important impact on the subsequent use of the model. Correspondence of marker points generally includes marker point selection, marking and registration. In the original PDM method [44], the selection, marking and registration of marker points are all realized simultaneously by manual marking.

In addition to marking the position of the mark point, the selection of the mark point, that is, the selection of the characteristic points with the shape expression ability from the original shape contour, also has an important influence on the construction of the model.

In the field of medical images, there are specific rules for the selection of marker points. The characteristic points that reflect the contour shape of organ tissues are usually divided into the following three categories:

- Feature points with anatomical significance: usually annotated by a medical doctor, reflecting the biological characteristics of the model.

- Feature points with mathematical significance: have obvious mathematical features (geometric, gray or gradient features).

- Pseudo-feature points: In order to keep the contour of the shape continuous and complete, the points inserted between the first two types of points to complete the shape.

In general, the selection of marker points needs to follow the following principles:

- For feature points with anatomical and mathematical significance, because they contain important information, they are usually selected as marker points.

- The selection of marking points should be distributed as evenly as possible.

- Control the overall density of marker points. Too high will affect the efficiency of the algorithm, while too low will affect the accuracy of the model and the ability to express the shape.

Early PDMs are mainly applied to 2-D targets, and the correspondence of marked points can usually be completed manually. However, due to the complicated surface shape of human organs, at least hundreds of markers are needed for expression. At the same time, the original 3-D shape has a high dimensionality, and the number of vertices on the surface is huge. It is difficult to find the correct position of the mark point. Manually obtaining the mark point is time-consuming and labor-intensive, and errors are prone to occur. Therefore, for a 3-D shape surface, an automatic method is usually used

to complete the correspondence of the marked points. In the process of automatic correspondence of shape markers, a registration algorithm needs to be used. According to the different expressions of the two parties used for registration, the registration modes can be roughly divided into the following four categories [52].

(1) Mesh-to-mesh

The method based on mesh-to-mesh registration is the most direct method of marking points. First, the shape surface is extracted from all the training data, and then the reference shape surface containing the mark points is used to find the corresponding mark points on the shape surface of the training set.

The Iterative Closet Point (ICP) method [67] is a typical point registration algorithm based on the surface shape features of the object. Essentially it is a least squares method based on optimal matching, which can be used for the registration of surface shapes containing different numbers of vertices and can transfer the optimal similarity transformation of one surface to another surface. By calculating the coincident point set obtained by the rigid body transformation of the target point set and the point set to be registered, the optimal rigid body transformation method is found, and the corresponding relationship between the points in the two shape point sets is obtained. This method is suitable for the registration of marked points between surfaces with similar shapes, such as bones and similar tissues with fixed shapes. However, the registration effect of the shape surface with a large change or a non-rigid body transformation relationship is poor, and it cannot cover the range of its shape change. In this case, if only using this method to determine the corresponding marker points, not only will the corresponding relationship of the marker points with obvious errors be obtained, but also the triangles in the non-isomorphic mapping surface net may be flipped.

(2) Mesh-to-volume

The mesh-to-volume registration method based on the network is mainly applied to the volume data type obtained through segmentation.

Firstly, the deformable surface net containing the mark points is adapted to the binary or gray-scale volume data, and the corresponding relationship is determined by the final position of the mark points in the image. The advantage of this type of method is that it does not require prior manual segmentation, but it has structural limitations and can only be used for the types of structures that can be segmented by the template, such as bone structures. In the mesh-to-volume method, if the template does not fold itself during the process of adapting to volume data targets, the homeomorphic mapping of the input shape can be well guaranteed.

(3) Volume-to-volume

The volume-to-volume registration method uses a reference voxel map containing the marker points to find the marker points in the volume data. First, a voxel map is registered to all binary or gray-level volume data, and then the marker point reference template is mapped to the volume data through the variable voxel map to realize the correspondence of the marker points.

The result of the method of processing the original gray volume data mainly depends on the type of the object to be modeled and the quality of the input image. Generally, the higher the similarity of the training samples in the original training set, the easier it is to extract the region of interest from the background, and the more accurate the final SSM will be.

(4) Parameter-to-parameter

The parameter-to-parameter registration method transforms the corresponding process of the marked points into the spherical parameter domain to form a bijective mapping between the surface net and the appropriate basic domain. After the shape surface is extracted, all the surfaces are mapped to the spherical domain, the mark points on the surface of the spherical domain of each shape are found by using the mark points on the reference sphere, and the shape space is inversely mapped.

In a 2-D space, a circle is used as the shared basic domain of the closed shape contour, and the registration of the two shape contours is achieved through their parameterization, which is similar to using the relative arc length to obtain the corresponding relationship between the shapes. In the 3-D space, the parameterization of the surface of the shape is relatively complicated, and the topological relationship of the shape generally needs to be used. Most methods are limited to the orientable closed 2-D manifold with zero genus (that is, the surface of the shape has no holes and no self-intersection), and their common basic domain is a spherical surface.

Since the research object of this thesis focus on human organs with uncertain deformation and complex surface contours, there are large differences between the shape contours in the training set, in this thesis, we use a method based on the registration of parameterized shapes to complete the landmarks correspondence in the spherical domain after conformal mapping.

2.2.3 Preliminary Localization for Multiple Objects

In addition to the structure of models, the initialization of SSM and other models before the segmentation is also very important, although it is in the supervoxel-based methods [68-69]. Before the segmentation searching, the shape of SSM is required to be initialized into a close range of the target organ in the image. Casually, the appearance model, as a part of SSM, only learns local profiles of intensity within a small range of the model surface. Thus, at the beginning of the searching process, if the model is placed far from the target, it is difficult to move the model to the target when the surrounding voxel intensities are far different from the local appearance model and the searching process may fail into a local optimum which differs a lot to the ground truth. Therefore, a robust initialization method that can initialize the model close to the target organ is very important for the SSM methods.

Yao et al. proposed a statistical location model for abdominal organ localization [70]. The position relevance of pre-segmented spinal and abdominal organs is obtained by

using the statistical location model (SLM). The locations of the organs are found by the maximum a posterior (MAP) estimation of the probabilistic density model. In addition to the position, Liu et al. also focuses on the organ pose and employed a distribution model and a MAP framework for automated abdominal multi-organ localization [71]. The orientations and scales, in addition to the locations, are obtained and thus the result is more accurate. The method show performs well for the liver, spleen, left and right kidneys, but is less adequate for the pancreas. In 2011, Criminish et al. introduced the machine learning method of regression forest for efficient anatomy detection and localization in CT studies [72]. The atlas of the organs is incorporated within a compact tree-based model. Compared with affine, atlas-based registration methods, they achieved a faster localization with smaller error mean and SD. Pauly et al. conducted a fast multiple organ detection and localization in whole body MR dixon sequences [73]. The random ferns method is used for regression with the 3-D LBP descriptors, which improved the accuracy of the anatomy localization with higher efficiency and robustness. Gauriau et al. proposed a global-to-local cascade of regression RF to realize accurate and robust localization of several organs in medical images [74]. A first regressor encodes the global relationships between organs and the localization of each organ are independently refined in the subsequent regressor. The confidence maps by combining regression vote distribution and organ shape prior to compute confidence maps, which enhanced the consistency and accuracy of multi-organ localization.

To overcome the problem of high requirements of Random Access Memory (RAM) and storage for offset vectors along each bounding wall and requirement of large number of leaf nodes, Samarakoon et al. propose the light random regression forests for automatic multi-organ localization in CT images [75]. The storage and RAM are reduced by Light Random Regression Forests (LRRFs), which describes the random variables inherent to the random processes. The LRRF comprising 4 trees with 17 decision levels is approximately 9 times faster, takes 10 times less RAM, and uses 30 times less storage space compared to a similar classic RRF.

2.3 Simultaneous Segmentation of Multi-objects

The multi-objects segmentation method developed on the basis of single organ can obtain the segmentation results of multiple organs in the body cavity at one time. In addition to using the information of individual organs, the multi-organ segmentation method can also combine prior information to make full use of the spatial information corresponding to each organ in the structure of multiple organs. Compared with the single-organ segmentation method, it can further improve the accuracy of segmentation. Degree and stability, reduce the target boundary overlap in the segmentation result. In the practical application of CAD, multi-organ segmentation can meet the needs of obtaining information about multiple organs in the body area and provide more and more valuable information for doctors' diagnosis and follow-up treatment.

The new requirement of multi-organ segmentation is rising along with the evolution from organ-based to organism-based approaches in modern medical diagnosis, and the analysis of multiple organs can also be helpful for comprehensive diagnosis or pre-operative planning and guidance in CAD systems [76].

2.3.1 Conventional Approaches

The multi-organ segmentation method is a segmentation framework for multiple organs in a certain area proposed based on a single organ segmentation method. Since Cootes et al. [44,45] proposed a segmentation method using PDMs, SSMs have been greatly developed in the field of single organ segmentation.

Many models involving multiple organs are mainly used for the segmentation of multiple organs (targets) in the brain, chest, and abdominal CT. Multi-organ segmentation based on gray registration of non-rigid maps [77] is not ideal for CT images acquired under different imaging conditions. The multi-organ segmentation method based on machine learning is mostly applied to the segmentation of the liver, spleen, kidney, and aorta, but not for organs with inter-subjectivity in shape and position [78].

Statistical model-based methods can be divided into two categories: one is multi-structured joint modeling, with very stable shape and position, but the segmentation accuracy needs to be improved; the other is modeling multi-organ correlation joints and multi-scale modeling [79] improves the accuracy of segmentation and the stability of segmentation. The multi-organ segmentation methods are generally developed from single organ segmentation in which field statistical atlas and shape models are widely used.

In 2007, Shimizu et al. [80] proposed an atlas-guided segmentation method on twelve organs with level-set refinement. Twelve organs are simultaneously extracted from non-contrast 3-D abdominal CT images, by using the abdominal cavity standardization process and segmentation of rough atlas guided segmentation with expectation maximization (EM) algorithm based parameter estimation and the following multiple levels set fine segmentation. In 2012, Linguraru et al proposed a multi-organ segmentation method from multi-phase abdominal CT via 4D statistical graphs using enhancement, shape, and location optimization [76]. The results significantly over those of the 4D convolution for all organs, the employed shape and location information bring a further improvement for liver and spleen. Wolz et al conducted an automated segmentation of abdominal organs in a multi-atlas registration framework in [81]. The hierarchical model with three levels, the global, organ, and voxel level, are designed to overcome the problem of subject-variability. Chu et al. used a spatially divided probabilistic atlas to reduce the inter-subject variance in organ shape and position with the global and local weight assigned and conducted the segmentation incorporating a MAP estimation and a graph cut method [82]. To realize the fast multi-organ segmentation, Lay et al. used context integration and discriminative models integrating both local and global discriminative information [83]. The result shows the time of segmenting up to 6 organs in either CT or MR data cost roughly one to three seconds, with segmentation in MR less than one second, while maintaining the accuracy. Okada et al. [68] constructed a hierarchical multi-organ statistical atlas with constraints for multi-

organ inter-relationships embedded by introducing prediction-based conditional shape–location priors from organ correlation graph (OCG). The predictor organs are pre-segmented and used to guide the segmentation of the remaining organs hierarchically by the conditional shape–location priors. This method increases the accuracy as well as extends the applicability to various imaging conditions without supervised intensity information.

These atlas and shape-model-based methods show better robustness in favor of the prior knowledge obtained from the training set. However, more accurate registration is required to match the pre-trained atlas to the test image, and shape correspondence within the training set is necessary for shape models.

As the initial position of the organs is required in the previous method, in 2016, Zografos et al proposed a method using a discriminative classifier trained based on multiple levels of super voxels features and obtain the final segmentation result by hierarchical conditional random field fusion [69]. Takaoka et al. also introduced the supervoxels and use them instead of voxels as the unit of graph cut model, with an energy function to minimize it [84]. The accuracy was improved in the energy minimization and the problem of missing organs and misshaped segments were alleviated.

2.3.2 Deep Learning Based Approaches

Instead of training models, deep convolutional neural network learning techniques are also introduced to this field, benefited from their outperformed semantic segmentation ability based on the mechanism of feature extraction using multiple convolution layers.

In 2017, Zhu et al. used an FCN to realize a semantic segmentation of nineteen anatomical structures from 3-D CT images with a simplified majority voting scheme over the semantic segmentation of multiple 2-D slices drawn from different viewpoints with redundancy [85]. In 2018, more attentions are focused on the field of deep-learning-based medical image segmentation. Roth et al. used a cascaded 3-D FCN to improve the

inaccuracies of smaller organs and vessels in a coarse-to-fine approach [86]. For the problem of a small, partially annotated dataset in the deep learning training process weakly supervised training is combined in recent methods. In the work of Gibson et al., the segmentation of multiple organs is conducted with Dense 3-D V-networks [87]. The batch-wise spatial dropout scheme is applied to lower the computation cost. In [88], Wang et al. proposed a novel framework for multi-organ segmentation of abdominal regions by using organ-attention networks with reverse connections (OAN-RCs) which are applied to 2-D views, of the 3-D CT volume. The output is estimated by a combination of statistical fusion exploiting structural similarity. In [89], a simple but effective sample selection method is introduced for training multi-organ segmentation networks. They focused on the training sample selection problem instead of network architecture. A Relaxed Upper Confident Bound (RUCB) strategy for sample selection was proposed to mitigate the influence of annotation errors during the training process and increase the segmentation performance. Bobo et al. applied the FCN to the segmentation of abdominal organs from MRI images [90]. Zhou et al. proposed a Prior-aware Neural Network (PaNN) using anatomical priors on organ sizes and domain-specific knowledge in the training process [91]. In 2020. Conze et al. introduced the cascaded convolution and adversarial deep network in the abdominal multi-organ segmentation from abdominal CT and MR images [92]. Fang et al. proposed a unified training strategy that enables a novel multi-scale deep neural network to be trained on multiple partially labeled datasets for multi-organ segmentation and achieved a promising performance [93].

The accuracy of the multi-organ segmentation tasks is tremendously improved by the deep neural networks, benefit by the automatically selected features. However, the training of deep networks requires large calculation resources and manually labeled training data, which is difficult to obtain. The overfitting and gradient vanishing are still or even more serious problems for deep neural networks applied to medical image segmentation tasks, especially for 3-D tasks.

As a conclusion, we use the SSMs for segmentation of human organs, which are of a relatively fixed patterns of surface shapes. The robustness of the segmentation can be guaranteed by the mean shape and limited deformation in the SMMs. And for the segmentation of small tumors with uncertain shape, size and appearance, the advantage of feature learning is shown by the convolutional layers in the deep neural networks.

Chapter 3

Statistical Shape Model Building

In this chapter, we introduce the basic process of building an SSM, which is prepared for the task of multi-organ segmentation. The quality of SSMs mainly depends on the accuracy of landmarks correspondence, which is the process of capturing deformation in the organ surfaces of individuals organs. We propose an SSM building method based on k-means clustering, which improved the model quality by providing more approximate prototype in the correspondence. To realize the construction of multi-organ SSMs, we extend the model building from single organs to multiple organs. A non-rigid iterative closest point (NIPC) approach is introduced to match the landmarks from the multi-organ surfaces.

3.1 Introduction

The SSM-based approaches are of excellent robustness and accuracy in the segmentation tasks. In the process of model building, the deformation patterns of the shape model are learned from the set of landmarks. This process highly depends on the accuracy of the corresponding landmarks which are marked on each of the shapes in the training set. Each landmark with the same serial number represents an identical position of the shape. The implied distribution patterns of the points can be recovered by the statistical analysis of these landmarks. The quality of the SSMs is also relevant to their ability to correct deformation, which contributes to the correct segmentation of the region of interest. In the following section, we will introduce the basic process of building an SSM, as well as its application to tasks of image segmentation.

In section 3.2, the relevant concepts of SSMs building are provided. Section 3.2.1

explains the basic steps of building an SSM. The indicators used to evaluate the quality of models, i.e., the generalization ability, specificity, and compactness, are described in section 3.2.2.

Section 3.3 provides an improved landmarks corresponding method for SSMs building. The details of the methodology, including the *k*-mean clustering which are used in the process of landmarks corresponding, are introduced in section 3.3.2. Two SSMs of left lungs and right lungs were built and their quality was illustrated in section 3.3.3. Section 3.3.4 contains the discussion of the experiment and the conclusion.

We also proposed a multi-organ SSMs building method which is based on non-rigid iterative closest point (NICP) registration method in section 3.4. The SSMs of single organs are extended to multiple organs and illustrated in the section. The detailed process of model building, introduced NICP approach, is explained in section 3.4.2. The proposed method was tested by an experiment of building multi-orang SSMs and their model quality is evaluated and shown in section 3.4.3. The results are discussed in section 3.4.4 with the conclusion and the future work.

3.2 Model Building and Quality Assessment

3.2.1 Model Building Framework

The aim of model building is to generate a mean surface mesh with patterns of defamation so that it can be deformed flexibly within a certain range and limitation. This process can be regarded as an information compression of the shapes in the training set by using statistical methods. In this section, we provide an introduction to the common process of SSMs building from a training dataset of labeled medical images. The flow chart of this process is shown in Figure 3.1.
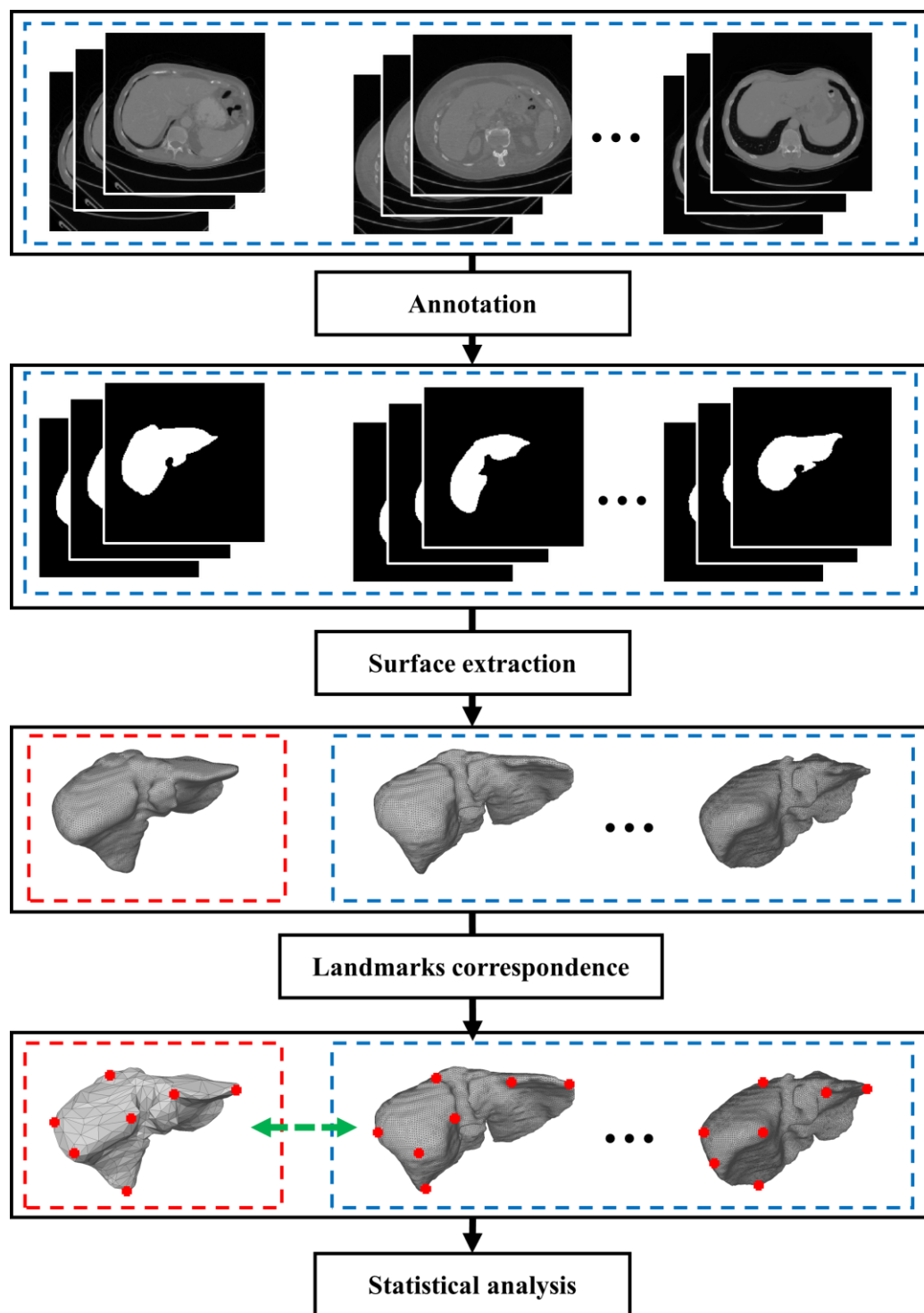
Figure 3.1 Framework of statistical shape model building.

(1) Datasets annotation

The basic material is a training set containing 3-D medical images from different patients. In each image, the voxels of the desired organ are manually labeled by professional physicians.

As the thickness of CT slices is usually different, the spatial resolution of the 3-D image in each direction needs to be unified isotopically so that the voxels in the reconstructed images are truly reflexing the real structure of the patients.

(2) Surface extraction

The process of surface extraction is to obtain the surfaces of the object organs from the labeled voxel data. In the 3-D CT images, the human organs are expressed by connected cubic voxels. To extract the surface of organs, the marching cubes (MC) algorithm is a commonly applied method [94]. In the algorithm using the divide and conquer strategy, eight pixels from two adjacent slices form a cube and the key is to find a surface intersection in the cube. 14 patterns from all 256 cases of possible intersection ways are precalculated to approximate the linear interpolations and output the triangular surface meshes. The algorithm traversed all cubes and produces a triangular surface mesh. For multiple organs, the marching cubes are performed organ by organ to avoid merging extremely closing organs to one the same surface.

As the original data from medical images are obtained from the imaging system under certain resolutions, the naturally smoothing human organs and tissues are inevitably transformed to discrete voxels data and the discreteness remains in the triangular surfaces obtained from the MC algorithm. To decrease the roughness of the surface and recover the smoothness in general, a Gaussian filter is applied to the surfaces. The smoothing procedure can also remove the noise of cube edges and benefic the following registration procedures. In surface registration, the similarity is a basis to find the correspondence. However, some of the similarities in local voxel structures are produced from the MC algorithm but not from the original anatomical features of human organs. The smoothing

can reduce the effect of the MC and restore the basic anatomical contours of the organs.

Although the quality of the surfaces is improved in the smoothing procedure, the quantity of vertices and faces on the surface remains the same and it is more than required. The vertices originated from MC are generated cube by cube and redundant vertices that describe the same geometry structures can be simplified to decrease the computational amount of the following processes on surface meshes.

(3) Landmarks correspondence

After the surfaces are reconstructed and smoothed, for each of the surfaces, a set of landmarks that represent the identical position in the organ surfaces correspond. Before the registration matches the vertices of the surfaces, it is necessary to prepare a set of standard landmarks. A set of ideal landmarks should be representative of the anatomical structure of the object organ surface with a small ratio of the vertices and faces in the original surface. Also, as the final aim of landmark corresponding is to build an SSM, it is important to capture the varieties that existed in the identical positions of landmarks from different patients. A set of satisfied reference landmarks is the first step of building an SSM in high quality.

The next step is surface-wise registration. The most straightforward way is to register the surfaces to the surface which is used to generate the reference landmarks. Then it can be treated as a surface registration problem, which aims at finding the special relation of identical vertices. However, there is a difference between the landmarks correspondence and the conventional surface registration. The purpose of surface registration is to find the correspondent points for all of the vertices, but the landmarks are highly representative vertices and only occupy a small proportion of the whole surface. This feature enables the surfaces used in landmarks corresponding to be tolerant lower resolution.

After the surface registration, the standard landmarks in the reference surface

correspond to all of the surfaces in the training set. The identical set of landmarks on each surface represent the distribution of the same anatomical structure in different patients. Thus, the accuracy of landmarks corresponding is very important to build shape models that can reflex the real deformation patterns of the human organ.

(4) Statistical analysis

The SSM is a deformable shape model, where the deformation patterns are learned by statistical analysis from a training set consisting of corresponding landmarks. The SSM describes the vertices distribution of the shape by a mean shape and deformations.

An initial alignment of the training shapes is required to remove the global linear difference of translation, rotation, and scaling among the shapes by the Procrustes analysis [95]. This process concentrates the shapes obtained from different images to a unified space and helps preserve the actual varieties of shape counters instead of their spatial distribution in the images.

After the alignment, the shapes and positions of all surfaces in the training set are the same. The shape differences that have nothing to do with shape changes are eliminated, and the change patterns of corresponding marked points among all samples can be counted.

Principle Component Analysis (PCA) is a statistical method that uses an orthogonal transformation to transform a set of observations that may have correlated variables into a set of linearly uncorrelated variables i.e., the principal components [96]. In the shape model, the PCA method is used to extract the main deformation pattern of the position change of the landmark. After the process of PCA, the model is expressed as a linear structure composed of an eigenvector reflecting the deformation pattern and an eigenvalue reflecting the degree of change.

In the training set of SSMs containing N instances, the ith landmark shape $\boldsymbol{x}_i$ is composed of n vertices $\boldsymbol{x}_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{n1}, x_{n2}, x_{n3})^T \in \mathbb{R}^{3n}$, and the training set

is $X = (x_1, x_2, \ldots, x_N)$. To capture the deformation patterns, PCA is employed to the training set. The mean shape $\bar{x}$ and covariance matrix S_c is firstly calculated:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{3.1}$$

$$S_c = \frac{1}{N-1} \sum_{i=1}^{N} (\bar{x} - x_i)(\bar{x} - x_i)^T \tag{3.2}$$

And the eigenvectors $\Phi = (\phi_1, \phi_2, \ldots, \phi_m)$ and their corresponding eigenvalues $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ can be calculated by eigenvalue decomposition, which satisfies:

$$S\phi_i = \lambda_i \phi_i \tag{3.3}$$

where $m = \max((s-1), 3n)$. The eigenvalues are sorted and the largest c eigenvalues and eigenvectors are retained to move noise of the data. A proportion δ (usually ranged from 0.95 to 0.995) is used to obtain C by:

$$\sum_{i=1}^{C} \lambda_i \geq \delta \sum_{i=1}^{m} \lambda_i \tag{3.4}$$

After obtaining the shape model, any shape x can be approximated as $\tilde{x}$ by adjust the weight parameter b of the model:

$$\tilde{x} = \bar{x} + \Phi b \tag{3.5}$$

3.2.2 Evaluation of Model Quality

After the PCA-based dimensionality reduction, an SSM that consists of a mean surface mesh and vectors of shape deformation is obtained. To evaluate the model quality quantitatively, three indicators that measure the ability of the new SSM are introduced by Davies [97], including the generalization ability, specificity, and compactness.

(1) Generalization ability

Generalization is the most basic ability of a model, which reflects the general expression ability of the model to learn the target shape characteristics from a limited number of shapes in the training set. A high-quality model can not only express the sample shapes included in the training set well but also effectively express the shapes that are not included in the training set. If the model overfits the training set, it will not be able to generate unknown shapes.

Suppose M is the number of modes which are used to build the model for fitting the current instance outside the training set as much as possible and number of cases inside training set is n_t, the generalization ability $G(M)$ is expressed as:

$$G(M) = \frac{1}{n_t,} \sum_{i=1}^{n_t} |x_i - x_i'(M)|^2 \tag{3.6}$$

where x_i denotes the excluded shape and $x_i'(M)$ denotes the shape best fitted to x_i by using M modes of the shape model. For model A and model B with $G_A(M)$ and $G_B(M)$ being their generalization abilities, if $G_A(M) \leq G_B(M)$ for all of the M and $G_A(M) < G_B(M)$ for some of the M, model A is considered better than B in the aspect of generalization ability.

The standard error of $G(M)$ is defined as:

$$\sigma_{G(M)} = \frac{\sigma}{\sqrt{n_t - 1}} \tag{3.7}$$

where σ is the sample standard deviation of $G(M)$.

(2) Specificity

The specificity reflects the ability to generate shapes which are similar to those in the training set. It is expressed as:

$$S(M) = \frac{1}{n_e} \sum_{j=1}^{n_e} |x_j(M) - x_j'|^2 \tag{3.8}$$

where x_j is the shape generated by the model with n_e different eigenvalues on M modes and x_j' is the most approximate shape to $x_j(M)$ in the training set. Similarly to the generalisation ability, for model A and model B with $S_A(M)$ and $S_B(M)$ being their specificity, if $S_A(M) \leq S_B(M)$ for all of the M and $S_A(M) < S_B(M)$ for some of the M, model A is considered better than B in the aspect of specificity.

The standard error of $S(M)$ is given by:

$$\sigma_{S(M)} = \frac{\sigma}{\sqrt{n_e - 1}} \tag{3.9}$$

where σ is the sample standard deviation of $S(M)$.

(3) Compactness

The compactness assesses the ability of constructing the instance with the minimum number of modes possible. Defining λ_i as the i th eigenvalue, the compactness is described by the accumulation of variance:

$$C(M) = \sum_{i=1}^{M} \lambda_i \tag{3.10}$$

For $C_A(M)$ and $C_B(M)$ being the compactness of method A and B, the comparison of the compactness is similar to the specificity.

The standard deviation of the λ_ith variation pattern of can be obtained from the sample mean of the sampling distribution:

$$\sigma_{\lambda_m} = \sqrt{\frac{2}{n_t}} \lambda_m \qquad\qquad (3.11)$$

where λ_m is the mth largest eigenvector in the covariance matrix. The standard errors of $C(M)$ can be obtained by:

$$\sigma_{C(M)} = \sum_{m=1}^{M} \sqrt{\frac{2}{n_t}} \lambda_m \qquad\qquad (3.12)$$

3.3 Optimization of Points Correspondence

Medical image segmentation is one of the important basic steps in a CAD system. The volume data obtained by segmentation is widely used in 3-D organ reconstruction, pathological analysis, disease tracking, treatment planning, surgical navigation, visual medical treatment, and medical training. Among many segmentation methods, SSM is a robust and accurate global a priori method for medical images with high diversity and complexity. The SSM is usually expressed with a PDM which consists of a basic mean geometric surface and statistical feature modes of geometric surface variation learned from a training set using the PCA method. The geometric surface in the PDM contains a set of labeled vertices, called landmarks, and triangular faces.

Landmark correspondence which entails finding landmarks that represent the same structure of the shape from each geometric surface in the training set is an important step in the SSM generation process. A successful generation of SSM leads to a well-learned mode of shape variation and subsequent segmentation. Manual landmark correspondence is feasible in 2-D SSMs generation. However, such an endeavor is both time-consuming and error-prone in the case of 3-D SSMs generation. Apart from these difficulties, a good quantitative definition of correspondence is problematic for automatic methods.

Davies et al. [97] proposed an automatic method of optimal 3-D SSM generation. In

their method, an optimal parameterization of shapes is represented explicitly and manipulated for the optimal model generation. The optimal object function was based on the minimum description length (MDL) principle. Styner et al. [98], compared three automatic landmark correspondence methods including SPHARM (spherical harmonic), MDL, and the covariance determinant (DetCov) with manually initialized subdivision surface method by using the evaluation method proposed in [47]. They concluded that the MDL method had better performance. Heimann et al. [50] present an optimized method for the description minimization which is easier and faster than [47].

In a previously proposed model building method [99-100], a quadric-based simplification of the original geometric surface is performed on the training set. These are then regarded as the reference landmarks to find the corresponding landmarks among all the other surfaces. Prior to the correspondence analysis, mapping to the spherical domain and diffeomorphic demons registration method was conducted on the training set. Improved segmentation accuracy was reported but the model quality is not improved.

In the following section, we present an improved SSM generation method that solves the drawback of the previous method [99-100]. The proposed method conducts the landmark correspondence procedure using a reference surface where the landmarks are found by the k-means clustering method. This is in contrast to the quadric-based simplification method. In section 3.3.2, two training sets respectively containing left lung and right lung geometric surfaces are tested by the proposed method to build two SSMs. An evaluation and discussion of the model qualities as well as the segmentation result are conducted in section 3.3.3.

3.3.1 Methodology

3.3.1.1 The main architecture

An overview of the proposed method is depicted as a flowchart in Figure 3.2. In the first step, the prototype landmarks are initially prepared. An instance of geometric surface, which contains numerous vertices and meshes, is selected from the training set

and a *k*-means clustering method [101] is used to find the prototype of the training set. Given the required number of landmarks, the vertices on the surface are divided into clusters based on a feature vector including the position information of the vertices on the original surface and spherical conformal mapped surface as well as their curvature. The cluster centroids are regarded as the prototype of landmarks selected from the geometric surface of the instance.
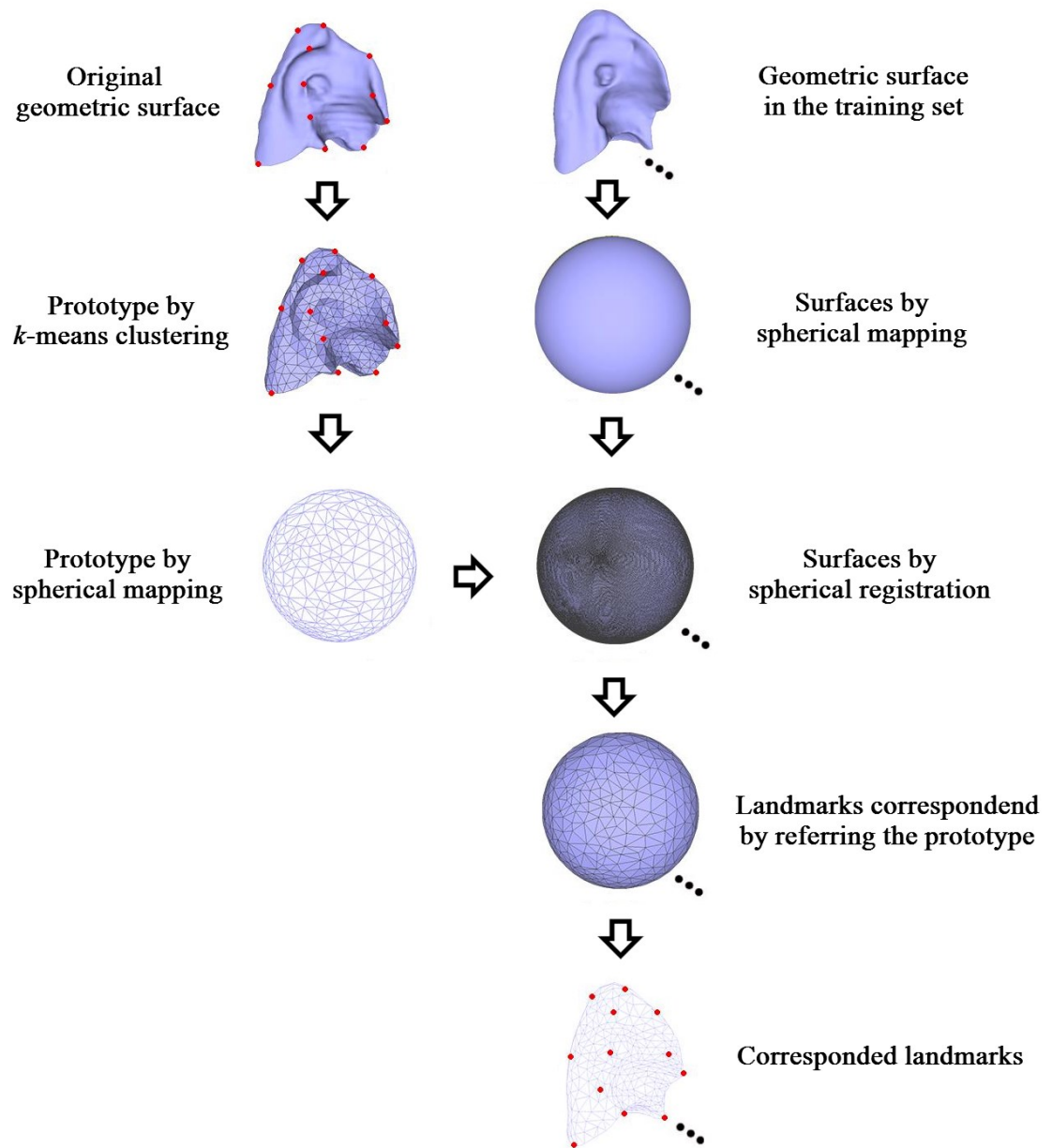
Figure 3.2 Flow chart of the proposed landmarks corresponding approach. The top line shows the process of building a prototype of a set of standard landmarks. The second line explains the process of each surface in the training set and their interaction with each other and the prototype.

Second, a spherical conformal mapping [102] is performed on each instance (with landmarks) of the surface in the training set to map the geometric surfaces from the Euclidean domain to the spherical parametric domain.

Next, a spherical demons registration [103] is performed on all of the mapped spheres, taking advantage of the gradient information of the voxels in the original image corresponding to each vertex on the surface. This is a group-wise registration among all the spherical surfaces.

After the registration, the relation of vertices between the reference surface and the other surfaces has been built. The corresponding landmarks of each instance can be indexed by finding the nearest vertices on the prototype of the reference surface. As the surface mesh is not changed in the previous spherical mapping process, the landmarks can be easily found on the original surface. With these landmarks on each instance of the training set, an SSM can be generated by the statistical analysis approach described in section 3.2.1.

3.3.1.2 *K*-means clustering

Before the registration and correspondence, it is necessary to prepare a set of standard landmarks as the prototype of a training sample. A set of ideal landmarks should be representative of the anatomical structure of the object organ surface with a small ratio of the vertices and faces in the original surface. Also, as the final aim of landmark corresponding is to build an SSM, it is important to capture the varieties that existed in the identical positions of landmarks from different patients. A set of satisfied reference landmarks is the first step of building an SSM of high quality.

The authors in [99-100], developed a quadric error-based metric surface simplification method to find the prototype of the surface by edge contraction idea. In this section, the generation of the prototype is posed as a clustering problem. Vertices of the original surface are clustered into different groups such that within-group similarities are larger

than between-group similarities. The resulting cluster centroids are regarded as the prototypes of the training set.

First, one of the surface cases is randomly chosen from the training set of surface meshes obtained before. Then, a k-means clustering-based surfaces simplification method [101] is performed to extract representative vertices from the chosen surface as the reference landmarks. The k-means algorithm is an expectation-maximization (EM) procedure that iteratively searches for the optimal centroids. Given a required number of landmarks k and a surface with n vertices, for each of the vertices, a set of spatial and geometrical feature vector v is extracted, and the vertices are divided into k clusters as $D = \{D_1, D_2, \dots, D_k\}$ with c_i the center of the ith cluster. The clustering is equal to find a stable division D with minimal within-cluster sum of squares (WCSS):

$$\arg \min_{D} \sum_{i=1}^{k} \sum_{v \in D_i} \|v - c_i\|^2 \qquad (3.13)$$

In the initial step of the algorithm, a set of k initial clustering centers is chosen, and an initial clustering is obtained by Voronoi diagram division. Then the algorithm turns into iterations where new centers c are updated by the new divisions D until it converges (the division is fixed).

In our implementation, seven features are used viz. the three coordinates in the Euler space; three in the sphere space; and the curvature calculated on the geometric surface. The key advantage of k-means clustering in the proposed method lies in its effectiveness and ability to distinguish geometric surfaces.

The obtained centroids of the vertices in each of the clusters are regarded as reference landmarks. The faces of the landmarks could be generated by spherical Delaunay triangulation which is performed on the sphere mapped from the chosen surface mesh. The landmarks from the clustering-based simplification method possess better global

representativeness, compared with local surface simplification methods, due to the operation of clustering being applied to all vertices at the same time within each iteration. A surface mesh of the right lung in the clustering process is shown in Figure 3.3.

(a)   (b)   (c)
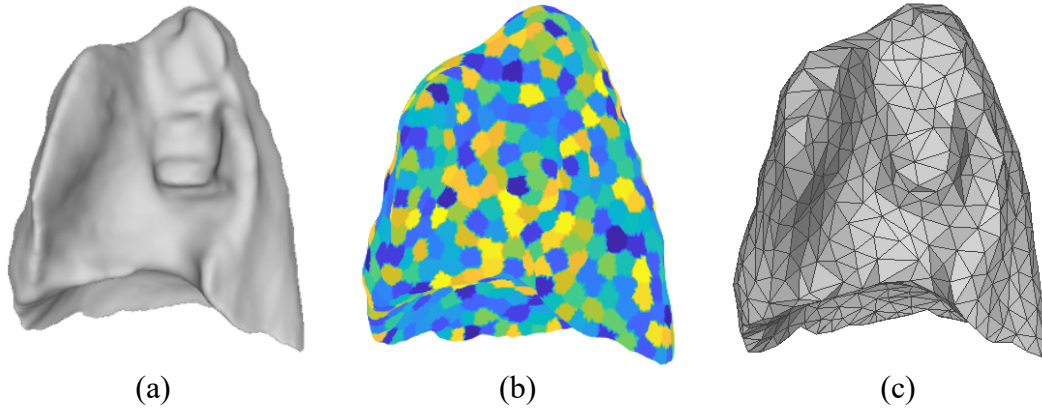
Figure 3.3 Illumination of a surface mesh of right lung in the clustering process. (a) Original surface of the reference instance. (b) Surface with marked regions represent the clustered vertices. (c) The generated prototype, where the new vertices are obtained from the centroids of each region in (b).

The simplification by the clustering is accomplished through a global search in contrast with the surface simplification method [99-100] where edges are contracted iteratively under the condition of the local geometry. Four cases of landmarks correspondence from the surfaces are shown in Figure 3.4. The geometric surfaces with blue landmarks in the first row are obtained by the correspondence method of [99-100] and the red landmarks in the second row are generated by the method proposed in this section.

44
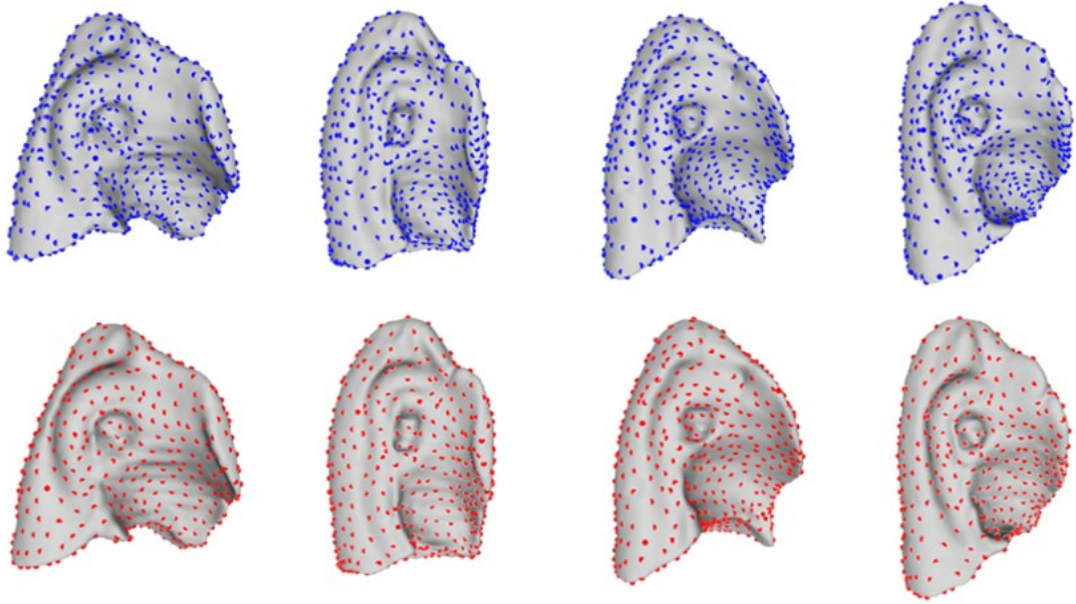
Figure 3.4 Examples of four cases of corresponded landmarks from the training set. The surfaces in the top line are marked with blue landmarks which are corresponded by previous method and the red landmarks in the second line are corresponded by our proposed method.

3.3.2 Experimental Results

In our study, 20 cases of thoracic CT from the Lung Image Database Consortium (LIDC) are used [104]. The size of the images varies from 144×144×127 to 180×180×205 voxels and the gap between each voxel is 2mm. The training set of left lungs and right lungs are obtained with the same method described in [100]. In our experiment, the geometric surfaces consist of about 69600 vertices and 139300 faces on average. After correspondence estimation, each mesh consists of 642 vertices and 1280 faces, which is the same as in the previous method for comparison.

The performance evaluation of model quality on the left lung and right lung training sets are shown in Figures 3.5, 3.6 and 3.7. The triangle symbol in blue represents models constructed by using the previous method [99-100] and the square symbol in red represents those of the proposed method.

Compared with previous methods, the generalization ability and specificity of SSMs generated by the proposed method increased. Landmarks corresponding with the prototype generated by *k*-means clustering are better approximation of original geometric surfaces.

(a)

(b)
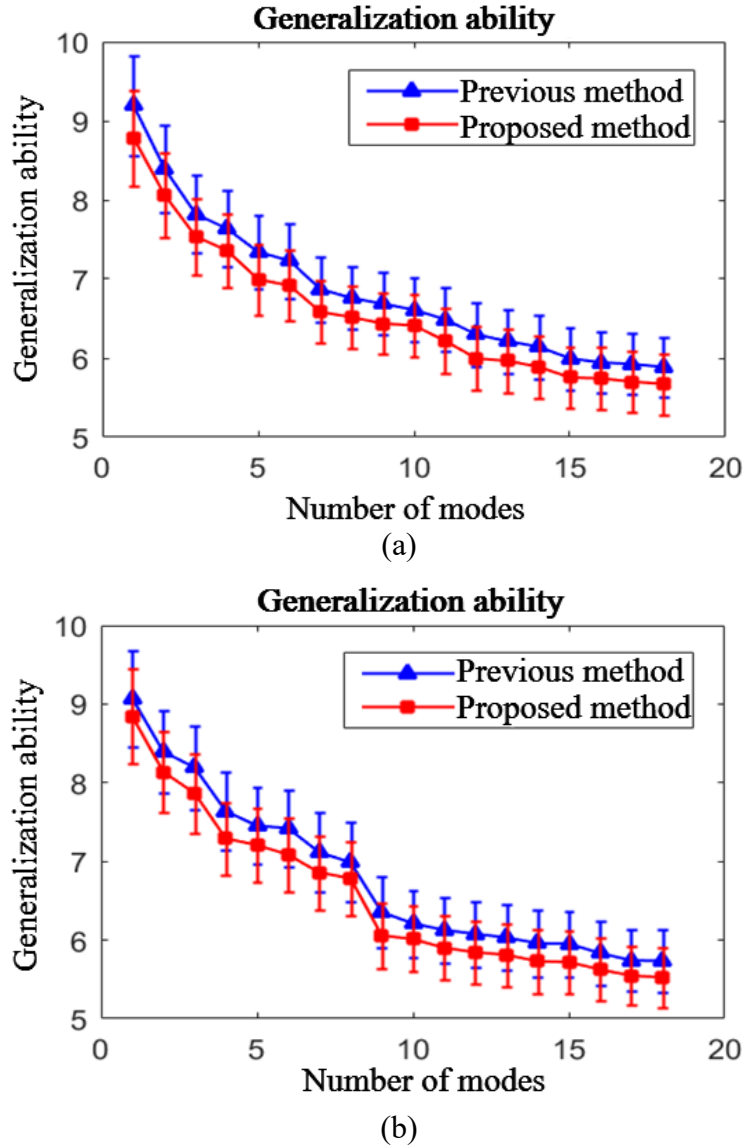
Figure 3.5 Generalization ability of (a) left lung SSMs and (b) right lung SSMs. The models are constructed by the previous method and our proposed method, respectively.
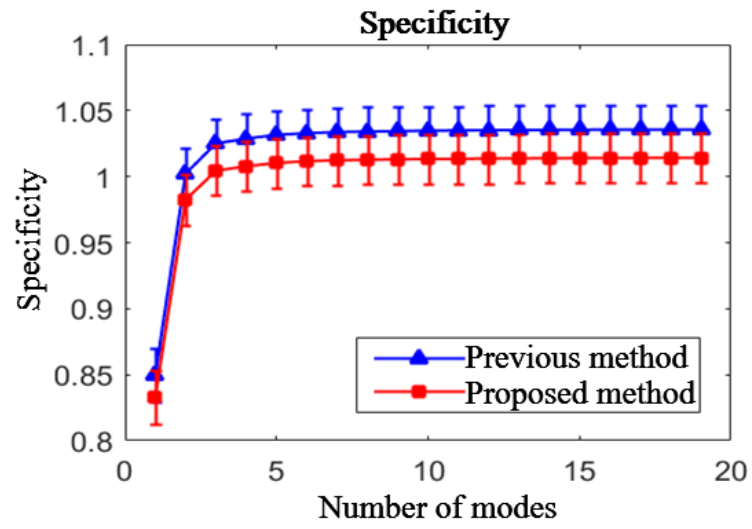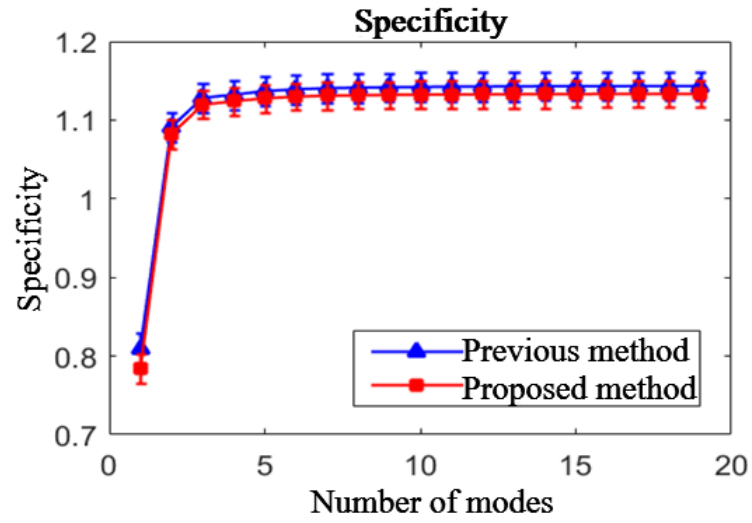
Figure 3.6 Specificity of (a) left lung SSMs and (b) right lung SSMs. The models are constructed by the previous method and our proposed method, respectively.
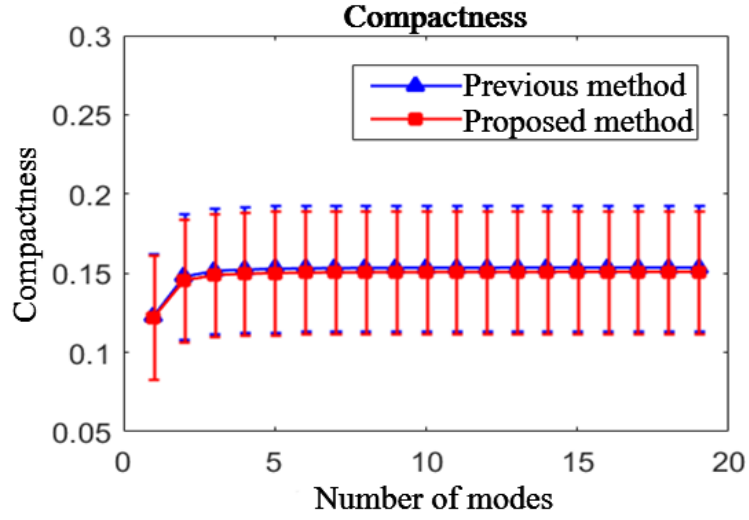
(a)
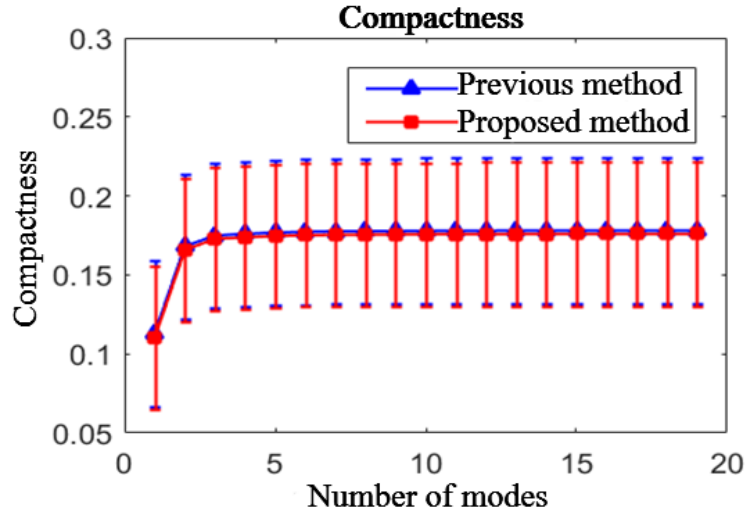
(b)

Figure 3.7 Compactness of (a) left lung SSMs and (b) right lung SSMs. The models are constructed by the previous method and our proposed method, respectively.

3.3.3 Discussion and Conclusion

Compared with the landmark corresponding method which is based on the surface simplification algorithm using the Quadric Error Metrics in [99-100], the new landmark which is generated by the clustering is more effective. The reason is that the landmarks are selected to form the PDMs, and each of the landmarks should be the most

representative vertex of their surrounding vertices on the original surfaces. The clustering algorithm is more efficient for this problem, as the similar vertices are collected within the same clusters and the centroids of each cluster are chosen as the landmarks, instead of the method using the iterative contraction of vertex pairs whose main object focus on minimizing the geometric error in the simplification process.

In this section, we proposed an automatic landmark correspondence method for the generation of 3-D SSM. The quality of the model built by using landmarks correspondence generated with the *k*-means clustering method is higher than that of the model built with a previous method in terms of generalization ability and specificity while maintaining the same compactness. The improvement of the model quality is originated from the prototype of the reference landmarks.

In the future, the robustness of the model should be improved for use in the segmentation of organs with severe lesions and the implementation of practical applications by introducing a multi-organ segmentation method.

3.4 Multi-organ Model Building by Non-rigid ICP Registration

SSM is a widely developed and applied tool in medical image analysis tasks, including segmentation, reconstruction of object organs or tissue [105]. The processed data and information in SSM can be used in computer-aided diagnosis (CAD), surgical planning and navigation, medical education, and so on fields. The SSM is composed of PDM, where the deformation patterns of the object shape are learned from the special varieties of corresponding landmarks on a group of aligned meshes represented by vertices and faces in a training set. The landmarks are a certain number of vertices picked from each of the dense surface mesh from different patients, which represent vertices with the same anatomic structure of shape. The correspondence of landmarks is an essential procedure in the SSMs building to accurately extract the information of variation across the surface mesh data. Although manually landmark corresponding is realizable for 2-D shapes with

a limited number of vertices and cases, it is difficult to find corresponding landmarks from 3-D surfaces due to the increasing quantity of candidate vertices and complexity of the geometric shape. Thus, automatic correspondence methods are more favorable for relevant tasks.

The problem of landmark correspondence can be cast as a combination of shape correspondence and landmark prototype building. The correspondence of shapes is usually stated as a problem of finding proper mappings between their elements (ordinarily vertices) and referred to as registration, alignment, or matching problems [106-107]. According to the completeness of the mapping, i.e., whether a full correspondence for each of the elements from the moving shape to the fixed shape is required, the correspondence problem can be classified into dense correspondence or sparse correspondence. The difficulties are almost the same because the searching space still covers the whole shape to find proper and meaningful correspondence, to ensure the quality of SSM built.

A fundamental distinguishment of the shape registration methods is the form of deformation, which can be roughly classified as rigid or non-rigid registration. The geometric deformation that matches one shape to another is differently chosen when adapted to shapes from different sources or for different tasks. One of the typical applications of rigid registration is the surface reconstruction from multiple point clouds partially scanned in different viewpoints of an identical object. Since the surfaces to be aligned are obtained from one object, only rigid deformation, i.e., translation, orientation, and scale are required in consideration. In rigid registration problems, noise, outlier, and the limited amount of overlapping are the main difficulties [108]. However, in the scope of biomedical surfaces, surface meshes are generated from medical images obtained from different patients, and some of the conditions and characteristics of the object surfaces are distinctive from those of ordinary surfaces. Thus, rigid deformation is not enough to describe the deformation of the organs or tissue surfaces and non-rigid registration is required to match the surfaces of individual differences.

Davies et al. [109] proposed an automatic landmark corresponding method that used the Minimum Description Length (MDL) principle to find the optimal parameterization of training shape. Three properties of the built shape model: generalization ability, specificity, and compactness are introduced as quantitative measures of model quality. In our previous research in section 3.3, the correspondence of landmarks is conducted across parameterized surfaces obtained from spherical conformal mapping and Demon registration. The parameterization that maps each of the original surfaces in the training set to spherical surfaces decreases the complexity of finding identity landmarks while preserving relevant information among vertices in the original organ surfaces to a certain extent. After a dense matching of the whole training set, landmarks with geometrical and anatomical representativeness are chosen from the original surface. In our previous methods, surface simplification using quadric error metrics and *k*-means clustering is implemented in one of the original surface mesh in the training set to obtain a set of reference landmarks. Ravikumar et al. [110] introduce a group-wise similarity registration using Student's t-mixture model for landmarks corresponding.

In the field of medical image analysis, more attention is paid to multi-organ models instead of organ-, and disease-specific methods [111]. The combination of inter-organ relations, including spatial, functional, and physiological relations, provides more accurate human anatomy information and benefits many medical procedures, including diagnosis, therapeutic assistance, radiotherapy planning, surgery simulation, or injury severity prediction. However, the jointed structure of multiple organ surfaces is unable to be parameterized into one single surface and the following spherical registration could not be conducted. Therefore, to extend our SSM building scheme and make it available for multi-organ models, a non-parameterization-based registration method is introduced and applied in the proceeding of landmark correspondence.

In the section, a non-rigid iterative closest point method is introduced for automatic surface registration of 3-D multi-organ surface meshes. The main architecture of the proposed method and the NIPC registration method are described is section 3.4.1. In the

experiment, different strategies of single organ and multi-organ combination are compared in the registration scheme on four abdominal organs from different patients, and the results are provided in the section 3.4.2. The results of surface registration and the SMMs building are discussion, and the conclusion is provide in section 3.4.3.

3.4.1 Methodology

3.4.1.1 Architecture

In our research, a NICP method is applied to landmark correspondence for multi-organ SSMs building from 3-D volume data of medical images. The flow chart of the proposed architect is shown in Figure 3.8. The total flow chart is similar to the process of model building architecture introduced in section 3.2.1. The difference mainly lies in the multi-organ structures and is reflected in the registration process.

In the first step, a series of preprocesses are necessary to obtain suitable surface meshes for landmarks correspondence from medical images. A marching-cubes algorithm is applied to the manually segmented voxel data, which is a process that transformed the labeled medical image of organs to triangular surface meshes. The following surface filtering is performed to improve the smoothness of the transformed rough surfaces. To decrease the computational amount of the following processes on surface meshes, a surface simplification algorithm is applied on the smoothed organ meshes.
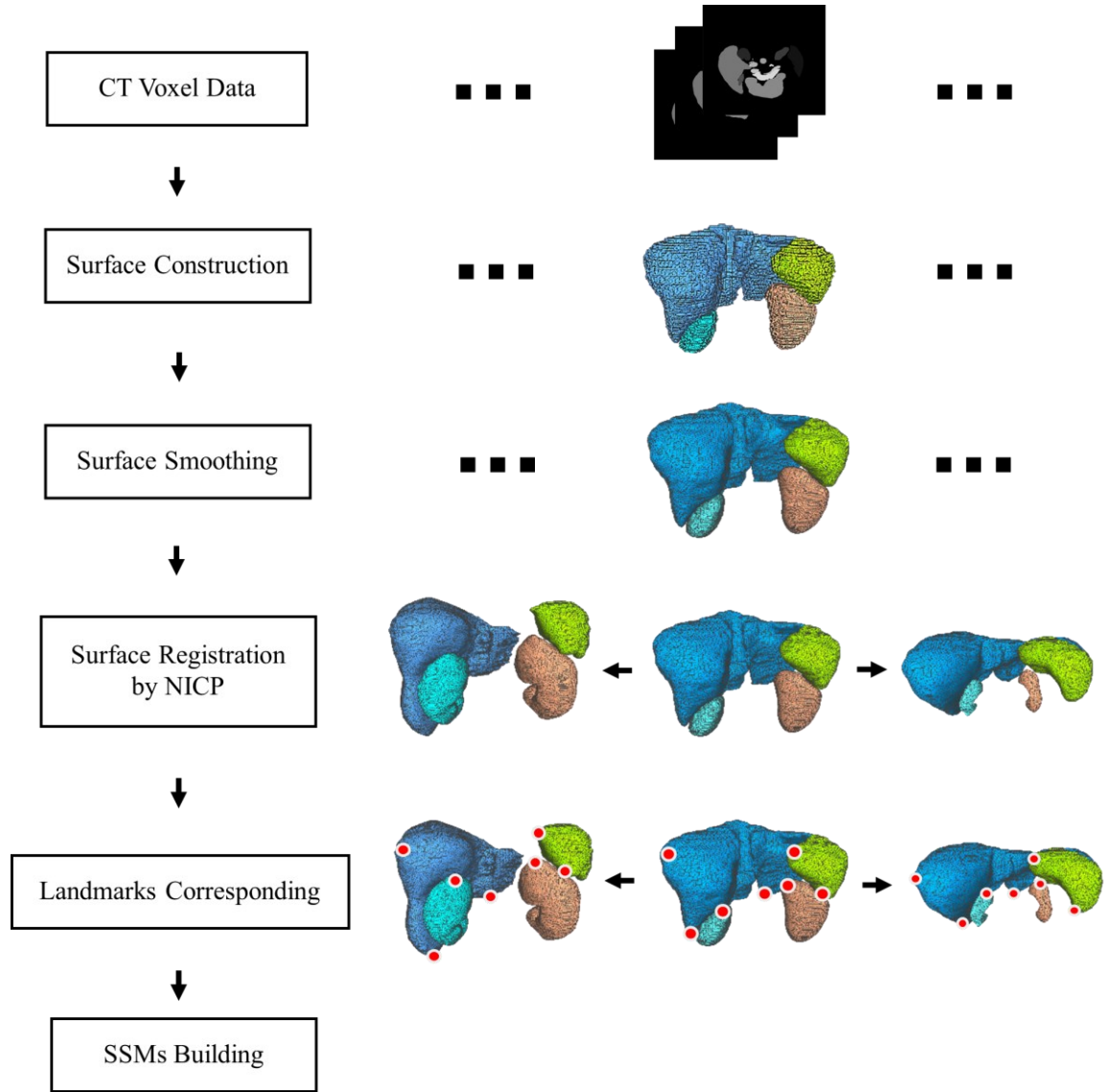
Figure 3.8 Flow chart of the proposed architecture. The landmark corresponding and SSMs building are conducted by NICP registration.

In the second step, a reference mesh of landmarks is prepared by a *k*-means clustering procedure which is performed on a chosen surface, as described in section 3.3.

Then, the chosen surface is regarded as the target surface and the other surfaces in the rest of the training set are registered to it. The landmarks of the other cases can be corresponded by finding the nearest vertices on the target surface from the deformed

surfaces.

After the correspondence, an SSM can be generated from the aligned landmarks by PCA. A series of surface deformation patterns are learned from the set of landmarks of human organ surfaces.

3.4.1.2 Registration and correspondence

Surfaces in the training set are extracted and reconstructed from CT images acquired from different patients and the states of body positions, which produces large diversity of identical organs and tissues and makes landmarks corresponding more difficult. Such differences could not be described and deformed within rigid registration and a non-rigid surface registration method is required. In this section, we introduce a NICP registration to match surfaces to the chosen reference surface throughout the training set.

Given two points sets $P^s = \{p_i^s, i\epsilon 1, \ldots, N^s\}$ of N^s vertices and $P^t = \{p_i^t, i\epsilon 1, \ldots, N^t\}$ of N^t vertices represent the source mesh and the target mesh respectively, a series of pairwise registration operations is described on them. At first, a rigid iterative closest point (ICP) algorithm using nearest neighbor (NN) are searched in the iterations. For each vertex p_i^s in the source mesh, one forward corresponding vertex and one or more backward corresponding vertex (vertices) can be corresponded in the target mesh. The bidirectional displacements from the vertices to each corresponding vertex of forward and backward are used to find rigid transforms of translation, rotation and scaling and get the new displacement $\boldsymbol{\delta}(p_i^s)$. After the rigid registration, a non-rigid deformation from the source mesh to the target mesh is approximated by a sum of N^g Gaussian Radial Basis Functions (G-RBF) with centers $\boldsymbol{c_j}$ and appropriate coefficient $\boldsymbol{\omega_j}$ [112]:

$$r(\boldsymbol{p_i^s}) = \sum_{j=1}^{N^g} \boldsymbol{\omega_j} \rho(\boldsymbol{p_i^s} - \boldsymbol{c_j}) \text{ with } \rho = e^{-\left(\mu\|\boldsymbol{p_i^s}\|^2\right)} \tag{3.19}$$

which is subject to the constraints as:

$$\boldsymbol{\delta}(\boldsymbol{p}_i^S) = \sum_{j=1}^{N^g} \boldsymbol{\omega}_j \rho\left(\boldsymbol{p}_i^S - \boldsymbol{c}_j\right), i = 1, \dots, N^g \tag{3.20}$$

The number N^g is smaller than the vertices in the whole surface. The $\boldsymbol{\omega}_j$ can be solved by minimizing:

$$\sum_{j=1}^{N^g} \|r(\boldsymbol{p}_i^S) - \boldsymbol{\delta}(\boldsymbol{p}_i^S)\|^2 + \varepsilon \|\boldsymbol{c}\|^2 \tag{3.21}$$

The additional Tikhonov L2-regularization term $\varepsilon \|\boldsymbol{c}\|^2$ is introduced in case of instability or ill-condition situations. The optimal deformation coefficients $\boldsymbol{\omega}.$ are obtained given displacement $\boldsymbol{\delta}(\boldsymbol{p}.)$:

$$\boldsymbol{\omega}. = (\boldsymbol{\Theta}^T \boldsymbol{\Theta} + \varepsilon \boldsymbol{I})^{-1} \boldsymbol{\Theta}^T \boldsymbol{\delta}(\boldsymbol{p}.) \ with \ \theta_{i,j} = \rho\left(\boldsymbol{p}_i^S - \boldsymbol{c}_j\right) \tag{3.22}$$

The above registration focuses on the single-to-single surface situation and the complexity increases when applied to a multi-organ structure surface. In this structure, the organs are jointed as a whole, and the candidates of deformation are increased as well, which makes the registration more difficult. To verify the feasibility of applying the non-rigid registration algorithm to landmarks corresponding to multiple organs, we designed a series of strategies employing the non-rigid registration to the single or multiple surfaces. The registration can be performed directly on the multi-organ structure, individually between single organs, or their combination.

3.4.2 Experimental Results

3.4.2.1 Data preparation

In the experiment, 30 cases of manually labeled 3-D volume data of human abdominal organs are included in the training from the "Multi-atlas labeling beyond the cranial vault-workshop and challenge". Four of the organs: the spleen, right kidney, left kidney, and liver is regarded as basic single organ element of the multi-organ SSMs. The surfaces

of the organs are generated from the volume data using the surface reconstruction method mentioned in section 3.2.1.2. 30 cases of surface mesh under four simplification levels are used in the registration.

3.4.2.2 Surface registration

The registration of multiple surfaces can be merged into one joint surface or decomposed into matchings between surfaces of single organs. To verify the effect of NICP performed on multi-organ structures, five surface registration strategies are implemented in the experiment. In method 1, the classical rigid ICP is performed to the jointed multi-organ surface, which is also a basic step among all the rest methods. In method 2, a NICP described in [112] is directly performed to the multi-organ structure. In method 3, after rigid ICP, non-rigid registration is firstly applied to the multi-organ structure and the single organs are individually registered. The flow in method 4 is similar to method 3, except for the procedure, where single organ registration is advanced before multi-organ. In method 5, only single organ registration is retained after the global rigid ICP.

Another variate is the simplification level of the surfaces. The deeper the simplification is, the fewer vertices left, and few computing resources are required. We prepared four level $\alpha \in \{1,2,3,4\}$, which retains a ratio of $2^{-(\alpha-1)}$ vertices after the simplification.

To evaluate the registration result, the Hausdorff distance is introduced here. Given two pointsets $A = \{a_1, a_2, \ldots, a_{N_a}\}$ and $B = \{b_1, b_2, \ldots, b_{N_b}\}$, the one-sided Hausdorff distance from A to B is defined as:

$$\tilde{\xi}_H(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \tag{3.23}$$

And the bidirectional Hausdorff distance between A and B is defined as:

$$\xi_H(A, B) = \max\left(\tilde{\xi}_H(A, B), \tilde{\xi}_H(B, A)\right) \tag{3.24}$$

The Hausdorff distance measures the maximum of the distances from each point in A to the closest point in B.

The five registration strategies described before are applied to the training set of 30 abdominal multi-organ surfaces under four simplification levels. The case of reference surface is regarded as the target surface and the rest 29 cases of surfaces in the training set are registered to the target surface and the mean Hausdorff distance of the registration pair is shown in Table 3.1.

The comparison of simplification levels under the same method shows that the simplification levels only have a negligible effect on the registration results. The introduction of non-rigid registration enormously increases the accuracy of registration. Compared with the former methods, the independent registration of single organs in method 4 deforms the surfaces to individual organs before mapping them to the wrong organs in the multiple organ structures and increases the accuracy. In method 5, using more iteration of NICP of single organs instead of multi-organ level registration also increases the registration result.

Table 3.1 Mean accuracy of surface registration

Hausdorff Distance	SL 1	SL 2	SL 3	SL 4
Method 1	21.44±6.36	21.26±6.31	21.12±6.25	21.12±6.25
Method 2	16.26±5.72	15.74±5.44	15.24±5.3	14.62±4.91
Method 3	16.24±7.40	15.55±7.34	14.97±7.17	14.89±6.55
Method 4	13.68±6.44	13.26±6.25	13.34±5.99	13.68±5.86
Method 5	12.37±6.20	11.92±5.90	11.73±5.50	12.00±5.47

Although the Hausdorff distance of registration is improved by introducing the NICP algorithm, some cases of the registration results yield bad matching which may lead to negative effects to the SSMs. An example of registration results with a Hausdorff distance of 28.15 is shown in Figure 3.9. By comparing the source mesh (cyan), the target mesh (green), and the moved source mesh (blue), a large difference after registration can be found in the bottom region of the liver organ.

(a)   (b)

(c)   (d)
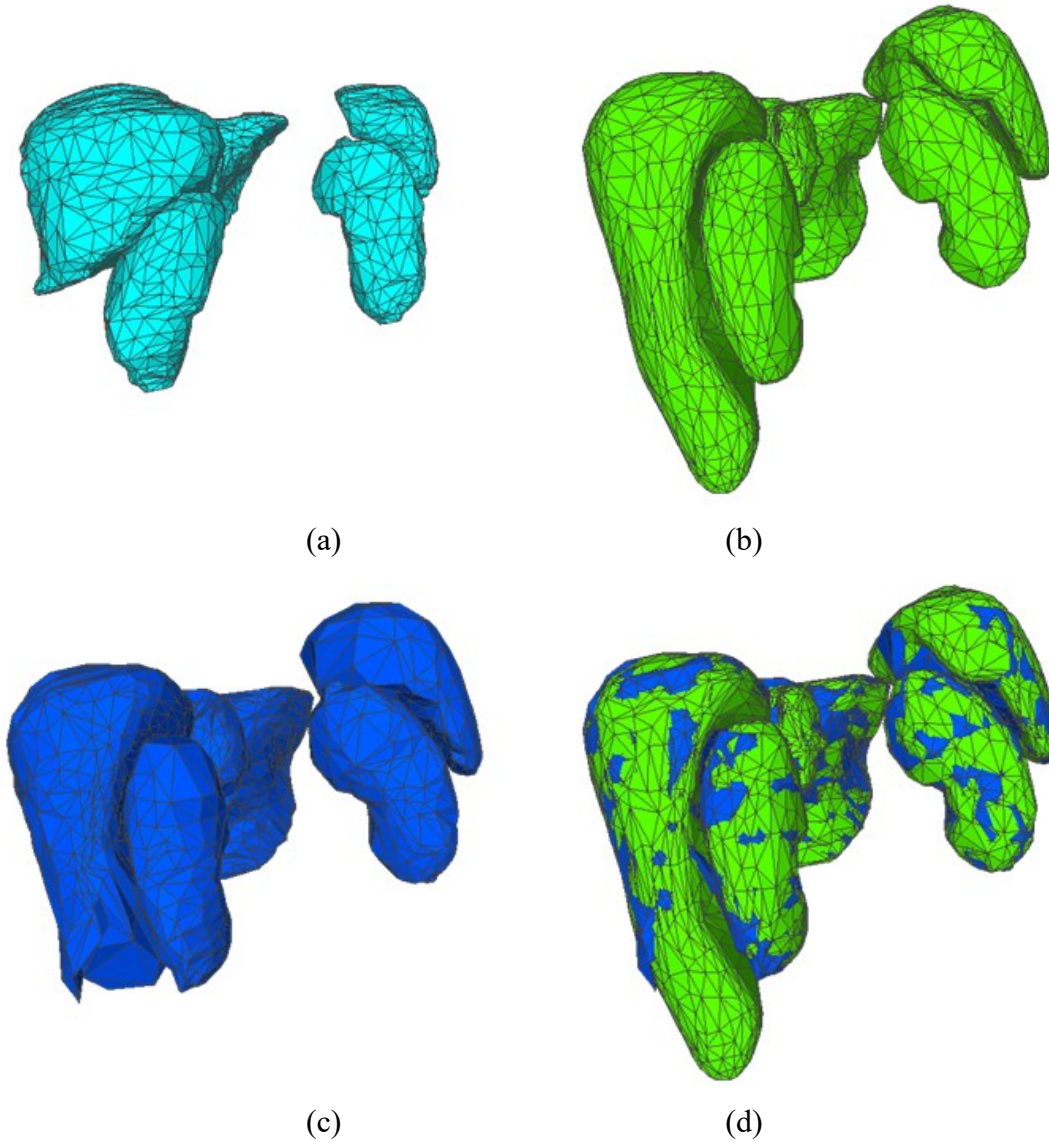
Figure 3.9 An example of registration results with large Hausdorff distance: four surfaces representing (a) source mesh, (b) target mesh, (c)deformed mesh, and (d) a combination of target mesh and deformed mesh. The order of organs shown in each surface is: liver, right kidney, left kidney, and spleen.

3.4.2.3 Shape model building

The SSMs are generated from the landmarks corresponding to the registration of multi-organ surfaces simplified in level 2. To evaluate the quality of SSMs built from the landmarks obtained by five different registration methods, three indicators of model

quality, namely generalization ability, specificity, and compactness, are introduced in the experiment. The generalization ability, specificity, and compactness of four multi-organ SSMs built from landmarks corresponded using five registration strategies are compared in Figure 3.10, Figure 3.11, and Figure 3.12, respectively.

(a)

Figure 3.10 Generalization ability of SSMs. They are built from landmarks corresponded by five registration strategies (1: rigid-ICP of multi-organ, 2: NICP of multi-organ, 3: NICP of multi-organ followed by single organs, 4: NICP of single organs followed by multi-organ, and 5: NICP of single organs); (a) Spleen, (b) right kidney, (c) left kidney, (d) liver. (Cont.)

(b)

(c)

Figure 3.10 Generalization ability of SSMs. They are built from landmarks corresponded by five registration strategies (1: rigid-ICP of multi-organ, 2: NICP of multi-organ, 3: NICP of multi-organ followed by single organs, 4: NICP of single organs followed by multi-organ, and 5: NICP of single organs); (a) Spleen, (b) right kidney, (c) left kidney, (d) liver. (Cont.)

(d)

Figure 3.10 Generalization ability of SSMs. They are built from landmarks corresponded by five registration strategies (1: rigid-ICP of multi-organ, 2: NICP of multi-organ, 3: NICP of multi-organ followed by single organs, 4: NICP of single organs followed by multi-organ, and 5: NICP of single organs); (a) Spleen, (b) right kidney, (c) left kidney, (d) liver.

(a)

Figure 3.11 Specificity of SSMs. They are built from landmarks corresponded by five registration strategies (1: rigid-ICP of multi-organ, 2: NICP of multi-organ, 3: NICP of multi-organ followed by single organs, 4: NICP of single organs followed by multi-organ, and 5: NICP of single organs); (a) Spleen, (b) right kidney, (c) left kidney, (d) liver. (Cont.)

(b)

9
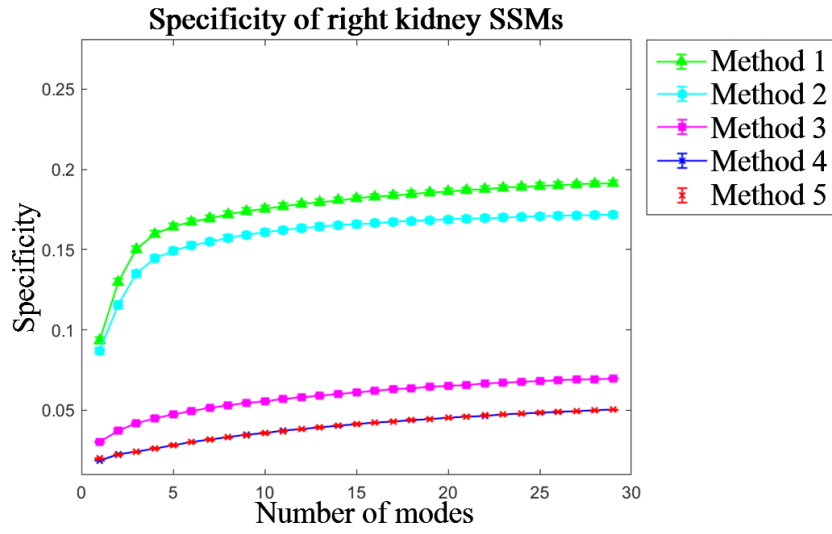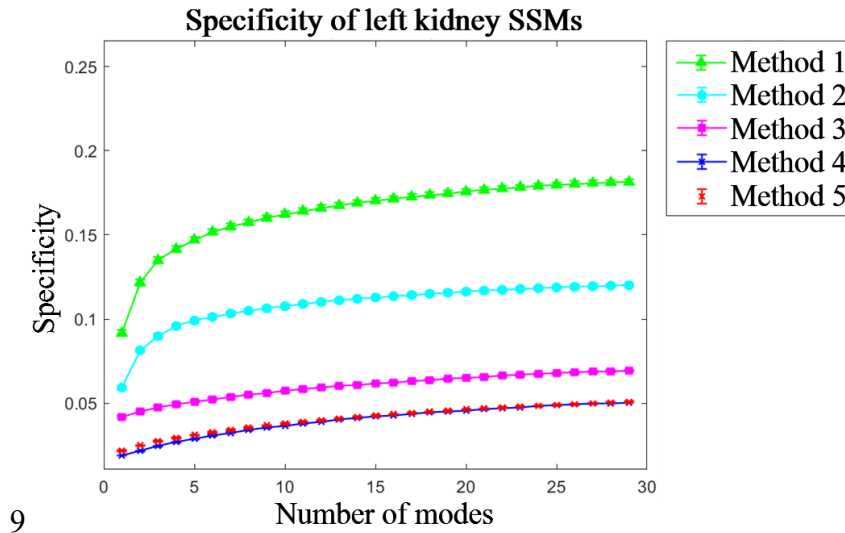
(c)

Figure 3.11 Specificity of SSMs. They are built from landmarks corresponded by five registration strategies (1: rigid-ICP of multi-organ, 2: NICP of multi-organ, 3: NICP of multi-organ followed by single organs, 4: NICP of single organs followed by multi-organ, and 5: NICP of single organs); (a) Spleen, (b) right kidney, (c) left kidney, (d) liver. (Cont.)

(d)

Figure 3.11 Specificity of SSMs. They are built from landmarks corresponded by five registration strategies (1: rigid-ICP of multi-organ, 2: NICP of multi-organ, 3: NICP of multi-organ followed by single organs, 4: NICP of single organs followed by multi-organ, and 5: NICP of single organs); (a) Spleen, (b) right kidney, (c) left kidney, (d) liver.

(a)

Figure 3.12 Compactness of SSMs. They are built from landmarks corresponded by using five registration strategies (1: rigid-ICP of multi-organ, 2: NICP of multi-organ, 3: NICP of multi-organ followed by single organs, 4: NICP of single organs followed by multi-organ, and 5: NICP of single organs); (a) Spleen, (b) right kidney, (c) left kidney, (d) liver. (Cont.)
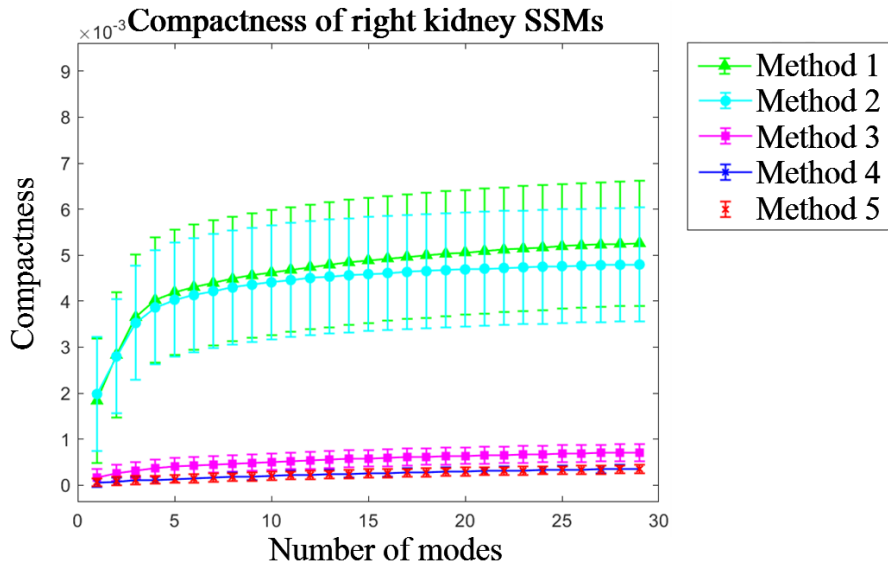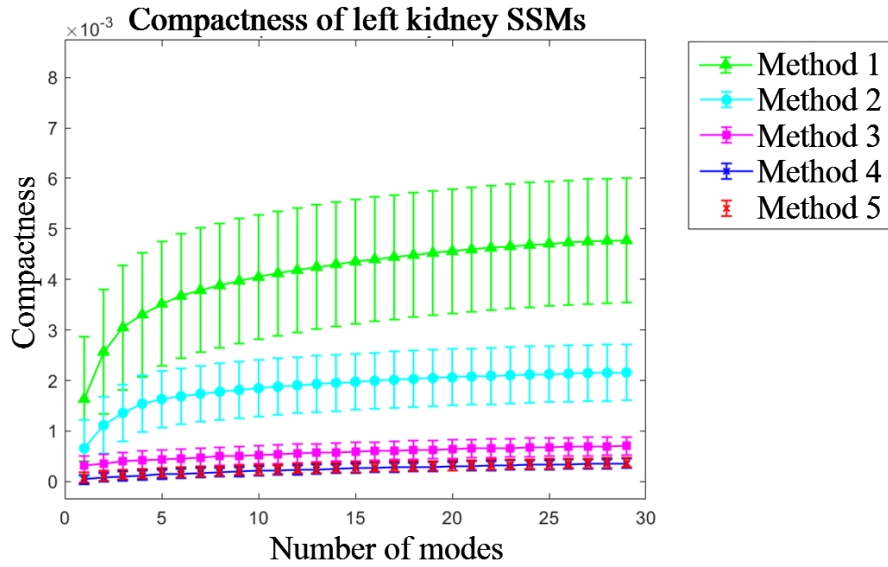
Figure 3.12 Compactness of SSMs. They are built from landmarks corresponded by using five registration strategies (1: rigid-ICP of multi-organ, 2: NICP of multi-organ, 3: NICP of multi-organ followed by single organs, 4: NICP of single organs followed by multi-organ, and 5: NICP of single organs); (a) Spleen, (b) right kidney, (c) left kidney, (d) liver. (Cont.)
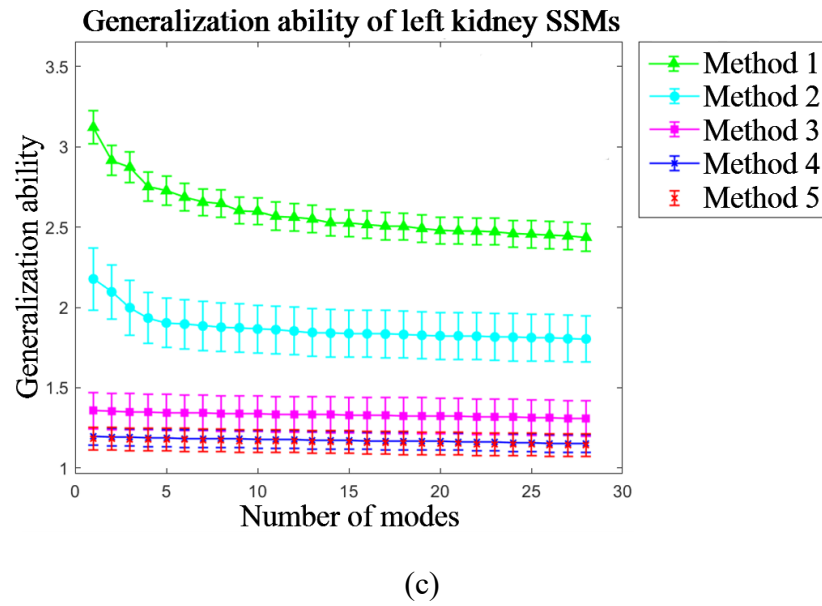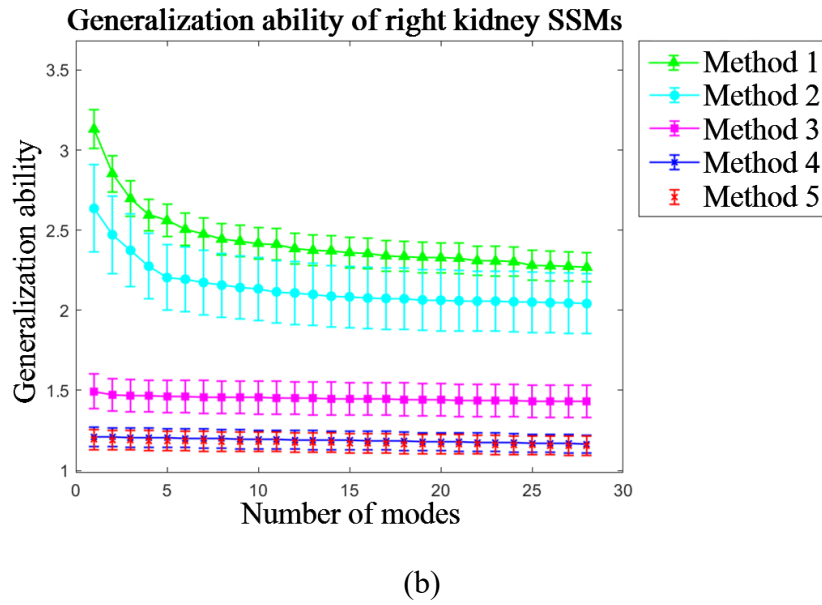
(d)

Figure 3.12 Compactness of SSMs. They are built from landmarks corresponded by using five registration strategies (1: rigid-ICP of multi-organ, 2: NICP of multi-organ, 3: NICP of multi-organ followed by single organs, 4: NICP of single organs followed by multi-organ, and 5: NICP of single organs); (a) Spleen, (b) right kidney, (c) left kidney, (d) liver.
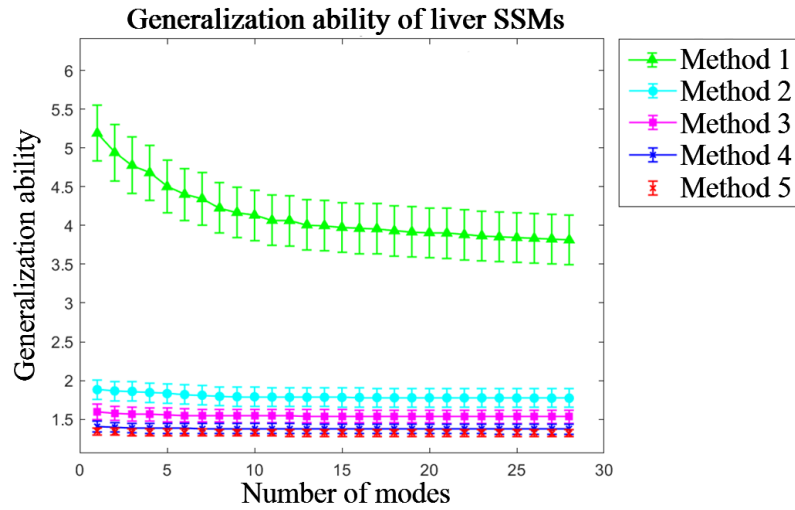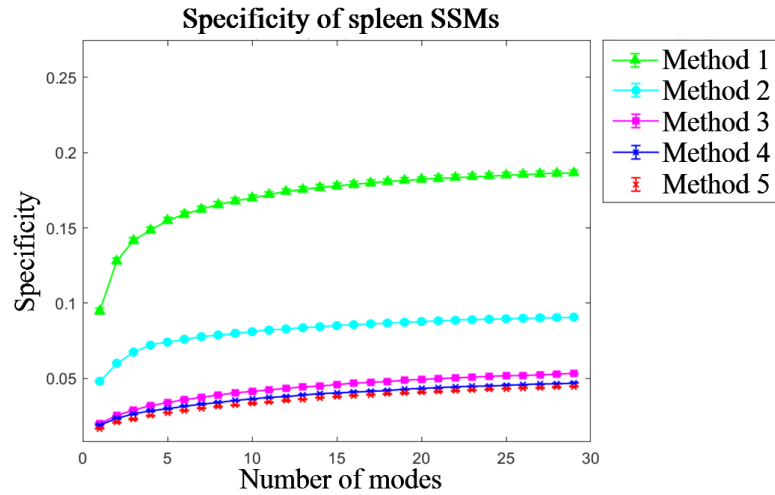
It can be found that the quality of SSMs built is relative to the registration accuracy in the landmarks corresponding procedure and the model qualities are increasing from method 1 to 5, as the accuracy of registration results. Thus, the introduction of the NICP method applied to single-to-single organs registration strategy can benefit the SSMs building process.

3.4.3 Discussion and Conclusion

3.4.3.1 Surface registration

In the experiments of multi-organ surface registration by using five different methods, the comparison of registration accuracy measured by Hausdorff distance shows that the strategy of method 5: a single organ NICP registration pairs, is more beneficial to the registration accuracy, as shown in Table 3.1.

In method 1, a classic ICP method is performed to the registration of jointed multi-organ structures. The large error of the result shows the necessity of introducing non-rigid transforms for the registration of human organs from different patients so that the deformation space is complex enough to describe the non-rigid distortion among different shapes. In method 2, the jointed surface containing four organs is directly matched to the target surface. Although there is no intersection among the four surfaces, as they are constructed from the voxel data of different organs, some of the organs are inevitably anatomically close. In our experiment, the right kidney is posterior to the liver, and the left kidney is posterior to the spleen, which may lead to misregistration when there is no distinguishment among different organs in method 2. The accuracy is improved by introducing the NICP, compared to the ICP method, but there is a problem of misregistration of different organs. In method 3, a single-to-single registration among individual organs is attached to the flow of multi-organ surface registration in method 3. However, the deformation of the surface is misled to the wrong organs in the previous multi-organ surface registration, and the attached single-to-single registration strategy is not able to fix the problem, which makes the accuracy of method 3 similar to method 2. In method 4, the order of the registration process is adjusted, and the single-to-single registration is brought forward. This operation improved the accuracy by the registration between single organs firstly and alleviates the problem of misregistration among anatomically close organs. In method 5, the registration of multi-organ structures is removed and replaced by a complete single-to-single registration strategy, and this strategy achieves the best result.

Compared with the registration of multi-organ structure, where the four organs are regarded as one united surface, the single-to-single strategy provides additional information of organ category and the misregistration of vertices belonging to different organs is forbidden, which improve the accuracy of the multi-organ registration.

In the experiment, four levels of simplification are performed to the original surfaces obtained from the voxel data. The registration accuracy by the same method under

different simplification levels is almost the same, which indicates that the simplification rarely affects the registration accuracy. It can be interpreted as that, in the simplification, the global geometry is nearly not changed despite the number of vertices and faces used to describe the surface being decreased rapidly as the simplification level raises. Although the accuracy cannot be improved by simplification, the use of fewer vertices in the registration can increase the efficiency of calculation and reduce the cost of computation resources.

3.4.3.2 SSMs building

The generalization ability, specificity, and compactness are three indicators to measure the quality of SSMs built from different training sets of landmarks, which also reflect the effectiveness of the corresponding method. The SSMs of four organs which are built from the landmarks corresponded by using five different registration strategies are evaluated and their generalization ability, specificity, and compactness are shown in Figures 3.10, 3.11 and 3.12 respectively. The qualities of SSMs built are increasing from method 1 to 5 and are positively relevant to the accuracy of registration. The SSMs built by using the proposed method of the highest accuracy also possess the highest quality, as the registration is important to the process of finding proper corresponding landmarks on each surface.

3.5 Conclusions

In this chapter, we firstly introduced the general process of SSM building and the assess method for model quality. We introduce the k-means clustering algorithms as a surfaces simplification process, which aims at finding spatially representative prototype of landmark in the reference surface, so that the landmarks corresponded from the other surfaces shares the benefit of such distribution of model mesh. We built SSMs of right and left lungs of and evaluated them with generalization ability, specificity, and compactness. The result of the model building shows that, the proposed method promotes the quality of SSMs building.

We also extend the object of SSMs building to the multiple abdominal organs. The landmarks corresponding based on the NICP registration method. A series of surface registration strategies combining NICP registration based on G-RBF are applied to the landmarks corresponding problem. Five registration strategies of registration involving different single or multiple organs matching processes are conducted in the experiment, which is performed on 30 cases of four abdominal organs. The result of registration and SSMs building shows that the proposed NICP registration strategy can improve the accuracy of registration and increase the quality of SSMs built by the corresponding landmarks from the registration.

Chapter 4

Multi-organ Segmentation

Simultaneous segmentation of multiple organs provides the regions of specific organs in the CT scans. It is essential for tumor detection and segmentation, as the other unnecessary regions in the CT scans are excluded, which can limit the searching space of tumors and eliminate the interference from the outside of the organs.

In this chapter, we introduce two statistical methods for the multi-organ segmentation problem in a coarse-to-fine scheme. A random forest regressor is trained with intensity features to localize the position of organs in the images. Then, a fast segmentation of multiple organs is conducted by the SSMs built in the previous chapter. To obtain precise regions of each organ, a supervoxel-based segmentation method is introduced, where the segmentation problem is conducted as the classification of supervoxels by using a random forest classifier. Four abdominal organs are used in our experiment and the accuracy of segmentation is improved.

4.1 Organ Region Localization Using Random Forest Regressor

4.1.1 Introduction

Medical image segmentation is one of the most basic processing steps in CAD systems, which helps the physician deal with a huge amount of medical image data. Also, the volume data obtained by segmentation is widely used in 3-D organ reconstruction, pathological analysis, treatment planning, surgical navigation, and so on fields. Due to the diversity and complexity of medical images, the ambiguity and inhomogeneity reduce the accuracy of segmentation.

The multi-organ segmentation method is a segmentation framework of multiple organs from a certain area of the medical images. Since Cootes et al. [113] proposed the segmentation method using the PDM, the SSM has been greatly developed in the field of single organ segmentation. Many segmentation models involving multiple organs are mainly used for the segmentation of multiple organs (targets) in the brain, chest, and abdominal CT. Multi-organ segmentation based on gray registration of non-rigid maps [77] is not ideal for CT images acquired under different imaging conditions. The multi-organ segmentation method based on machine learning is mostly applied to the segmentation of the liver, spleen, kidney, and aorta, but not for organs with intersubjectivity in shape and position [78]. Statistical model-based methods can be divided into two categories: one is multi-structured joint modeling [76], with very stable shape and position, but the segmentation accuracy needs to be improved; the other is modeling multi-organ correlation joints and multi-scale modeling [79] improves the accuracy of segmentation and the stability of segmentation. Okada et al. [68] proposed a sequence of multi-organ model architecture, which maps the spatial structure between organs to define the segmentation order of multiple organs in the abdominal CT image, using a probabilistic atlas. And the SSM as a segmentation method, which preferentially divides the organ with good segmentation effect in the multi-organ structure, and then uses the segmentation result to guide the organ with higher segmentation difficulty, improves the accuracy of multi-organ segmentation, and CT images for different conditions.

In addition to the structure of models, the initialization of SSM at the beginning of the segmentation is also very important. Before the segmentation searching, the shape of SSM is required to be initialized into a close range of the target organ in the image. Casually, the appearance model, as a part of SSM, only learns local profiles of intensity within a small range of the model surface. Thus, at the beginning of the searching process, if the model is placed far from the target, it is difficult to move the model to the target when the surrounding voxel intensities are far different from the local appearance model

70

and the searching process may fail into a local optimum which differs a lot to the ground truth. Therefore, a robust initialization method that can initialize the model close to the target organ is very important for the SSM methods.

In the following section, we present an improved SSM segmentation method that solves the drawback of the initialization problem. The proposed method initializes the SSM by using a random-forest-based regression method, which is pre-trained by bounding box parameters and image volumes. In section 4.1.2, the flowchart of the architecture, including the RF regressor and the SSMs searching algorithms are introduced. In section 4.1.3, a multi-organ SSM containing four organs (spleen, right kidney, left kidney, and liver) is built and used for abdominal CT segmentation. A brief discussion of the segmentation result and a conclusion are conducted in section 4.1.4.

4.1.2 Method

4.1.2.1 Processing flowchart

The flowchart of our multi-organ segmentation method is shown in Figure 4.1. First, an RF predictor which can regress the organ position from an image is built and trained with the parameters of the bounding box extracted from the labels and images in the training set. Then, an SSM is built with landmarks generated from labels volume of ground truth and CT image data in the training set, by using the approaches described in chapter 2. After the preparation of the regressor and the SSM, the test image is input. The position of organs in the test images can be predicted by the RF regressor. Then, the SSM is moved the initialized position and matching to the real surfaces by the deformation of its surface mesh.
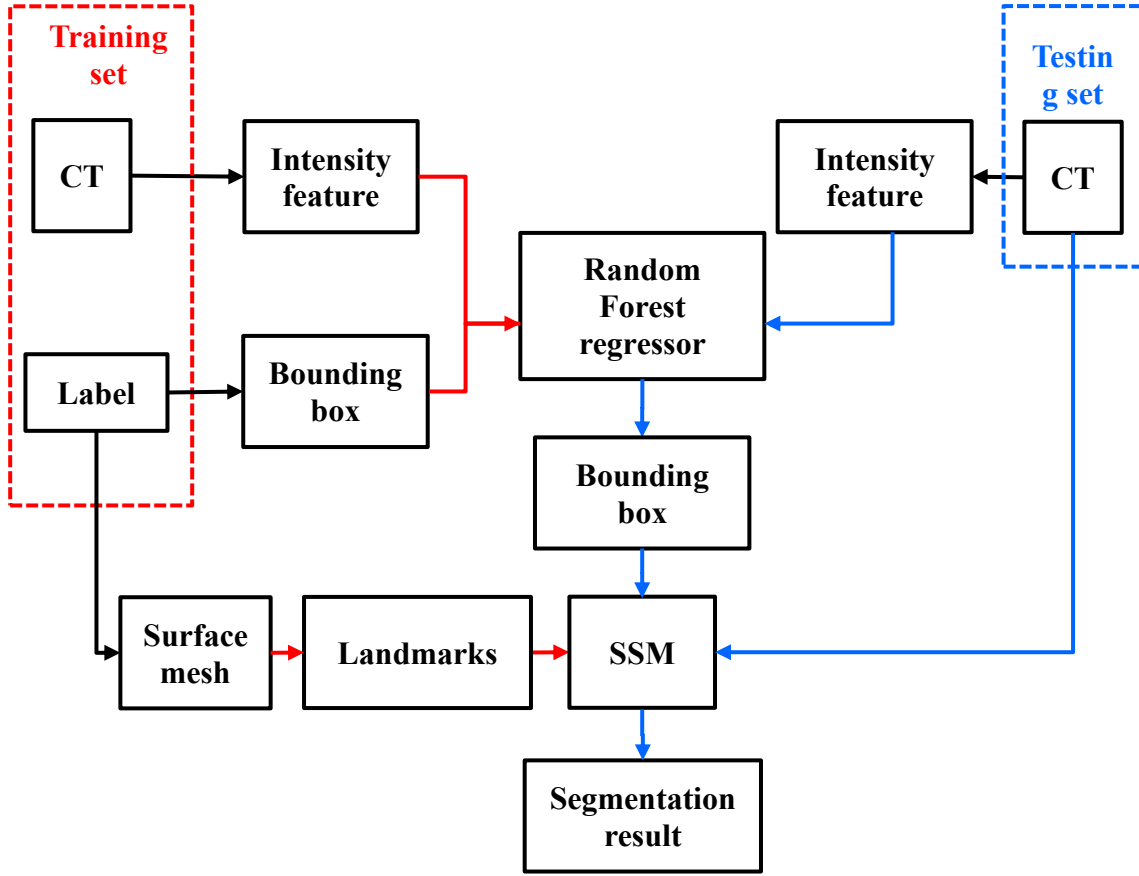
Figure 4.1 Flowchart describing the whole workflow of proposed segmentation method. The red arrows represent the information extracted from the training set, and the blue arrows describe the flow of information from the testing set.

4.1.2.2 Random forest regressor

RF is an ensemble learning method, which combines many base estimators. Based on the bagging algorithm built by decision trees, the random attribution selection is introduced in the training process of training [114]. The process of combining base estimators is very important. A basic method is simple averaging, where similar weight is distributed to all estimations. The manner can be described as:

$$p(\boldsymbol{y}_p|\boldsymbol{v}) = \frac{1}{N_r} \sum_{r=1}^{N_r} p_r(\boldsymbol{y}_p|\boldsymbol{v}) \tag{4.1}$$

where $p_r(y_p|v)$ is the posterior distribution estimated by the rth inducer given evidence v and r is the number of inducers. The final output of the RF is determined by the combining of the decision trees, as shown in Figure 4.2.
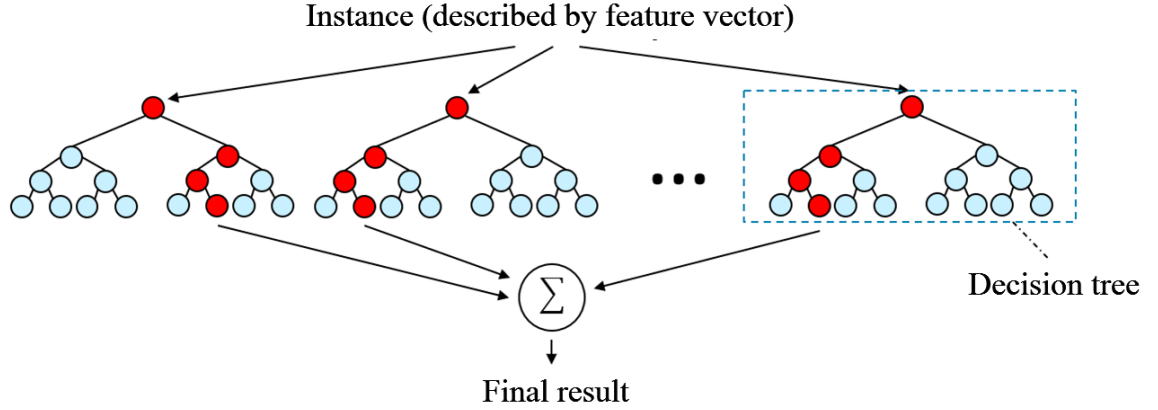
Figure 4.2 Basic structure of a random forest. The final class of the input instance are referring each of the result from individual decision tree.

In our study, the RF method is used as a regression that predicts the location of abdominal organs. In the training process of the RF regressor, a 6-D displacement vector that describes a bounding box for each organ is firstly obtained from the labeled ground truth. Then, the intensify feature of the resized images are synthesized to the RF baggers to find the relationship between images features and bounding boxes of each organ. In the process of segmentation, the trained repressor is used to find the initial location of organs in the given image, before the searching process of SSMs.

4.1.2.3 Segmentation by SSMs

The image segmentation problem based on the SSMs can be equivalent to finding an optimal model similarity deformation parameter and shape model parameter in the image to be segmented so that the deformed model matches the real organ contour.

In the shape space coordinate system, the similarity deformation of the model, including displacement, rotation, and scaling, are all related to each other. The shape of a

model x' in an isotropic image domain can be obtained by the similarity transformation of a model:

$$x' = T_{t,s,\theta}(\bar{x} + \Phi_s b_s) \tag{4.2}$$

where t, s, and θ represent the translation, scaling, and rotation parameters respectively. The Φ_s are the eigen vectors that describe the pattern of deformation and b_s are the weight of the Φ_s. The process of segmentation is to find a set of optimal similarity deformation parameters and shape model parameters, by which the shape model y matches the object organ in the image. This is solved by minimizing the distance between the two shapes:

$$|y - x'| \tag{4.3}$$

In the search process of the model, the similarity deformation parameters are initially adjusted, and the rest of the differences are used to adjust the model parameters. dx_0 is defined as the error between the actual position of the target organ y and the initial model shape x, dt is the displacement error between the center of the mean shape and the center of the target, dx is the error between the target and the shape model after similarity deformation. The geometric transformation method can be used to obtain the displacement error dt, the rotation error $d\theta$ and the scaling factor $(1+ds)$, and the adjustment of the shape model can be calculated:

$$dx = T_{s,\theta}\left(\left(s(1 + ds)^{-1}, -(\theta + d\theta)\right)\right)[y] - x \tag{4.4}$$

where the $x = T_{s,\theta}[x] + dx' - dt$. $T_{s,\theta}$ and $T_{t,s,\theta}$ are the same except for the translation.

The model parameters db can be obtained from dx, i.e., the residual error between the target and the shapes after the similarity deformation by the eigen vector of the model Φ:

$$dx \approx \Phi(db) \tag{4.5}$$

$$db = \Phi^T(dx) \tag{4.6}$$

The two processes continue to iterate until they converge. The segmentation process of the SSM is shown in Figure 4.3 (the image and model is simplified to the condition of 2-D, and the number of landmarks are also simplified to seven), and can be expressed as:

(1) Place the mean shape ($b = 0$ in formula 3.5) of the SSM in the image space, by referring to the position information from the random forest regressor, as shown in Figure 4.3 (C).

(2) Sampling the gray profile of the image along the average normal vector direction of each landmark on the contour of the shape, as shown in Figure 4.3 (D).

(3) Find the optimal location of the landmark with the smallest Mahalanobis distance of voxel intensity in the candidate position and move them to the new positions to obtain a new shape.

(4) Match the old shape model to the new shape, use the geometric transformation method to obtain the similarity deformation parameters, and get the residual error.

(5) Calculate the change of model shape parameters and get the error dx in formula 4.4. The errors are then converted to the model space by using formula 4.6.

(6) Limit the weight parameters within the range $|b_i| \leq 3\sqrt{\lambda_i}$, in case of unexpected deformation.

(7) Remap the model parameters back to the shape space and add the special transforms, as shown in Figure 4.3 (E).

(8) Jump back to step (2) until the shape converges, i.e., the difference between the old shape and the new shape is under a certain value, as shown in Figure 4.3 (F).
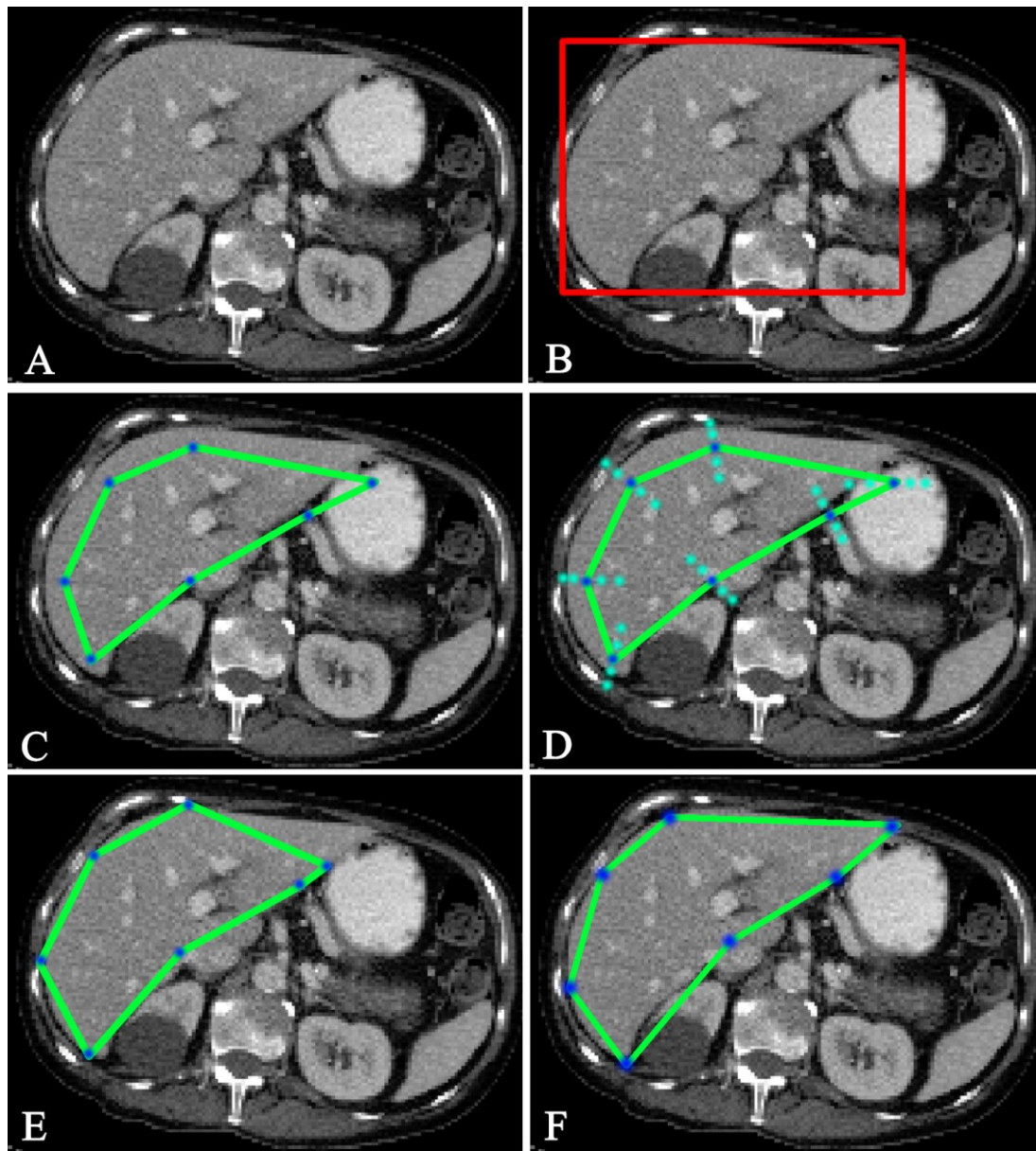
Figure 4.3 A general sketch map of SSMs searching process in 2-D. (A) original image, (B) the bounding box of initial position for the organ, (C) mean shape is placed in the image, (D) the candidates of landmarks to be matched by referring the Mahalanobis distance, (E) the new shape after the matching, (F) the final segmentation result is obtained at the end of the iteration.

4.1.3 Experimental Results

4.1.3.1 Date preparation

The training set of abdominal 3-D CT images, as well as the labeled binary images, are obtained from the "Multi-atlas labeling beyond the cranial vault-workshop and challenge" [115]. The CT image is acquired from the Vanderbilt University Medical Center (VUMC). The label of abdominal organs has been manually segmented by trained raters and the accuracy is checked by a radiologist or radiation oncologist. The CT images were captured during the portal venous contrast phase and the sizes range from $512 \times 512 \times 85$ to $512 \times 512 \times 198$.

The data set used in our experiment contains 26 cases of abdominal CT images and the corresponding labels of each organ. Leave-one-out cross validation is used to build SSMs and RFs, i.e. the dataset is divided into 26 sub sets. In the ith set, the ith case in the data set is chosen as test case and the rest 25 sets are in training set, of which images and labels are used in the SSM building and RF regression process.

The landmarks of organ surfaces are generated by a series of manipulation on the given labels. First, the marching cube algorithms are used to find the boundaries of the labels marked in the images. From this algorithm, the surface of the organs is extracted from the label volumes. Each surface is expressed with a set of sequenced points, called vertices, and a set of triangular faces which are represented by three vertices. Then, the squared edge of the result shapes is smoothed by a Gaussian filter. Then the landmarks are found with a spherical conformal mapping-based corresponding method [102].

The original images are resized to isotropic volumes whose element spacing each is equal to 2mm. Then, a linear mapping from (-128,384) H.U. to (0,255) is performed to remove the irrelevant tissue and emphasize the intensity of interests as well as compact the size of the images, which also help lower the cost of storage and searching space of the algorithm.

4.1.3.2 Evaluation of segmentation

Image segmentation usually includes target recognition and contour delineation. In the evaluation of segmentation accuracy, since it is often impossible to establish a true segmentation result, the gold standard is usually used as a substitute for the true segmentation result. The Dice Similarity Coefficient (DSC) is a commonly used segmentation evaluation standard, which defines the overlap between two sets of binary matrix masks as:

$$DSC = \frac{2|A_s \cap A_g|}{|A_s| + |A_g|} \tag{4.7}$$

where, A_s represents the area of the automatic segmentation result, and A_g represents the ground truth of the manual segmentation result by experts. The DSC ranges in $[0,1]$, and it is generally considered that a value closer to 1 indicates that the two binary masks of the segmentation and the ground truth have a better consistency.

4.1.3.3 Results and Discussion

The segmentation results get from multi-organ SSMs are jointed single organs. The voxels of each organ are marked separately with different labels and compared with labeled voxels' ground truth from each organ.

The mean DSC of segmentation result of abdominal organs, which are the spleen, right kidney, left kidney, and liver, are respectively shown in Table 4.1. The mean accuracy of the segmentation result of each organ is measured by DSC. Compared with the initialization method employed with SSM of which the initial position found by registration method, the accuracy of SSM segmentation initialized with the proposed method is improved.

Table 4.1 Mean DSC of each organ

Hausdorff Distance	Spleen	Right kidney	Left kidney	Liver
Registration method [116]	0.328	0.263	0.355	0.450
Proposed method	0.351	0.310	0.343	0.571

By using RF regression trained by image features, the bounding box of each organ can be found as the initial searching position for the SSM segmentation. The segmentation result shows that the SSM segmentation accuracy is improved by the initialization information provided by the RF method. Future work will focus on more efficient features and parameters of shapes to find more accurate local initial position correction for the SSM searching. A more flexible SSM structure for multi-organ is required as well.

4.2 Refining Segmentation Using Supervoxel

After the coarse segmentation of multiple organs by the SSMs, A precise segmentation is required to provide ROI for the following step of tumor detection and segmentation in the CAD system. We proposed a novel method for automatic abdominal multi-organ segmentation by introducing spatial information in the process of supervoxel classification. Supervoxels with boundaries adjacent to anatomical edges are separated from the image by using the simple linear iterative clustering (SLIC) from the images. Then a RF classifier is built to predict the labels of the supervoxels according to their spatial and intensity features. Thirty abdominal CT images are used in the experiment of segmentation task for the spleen, right kidney, left kidney, and liver region. The experiment result shows that the proposed method achieves a higher accuracy of segmentation compared to our previous model-based method.

4.2.1 Related Works

Medical images are an important information source for the clinical diagnosis of physicians. However, the examination of the medical images is a time-consuming task for

physicians, which may furtherly lead to misdiagnosis and missed diagnosis. With the development of computer techniques and machine learning, CAD has developed into a practicable technique to extract useful information of patients as well as providing more objective opinions for diagnosis of physicians, in addition to their experiment. Organ segmentation is one of the key tasks in CAD among many fundamental medical image processing tasks. The volume data obtained from the segmentation can be used in 3-D organ reconstruction, pathological analysis, disease tracking, and so on in clinical practice. In recent years, a new requirement of multi-organ segmentation is rising along with the evolution from organ-based to organism-based approaches in modern medical diagnosis, and the analysis of multiple organs can also be helpful for comprehensive diagnosis or pre-operative planning and guidance in CAD system [76].

Multi-organ segmentation methods are generally developed from single organ segmentation in which field statistical atlas and shape models are widely used. Shimizu et al. [80] proposed an atlas-guided segmentation method on twelve organs with level-set refinement. Twelve organs are simultaneously extracted from non-contrast 3-D abdominal CT images, by using the abdominal cavity standardization process and segmentation of rough atlas guided segmentation with expectation maximization (EM) algorithm based parameter estimation and the following multiple level set fine segmentation. In [82], Chu et al. used a spatially-divided probabilistic atlas to reduce the inter-subject variance in organ shape and position with the global and local weight assigned and conducted the segmentation incorporating an MAP estimation and a graph cut method. Okada et al. [68] constructed a hierarchical multi-organ statistical atlas with constraints for multi-organ inter-relationships embedded by introducing prediction-based conditional shape–location priors from organ correlation graph (OCG). The predictor organs are pre-segmented and used to guide the segmentation of the remaining organs hierarchically by the conditional shape–location priors. This method increases the accuracy as well as extends the applicability to various imaging conditions without supervised intensity information. These atlas and shape model-based methods show

better robustness in favor of the prior knowledge obtained from the training set. However, more accurate registration is required to match the pre-trained atlas to the test image, and shape correspondence within the training set is necessary for shape models.

Instead of training models, deep convolutional neural network learning techniques are also introduced to this field, benefiting from their outperformed semantic segmentation ability based on the mechanism of feature extraction using multiple convolution layers. Zhou et al. [85] used an FCN to realize a semantic segmentation of nineteen anatomical structures. Roth et al. [117] used a cascaded 3-D FCN to improve the inaccuracies of smaller organs and vessels in a coarse-to-fine approach. For the problem of a small, partially annotated dataset in the deep learning training process weakly supervised training is combined in recent methods. Zhou et al. [91] proposed a Prior-aware Neural Network (PaNN) using anatomical priors on organ sizes and domain-specific knowledge in the training process. In [89], Wang et al. focused on the training sample selection problem instead of network architecture. A Relaxed Upper Confident Bound (RUCB) strategy for sample selection was proposed to mitigate the influence of annotation errors during the training process and increase the segmentation performance. To relieve the inaccuracy on small organs and vessels, which is caused by the imbalance of background and foreground differentiation, and lower layers, the coarse-to-fine strategy was employed in [87] and [88]. The accuracy of the multi-organ segmentation tasks is tremendously improved by the deep neural networks, benefit by the automatically selected features. However, the training of deep networks requires large calculation resources and manually labeled training data, which is difficult to obtain. The overfitting and gradient vanishing are still or even more serious problems for deep neural networks applied to medical image segmentation tasks, especially for 3-D tasks.

To reduce the complexity of methods that directly operated on massive voxels in 3-D images, supervoxel pre-segmentation is introduced to the image segmentation field. Supervoxel is a set of voxels with similar intensities locations and textures, which is separated from a 3-D image volume [14]. In [84], Takaoka et al. proposed a supervoxel

based graph cut method for multi-organ segmentation. Tong et al. proposed a patch-based segmentation framework for the abdominal multi-organ segmentation. Dictionaries and classifiers are used to generate a subject-specific probabilistic atlas and the graph-cuts method is combined. Local information is obtained from local voxel-wise atlas selection to inter-subject variability problems [16]. In [118], Soltaninejad et al. proposed a brain tumor segmentation method from MRI brain image using RF classifier for supervoxel textures. The individually over-segmented supervoxel are the tiniest elements in these methods for a further process, which merges the similar voxels and lowers the requirement of calculation for 3-D image processing.

In this section, we present a supervoxel and RF-based method for automatic multi-organ segmentation from abdominal CT images. After an adjustment of quality, each image is separated into a group of supervoxels and a classifier for them is trained by extracted spatial and intensity features to label the image volume and conduct the segmentation. In section 4.2.2, the proposed method is described, including preprocessing, supervoxel clustering, feature extraction, and RF model. Section 4.2.3 presents our experiment on RF classification and image segmentation results. In section 4.2.4, the method and experiment result are discussed with a conclusion of the research followed.

4.2.2 Outline of Procedure

In this section, an automatic multi-organ segmentation method is introduced, as shown in Figure 4.4. Firstly, a preprocessing on original image data is performed to acquire equidistant volumes and the unified intensity of each voxel. Then, the 3-D images participate into supervoxels, each of which can be categorized as a label representing the background, or other organs. Then, some features of distinctiveness are extracted from the participated supervoxels and fed to a RF trainer to obtain a classifier for supervoxels
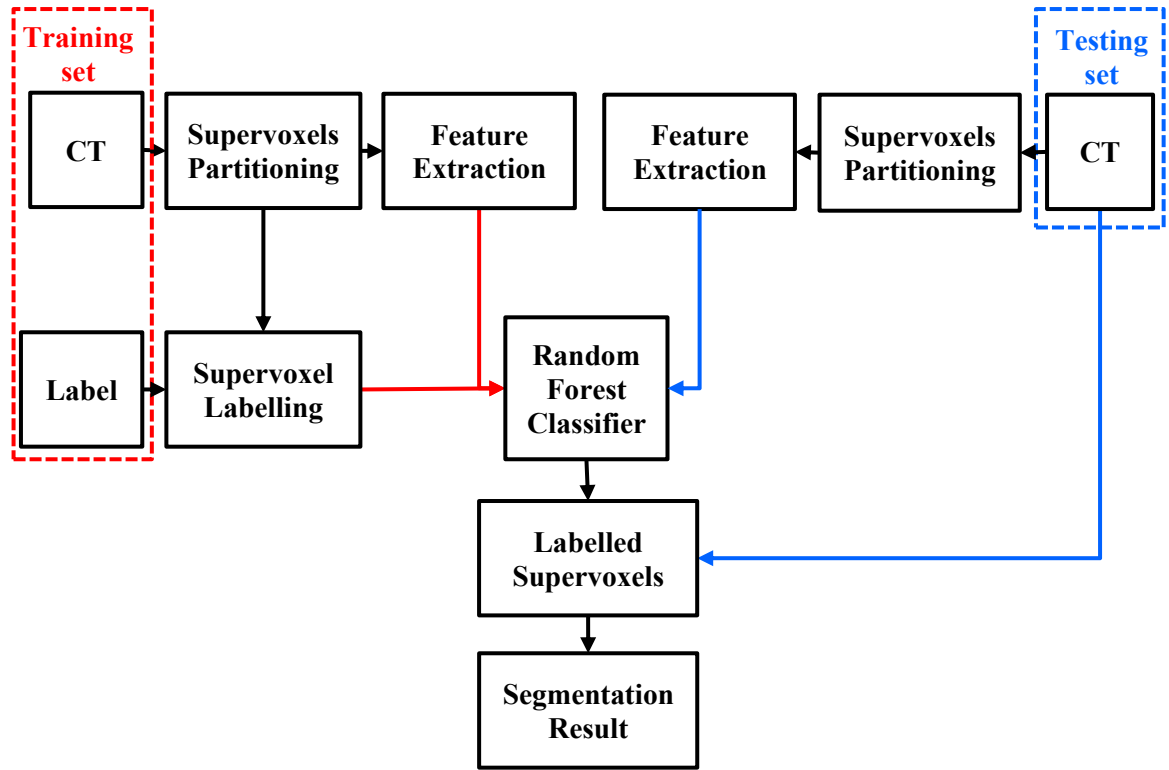
Figure 4.4 Flow chart of the proposed method. The information from the training set are marked with red arrows, and the flow including the information of testing set are represented by blue arrows.

label. Finally, given a test image, each of the participated supervoxels can be labeled by the classifier as background or other organs and the final segmentation result can be obtained by merging the supervoxels of the same label.

4.2.2.1 Supervoxel clustering

After the preprocessing, the images are separated into smaller units of supervoxels. As the CT images data are organized in tensors of rank 3, it is natural and convenient to develop algorithms based on the smaller cubic patches. However, the shapes of anatomical structures are irregular, and the cubic division of the image may lead to the isolation of identical tissues and organs. In this process, many useful connectivity information and communal features of the same region are ignored. Instead of using cubic patches, supervoxel that formed by adjacent voxels of similar intensity can be a

better minimal unit for medical image analysis.

In our research, simple linear iterative clustering (SLIC) [119] is used to generate the supervoxel division. The voxels of an image are clustered into groups of supervoxels by using a k-means method, in which the voxel intensities and spatial positions are used to measure the distance between voxels. The distance of intensity and position between the ith and jth voxel can be defined as

$$d_c = \sqrt{(c_i - c_j)^2} \tag{4.8}$$

$$d_p = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \tag{4.9}$$

where c_i and c_j are the grayscale intensity of the ith and jth voxel; (x_i, y_i, z_i) and (x_j, y_j, z_j) the position of the ith and jth voxel respectively. A comprehensive measurement of the intensity distance d_c and position distance d_p, represented as D is used in practice

$$D = \sqrt{(d_c)^2 + \left(\frac{d_c}{S_{int}}\right)^2 (\varepsilon)^2} \tag{4.10}$$

where S_{int} represents the sampling interval. The ε in (4.4) is a constant that helps balance the importance of intensity and spatial distance, which also influences the irregularity of the obtained supervoxels. When ε is smaller, the edges of the obtained supervoxel are more coincided with the real boundaries of existing tissues or organs, while ε is larger the edges are tended to approach regular grids and the shape of the supervoxel would be more regular.

The number of supervoxel k is decided by the floored quotient of voxel number in the image, N_v, and the interval S_{int}:

$$k_s = \left\lfloor \frac{N_v}{S_{int}{}^3} \right\rfloor \tag{4.11}$$

The clustering of supervoxel is an iterative process. At first, centers of supervoxel are initialized at the voxel of minimal gradient within each equally divided grid of the input image. Then for each center of the initial grids, its distance to each voxel within a range (set as a region) is calculated and the voxels are assigned to the nearest cluster. In each iteration, the centers are recalculated, and the voxels are reassigned until the residual error converges.

4.2.2.2 Feature extraction

After the supervoxels are obtained, suitable features are required to distinguish supervoxels as the organs or background. In our method, the spatial position and statistical intensity features are considered beneficial for the supervoxel classification.

As the abdominal organ of interest shares, similar anatomical relations in the abdominal cavity, the same organ in different cases of patients processes an approximate position. We introduce the position of the central voxel from each supervoxel as the spatial feature, which is previously normalized according to the image size. Also, the intensity features are varied for supervoxels that belong to different tissues, organs or so on matters. While the difference of intensity among different images is reduced in the preprocessing, a gradient map of the image can furtherly help decrease the influence from the data diversity caused by many imaging environments. The maximum, minimum, and mean of the intensity value and gradient value are calculated as intensity features. Three position features, three intensity features, and three intensity gradient features form the feature vector for each supervoxel.

4.2.2.3 Random forest classifier

After the supervoxels are obtained, suitable features are required to distinguish supervoxels as the organs or background. In our method, the spatial

In this section, the RF [120] is used as a classification model, which is different from the regressor described in section 4.2.2.2. The whole training set are firstly sampled into sampling sets, each of which are composed of random samples. Then decision trees are trained separately from the sampling sets. In the training process of the decision trees, the optimal attributes in each node are selected from a random sampled subset of the attribute set. For the multiclass classification task, the result can be obtained by a voting from the results of each decision tree. The randomness from the bagging and attributes helps increase the generalization ability.

In our method, each set of training data for a RF classifier contains supervoxels obtained from all the image data. The attribute space is a nine-dimension vector containing nine features extracted by using the method described in section 4.3.2.2. The labels of each supervoxel, which indicate the category that the current supervoxel belongs to, are decided by the maximum of the voxel labels within the supervoxel. The label space contains five labels: background, spleen, right kidney, left kidney, and liver. In the segmentation procedure, an abdominal image is firstly separated into supervoxels by using the SLIC algorithm, and the supervoxels are classified and labeled as corresponding background or organs with the trained RF classifier. The labels of the supervoxels from the RF are distributed to their voxels and a labeled multi-organ segmentation result is obtained.

4.2.3 Experimental Results

4.2.3.1 Date preparation

In our experiment, a dataset from the "Multi-atlas labeling beyond the cranial vault-workshop and challenge" was used to evaluation the proposed method. In the dataset, 30 cases of abdominal CT images were acquired from the Vanderbilt University Medical Center (VUMC). Thirteen abdominal organs were manually labeled and the labels of voxels which were not organs of segmentation targets were excluded in the experiment, i.e., organs labeled except for spleen, right kidney, left kidney or liver, are relabeled as

background.

4.2.3.2 Supervoxel classification

The supervoxels partitioned from each image were classified by the RF multiclass classifier into five labels: 'background', 'spleen', 'right kidney', 'left kidney', and 'liver' from the training set. To measure the performance of the classifier, several evaluators were used on the classification results. As the research focuses on organ segmentation and there is an extreme imbalance of supervoxel number of organs or background, we only analyze the relative measurement on four organs, except for the background. For each organ, four basic measurements counting numbers of correctly or wrongly classification samples (supervoxels) are used: TP (true positive), FN (false negative), FP (false positive), and TN (true negative). The definition is as below:

TP: number of supervoxels belong to the organ and were correctly classified as the organ; FN: number of supervoxels that belong to the organ but were wrongly classified as the other organ or background; FP: number of supervoxels that do not belong to the organ but were wrongly classified as the organ; TN: number of supervoxels do not belongs to the organ and were correctly classified as the other organ or background.

The accuracy and specificity are used to evaluate the classification of supervoxels. The evaluators referred before are calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.12}$$

$$Specifity = \frac{TN}{TN + FP} \tag{4.13}$$

The mean accuracy and specificity are shown in Table 4.2

4.2.3.3 Organ segmentation

To evaluation the result of the segmentation task, the DSC is used to gauge the similarity of the segmentation by our method and the ground truth from the training set. In the experiment, we compared the segmentation result with our previous research using a RF regressor and SSM. The mean DSC of the spleen, right kidney, left kidney, and liver segmentation results are shown in Table 2. Compared with the previous method in section 4.1, the segmentation accuracy is increased by using the proposed method.

Table 4.2 Mean accuracy and specificity

Organ	Accuracy	Specificity
Spleen	0.6711	0.9990
Right Kidney	0.3822	0.9997
Left Kidney	0.3168	0.9997
Liver	0.7891	0.9997
Spleen	0.6711	0.9990

Table 4.3 Mean DSC of each organ

Organ	Previous method	Proposed method
Spleen	0.351	0.635
Right Kidney	0.310	0.443
Left Kidney	0.343	0.370
Liver	0.571	0.808
Spleen	0.351	0.635

4.2.4 Discussion

The supervoxel clustering is a crucial step in the method. A fundamental requirement for an appropriate supervoxel is that its boundaries must fully cover the edges of organs. As the supervoxels are the tiniest inseparable element in our method, voxels within the supervoxel must process a communal label. That is to say, each voxel on the edges of the ground truth must be included in the set of all the supervoxel boundaries. Hence, a larger number of over-segmented supervoxel is required.

The preprocessing for supervoxel separation focuses on preserving edge information and individual intensity information of each small supervoxel, as well as dividing the region of diverse intensities so that the supervoxel clustering can capture tiny edges and distinguish regions of different organs. Hence, image smoothing operations that can remove noise are not performed here but introduced in the feature extraction stage.

Feature extraction is another key step in supervoxel classification. Proper features can be used to distinguish supervoxels belonging to different organs. In abdominal CT images, the anatomical knowledge is learned by the correspondent relevance between fixed adjacent tissue and organs which are represented by the normalized spatial coordinate feature. However, with the widespread existence of physical differences and body postures, the current spatial feature can only provide information within limited precision. More complicated structural relativities combining adjacent supervoxels can be more effective.

Two groups of segmentation results (on the right) and ground truth from manual delineation (on the left) are compared in Figure 4.5. In the top group, a redundant tissue belonging to the background is misclassified as left kidney, which is adjacent in position and similar in intensity. In the bottom group, some of the regions belonging to the liver and spleen are misclassified as background, even the texture of the missed regions is similar. This is probably influenced by the inaccurate spatial features that arise from the unaligned images, where improvement is required in the future.
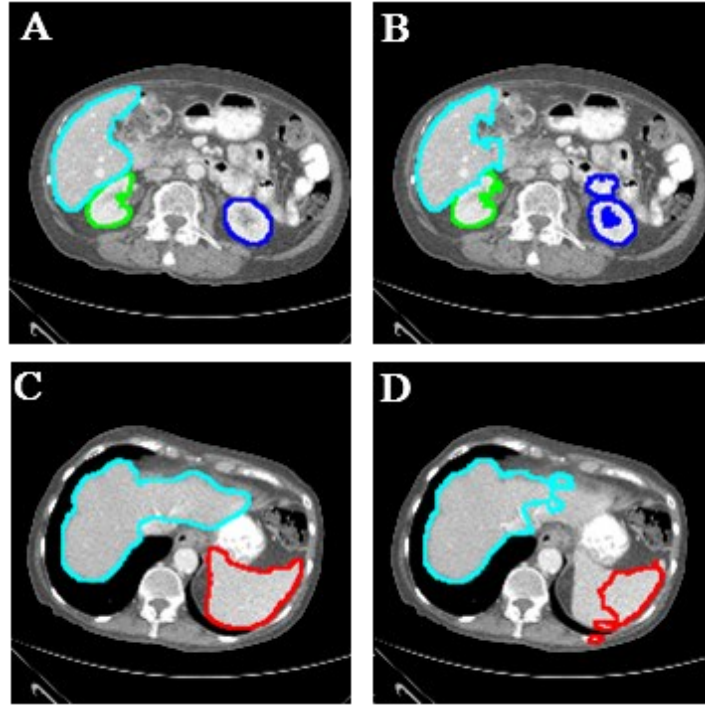
Figure 4.5 Comparison of ground truth (A) (C) and segmentation result (B) (D). The spleen, right kidney, left kidney, and liver are declined in red, green, blue, and cyan, respectively.

Except for alignment in the preprocess of images, feature extraction is another important process in this frame. The feature used in the proposed method relies on statistical intensity and their gradient value with limited size. To extract more representative features and distinguish supervoxels better, some more complicated features extraction method that combines local intensity and gradient from different directions can be applied to the supervoxel. As the 3-D supervoxels are irregular shapes that contain voxels of uncertain quantity, a proper improvement is required to fit the traditional feature extraction method that is regularly applied to square or cubic cells of images.

When not focusing on the feature of individual supervoxels, the global anatomical priors can also be introduced to the classification and segmentation scheme. In the proposed method, only the spatial position of every single supervoxel is considered as a

feature. However, for each of the supervoxels that belong to a specific organ, it is adjacent to at least one supervoxel of the organ. The spatial relevance of these adjacent supervoxels can be taken into consideration, which helps maintain the integrity of the organ.

The anatomical priors can also be used to refine the segmentation result after the supervoxel classification by referring to statistical atlas which comprises the shape or appearance of organs. Given a coarse segmentation result from the classification, like shown in Figure 4.5, a matching is required to fusion the anatomical shape models or atlases. In this procedure, the inter-individual variability is still a challenge and statistical atlas with adequate specification ability are required.

In this section, a RF classifier-based method is proposed for multi-organ segmentation from abdominal CT images. The images are clustered into small units of supervoxels with similar intensities and positions. A group of spatial and intensity features is extracted to distinguish supervoxels by using a RF classifier. The experiment result shows an improvement in segmentation accuracy especially for the spleen and liver compared to our previous shape model-based method. To reduce missed segmentation, more distinguishable features are required to represent the differences of object organs and background in future work.

4.3 Conclusions

In this chapter, we discuss the problems of multi-organ localization and segmentation. An RF regressor is trained to find the initial position of the object organs, which is coded in a bounding box. Then, the segmentation of object organs is accomplished by the deformation of SSMs. In the experiment, the proposed method was applied to locate the position of multi-organ structure, containing spleen, left kidney, right kidney, and liver, from the upper abdominal CT images. The segmentation result shows that the accuracy of segmentation is improved by the initialization information provided by the RF regressor.

The RF are used to localize the organs and the SSMs provide a fundamental result of fast but coarse segmentation. To obtain finer segmentation of organ regions for further steps in the CAD system, we also propose a supervoxel-based multi-organ segmentation architecture. The segmentation is conducted by the classification of supervoxels based on their intensity and position feature. The experiment result shows an improvement in segmentation accuracy, especially for the spleen and liver compared to the previous SSM-based method.

Chapter 5

Liver Tumors Segmentation

The segmentation of liver tumors provides the physician with aiding information for the diagnosis of liver cancer. As the difference of human tissues, including the vessels and tumors in the liver regions, can be enhanced in the multiphase scans, the analysis of multiphase CT images enabled an improved detection of liver tumors. However, tumor regions and peripheral tissues are difficult to distinguish and delineate owing to their highly similar image features. Moreover, their characteristics vary significantly in different phases. This is challenging when using segmentation methods that are based on unique training models. Herein, a hybrid statistical framework is proposed for liver tumor segmentation in multiphase images. We first develop a cascade region-based convolutional neural network with a refined head to locate the tumors. Meanwhile, phase-sensitive noise filtering is introduced to refine the segmentation conducted by a level-set-based framework. This method is sensitive to the intensity contrast but not to the regions of interest, thereby affording better performance in delineating adjacent tumors.

5.1 Introduction

According to data from the Global Cancer Statistics 2020, liver cancer ranks third among the causes of neoplasm-related deaths worldwide [121]. Medical-image-based technology provides a conventional and valid non-invasive procedure for liver cancer diagnosis. CT imaging is one of the most typically used modalities for this task. The detection of liver tumors from CT is essential step in the diagnosis. However, tumors appear similar to the surrounding tissues, which increases the difficulty of tumor detection. To enhance the appearance of different tissues in a CT image, a multiphase CT angiography technique is introduced to provide physicians with multi-degree information

as well as tissue differences based on the temporal resolutions [122]. The imaging time of CT is sufficiently short, thereby allowing each of the phases to be captured at a high resolution within a few seconds. A series of phase signals, i.e., non-contrast-enhanced (NC), arterial (ART), portal venous (PV), and delay phase signals, are obtained successively after contrast injection [123]. These characteristic variations from multiphase CT images allow radiologists to distinguish liver tumors [124-126].

Although the appearance of tumors is enhanced in a multiphase CT image, when compared with CT images without contrast enhancement, the detection of liver tumor regions remains challenging because of the wide variety of tumors appearing in the images from different patients. In fact, the intensity, texture, shape, size, and other features of the tumors are similar to their surrounding tissues or other independent tumors.

Recently, deep learning techniques have been introduced to medical image analysis and have provided new solutions for the segmentation task of liver tumor regions in multiphase CT. Conze et al. used the RF classifier and multiphase supervoxel-based features to segment hepatocellular carcinoma liver tumors [127]. The multiphase visual features extracted from registered multiphase CT images have been proven to improve the accuracy of the classification result. Sun et al. [128] considered distinctive pathological information from different phases. They proposed a multichannel fully convolutional network with separate FCN channels for each phase individually. The features of liver tumors, which involve different phases, are extracted in the deeper layers to improve the accuracy of segmentation. The efficacy of information and the features pertaining to the liver tissue from different phases can be corroborated in the study by Ouhmich et al. [129], where a dimensional multiphase was implemented on NC, ART, and PV slices. Linked single-phase images were formed as a feature map of multiple dimensions. Meanwhile, multiphase fusion was applied. The outputs from the single phases were combined to obtain the final results. To mine complementary information across images of multiple phases, a modality-weighted U-Net [130] was proposed to combine features learned from different phases using dynamically weighted feature maps.

94

Meanwhile, Xu et al. substantiated the efficiency of information extracted from multiple phases in a residual network [131]. They used a phase-attention residual network to exploit the ART phase and then extracted tumors from the PV phase. The phase attention modules were separated to obtain channel-wise and cross-phase interdependencies. In addition, to enhance sensitivity to the edges, a 3-D boundary-enhanced loss was used in the training.

However, these methods present two disadvantages. First, because the tumors appear tuberous, multiple tumor regions are likely to aggregate and form a single region. When the gap between the tumor regions is extremely narrow, as shown in Figure 5.1, the segmentation method may fail. Second, insufficient data quantity restricts the application of relatively deep learning methods in medical image processing, particularly in abdominal multiphase CT images. Herein, a hybrid method that incorporates a deep learning method is proposed to improve the accuracy of liver tumor segmentation. A cascade region-based convolutional neural network (R-CNN) with expanded head sections was introduced to detect tumor regions separately. Based on the regions, tumor boundaries were segmented using an improved level-set algorithm.
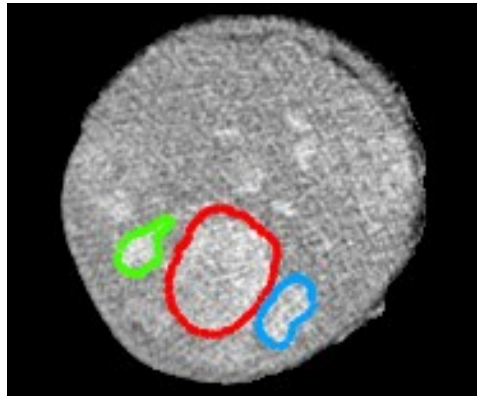
Figure 5.1 An example of adjacent tumors which are difficult to separated. There are three tumors marked in green, red, and blue. The distinguishment between the pair of adjacent tumors (in red and blue) are difficult as their boundaries are of large similarity.

In section 5.2, the approaches relative to tumor detection are provided. Section 5.2.1 reviews research that focuses on liver tumor segmentation. In section 5.2.2, a five cascade R-CNN is introduced to realize the detection of regions containing live tumors. Section 5.3 introduced the tumor segmentation approach. The level set method by fast marching is applied to delineate the contours of the tumor regions in section 5.3.1. The strategy of bilateral filtering is explained in section 5.3.2, to improve the edge of the tumors in the images. The Experiments and results are listed in section 5.4. In section 5.4.1, the dataset of multiphase slices is introduced. The evaluation approaches of liver detection and segmentation are described in section 5.4.2 and their results are listed in section 5.4.3 and section 5.4.4, respectively. The relevant results are discussed in section 5.5. Section 5.6 provided the conclusion of the proposed architecture and the future works.

5.2 Coarse Segmentation Using Improved R-CNN

5.2.1 Related Works

In the early stage, conventional methods, including region growing, the watershed algorithm, and machine-learning-based methods, have been introduced to solve problems pertaining to liver tumor segmentation from CT images [132-137]. However, the segmentation performance is affected by the custom-developed feature extractors. Recently, deep learning methods, which are characterized by convolutional neural networks, have been applied extensively and outperform state-of-the-art methods in terms of tumor segmentation. Christ et al. [138] used cascaded FCNs to segment liver regions and lesions separately, and they discovered that pathological characteristics can be acquired from medical data of different phases. Li et al. [139] proposed a hybrid framework for segmenting liver regions and tumors, where intra-slice features were extracted using a 2-D DenseUNet and merged with inter-slice features learned from a 3-D DenseUNet. By integrating 2-D feature maps, Yan et al. [140] proposed an R-CNN, in which 3-D context information is incorporated with 2-D feature maps. However, these

methods are conducted in multiple detection stages. To reduce the uncertainty of tumor edge effects from the U-Net, Seo et al. [141] added a residual path to the skip connection of the original network. This method enables the pooling quantity to be refined such that advanced features representing global information can be learned regardless of the object size. In [142], a volumetric attention model was proposed for 3-D liver and lesion segmentation. Meanwhile, the authors of [143] segmented the liver region using a 3-D U-Net. Subsequently, the candidates of tumors were obtained using the superpixel-based multi-scale candidate generation method. The candidates were classified by a 3-D fractal residual network, and the combined result were refined using an active contour model. However, the superpixel-based method results in mis-segmentation when multiple tumors are extremely close to each other. Chen et al. proposed an adversarial densely connected network for liver tumor segmentation, and the accuracy was improved by introducing an adversarial training strategy [144]. Based on the basic architecture of the U-Net, Ayalew et al. adjusted the filter and layers; however, small and irregular tumors still affected the accuracy [145]. The multiscale features proposed in [146] provided a better description of liver tumors with additional contextual information. In a previous study [147], the U-Net was set as the backbone, and a bi-directional convolutional long short-term memory was used for feature extraction and fusion across different slices.

However, the methods mentioned above were designed and applied to single-phase CT. In multiphase CT, the appearance of tumors and tissues is enhanced, and the consequent errors caused by the misdetection of multiple adjacent tumors affect the accuracy of liver cancer diagnosis. Hence, the region of interest (ROI) of individual tumors can be extracted from the liver region prior to the precise delineation of tumor regions.

R-CNNs are object detection frameworks comprising two main steps: region proposal and object classification [148]. The image features of different objects are learned from a convolutional neural network, and their offsets are adjusted via bounding box regression. In the first study pertaining to the R-CNN, the maximum number of region proposals was 2000, and these region proposals were classified using a support vector machine. Object

detection using the R-CNN is unsuitable for real-time applications as it achieves a forward pass of the convolution network for each of the proposals. Some regions of object proposals are overlaid, and the computation of features must be duplicated [149]. The fast R-CNN [149] computes a feature map of the entire image, and the proposal is classified directly by the local section of the feature map. To improve the region proposal generation, the convolution features of the entire image are shared in the region proposal network (RPN) of the faster R-CNN [150]. The region proposals are mapped to a set of reference anchor boxes, which are represented by rectangles of different sizes. This procedure further reduces the computational time efficiently.

Based on an image as the input, an RPN structure provides a series of region proposals and corresponding assessments of similarity. A common set of RPN computations is shared with the main object detection network branch. A mini network slides along the feature map shared by the previous CNN layers. In Figure 5.2, the red sliding windows are slid on the feature map extracted from the main flow of the CNN. Subsequently, k anchor boxes are applied to the sliding window, and "whether or not it is regarded as an object" is output from the "cls layer," and "how much it deviates from the correct area" from the "reg layer." The "cls layer" outputs the probabilities pertaining to the existence and non-existence of the object in the current anchor box; if it exists, the "reg layer" outputs the position information of the box, i.e., the center coordinates, width, and height. The number of anchor boxes is nine (combination of scale: 8, 16, and 32 pixels; aspect ratio: 1: 2, 1: 1, and 2: 1).

According to Cai et al. [151], as compared with the solo R-CNN, a cascade architecture can solve the problems of overfitting during training caused by vanishing positive samples. They indicated that the effectiveness of detection deteriorated when the intersection over union (IoU) threshold increased. Hence, we herein propose an object detection architecture for liver tumor segmentation involving a multistage process, which comprises a series of detectors with different IoU thresholds to prevent close false positives.

5.2.2 Five Cascade R-CNN Structure

The CNN architecture used in the proposed network for feature extraction is presented in Table 5.1. This architecture involves an $n \times n$ convolutional layer (CL), followed by two 1×1 sibling CLs. Rectified linear units are connected to the results from $n \times n$ CLs. In the head section, the position and category of the object are specified using the result obtained from the pooling layer of the ROI. As shown in Figure 5.2, the two FC sections are fully connected with 1024 units. In the classification block, the same quantity of units as classes are fully connected, and the probability of each region proposal is obtained using the softmax function. In the box regression block, the number of fully connected units is four times the number of classes, and the position vector (center coordinates and size of the box) for the region proposals is obtained.
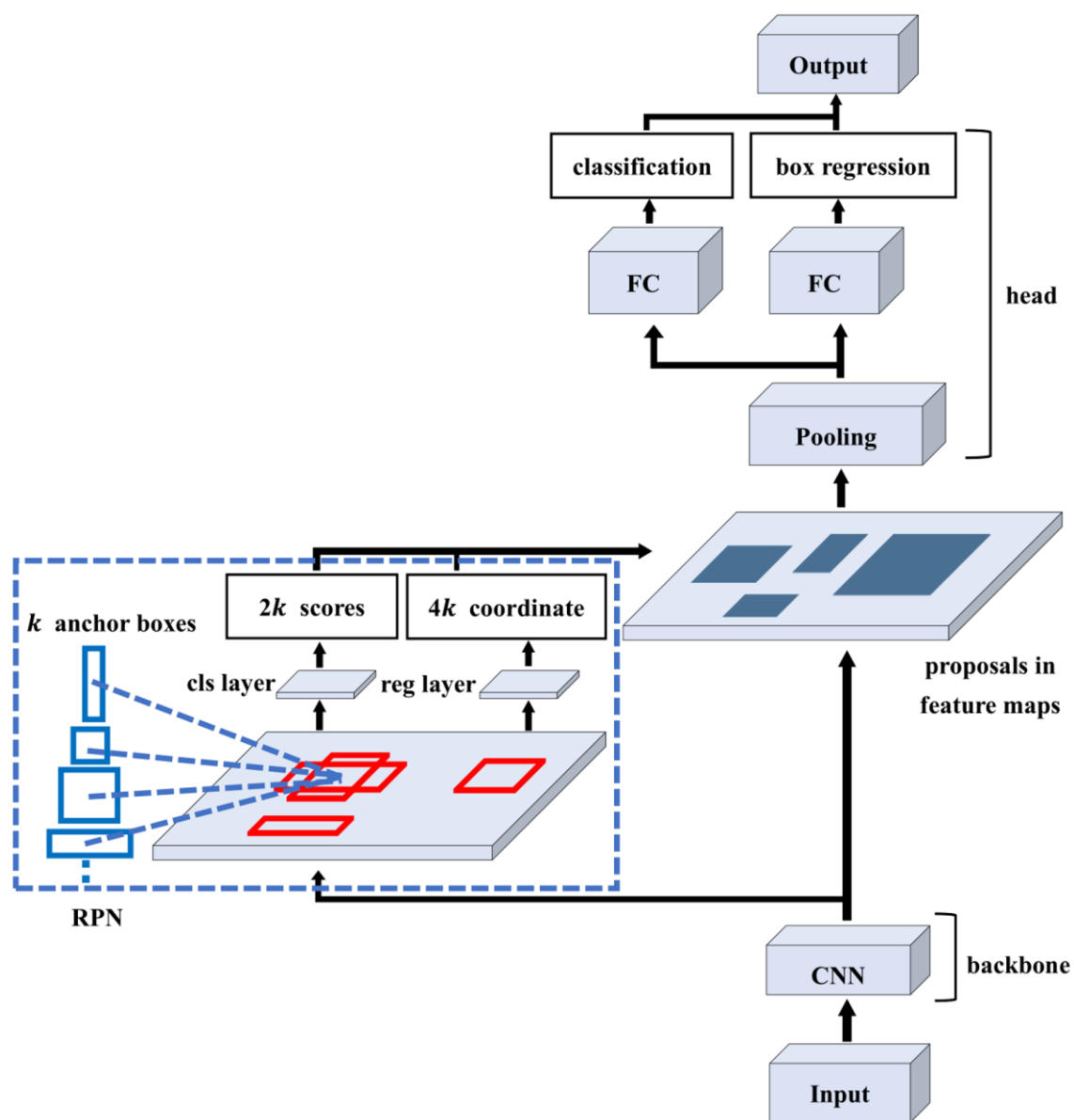
Figure 5.2 Architecture of RPN in faster R-CNN. (FC: fully connected layer).

Table 5.1 Architecture of ResNet101 used in proposed network

Layers	Output Size	Structure
CL-1	112×112	$[7 \times 7, \text{channel } 64, \text{stride } 2] \times 1$ $[3 \times 3 \max \text{pooling}, \text{stride } 2] \times 1$
CL-2	56×56	$\begin{bmatrix} 1 \times 1, \text{channel } 64 \\ 3 \times 3, \text{channel } 64 \\ 1 \times 1, \text{channel } 256 \end{bmatrix} \times 3$
CL-3	28×28	$\begin{bmatrix} 1 \times 1, \text{channel } 128 \\ 3 \times 3, \text{channel } 128 \\ 1 \times 1, \text{channel } 512 \end{bmatrix} \times 4$
CL-4	14×14	$\begin{bmatrix} 1 \times 1, \text{channel } 256 \\ 3 \times 3, \text{channel } 256 \\ 1 \times 1, \text{channel } 1024 \end{bmatrix} \times 23$
CL-5	7×7	$\begin{bmatrix} 1 \times 1, \text{channel } 512 \\ 3 \times 3, \text{channel } 512 \\ 1 \times 1, \text{channel } 2048 \end{bmatrix} \times 3$
-	1×1	Average pooling, fully connected layer, softmax

Owing to overfitting in deep learning frameworks, the distribution of hypotheses generated from a proposal detector is likely to be of low quality, even when the IoU thresholds are set larger. Moreover, the mismatch between the detector performance and the quality of the testing hypotheses at inference is difficult to solve. In the conventional faster R-CNN, an anchor box with IoU ≥ 0.5 or higher output from the RPN is learned by one head. However, the accuracy from one head is not necessarily high for all IoUs. In other words, if an anchor box with a low IoU (e.g., 0.5) is learned with the head, it becomes difficult to predict the position with a high IoU. In addition, if learning is performed only with an anchor box with a high IoU (e.g., 0.7), a prediction with a low IoU will be difficult to perform. Therefore, in the cascade R-CNN, three heads with different IoU thresholds are prepared, and the anchor boxes learned by the lower threshold are passed to the higher threshold such that both high- and low-IoU position prediction can be achieved.

In our proposed method, the number of heads was expanded from three to five, and the

corresponding threshold values of each IoU were set from 0.5, 0.6, and 0.7 to 0.5, 0.6, 0.7, 0.8, and 0.9, as shown in Figure 5.3. This improvement enables the bounding box to be predicted with a higher accuracy. The ROI of the tumor was set based on the coordinates of the bounding box output from the improved cascade R-CNN. Subsequently, the size of the ROI was adjusted by a certain number of pixels. Specifically, the ROI was refined by setting approximately a quarter of its length on its width and height to prevent the tumor area, which may extend beyond the ROI, from being overlooked.

The detector was trained end-to-end and sequentially, leveraging the fact that the output of a detector is sufficient for training the following level. Each detector possesses a positive set of examples of the same size. The resampling progressively improved the hypotheses to avoid overfitting. The same cascade procedure was conducted at the inference to ensure that the hypotheses matched more closely to the increasing detector quality.
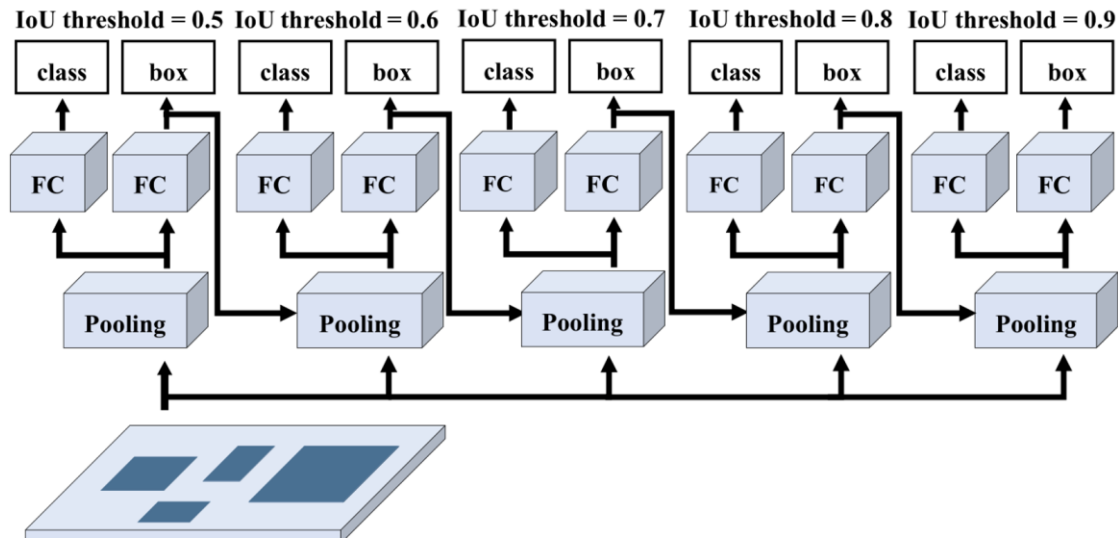
Figure 5.3 Structure of head section in improved cascade R-CNN. Number of heads was expanded from three to five.

5.3 Boundary Refine Using Fast Marching

5.3.1 Fast Marching

The originally proposed fast marching method was derived from the level-set method (LSM) [33]. It considers a curved surface model that is one dimension higher than the dimension of the target image, evolves the curved surface model over time, and eliminates the curved surface model at a point with increased dimensions to extract the target object region. Considering the case where a closed curve γ $(t = 0)$ appears as the outline of an object on the image, the level-set function ϕ (x, y, t) that matches γ at $t = 0$ is expressed as follows:

$$\phi(x, y, t = 0) = \pm d \tag{5.1}$$

where d is the distance from the closed curve γ and t is the time; the initial level-set function ϕ $(x, y, t = 0)$ is equal to the distance transformation of the initial contour. The inside and outside of the closed curve γ is expressed as a positive and negative number, respectively. After obtaining the initial level-set function, the target object region is extracted by solving the following level-set function evolution equation:

$$\frac{\partial \phi}{\partial t} = F \cdot |\nabla \phi| \tag{5.2}$$

where $\nabla \phi$ and F represent the normal vector and velocity function, respectively. The velocity function indicates the propagation velocity of the level-set function. The results can be obtained by solving a partial differential equation.

In addition, the initial contour of the LSM in the proposed method was created based on the ROI. Specifically, the size of the ROI was reduced to a quarter of its size and regarded as the initial contour. This initial contour is used because when the original ROI is regarded as the initial contour, the black region of the background other than the liver cannot be managed because it the segmentation boundaries are likely to diverge.

Therefore, segmentation is performed using the reduced initial contour, which is not affected by the black background.

5.3.2 Edge Preserving Optimization

In the original LSM, when the curve evolves, the difference in value with the neighboring pixels is not considered. In this study, we adopted the strategy of bilateral filtering [37]: for a pixel to affect another pixel, the smoothing weight is determined by the difference in brightness, not the difference in distance between pixels. When the difference is significant, the contour is retained by weight reduction. The filter is shown in formula 5.3, where $g(i,j)$ and $f(i,j)$ represent the output and input images, respectively; i and j are the corresponding pixel indexes; w, σ_1, and σ_2 represent the kernel size, smoothing parameter of the displacement, and intensity, respectively. $h_d(m,n)$ is a Gaussian function that describe the spatial kernel for smoothing differences in displacement and $h_v(i,j,m,n)$ is the range kernel for smoothing differences in intensities.

$$g(i,j) = \frac{\sum_n^w \sum_m^w f(i+m,j+n) h_d(m,n)\, h_v(i,j,m,n)}{\sum_n^w \sum_m^w h_d(m,n)\, h_v(i,j,m,n)} \tag{5.3}$$

$$h_d(m,n) = \exp\left(-\frac{m^2+n^2}{2\sigma_1^2}\right) \tag{5.4}$$

$$h_v(i,j,m,n) = \exp\left(-\frac{\big(f(i,j)-f(i+m,j+m)\big)^2}{2\sigma_2^2}\right) \tag{5.5}$$

Furthermore, m and n are the displacements of the neighbor pixels, satisfying $|m| \le w$ and $|n| \le w$.

5.4 Experiments and Results

5.4.1 Dataset and Environment

Information regarding the testing images used in the experiment is shown in Table 5.2. Eleven cases of the ART phase, nine cases of the PV phase, and six cases of delayed phase were included in the training set. At least one tumor area was present in each case. The size of all images was 512×512 pixels. Window transformations were preprocessed on the CT images in the dataset. In the ART phase, we set the window width and windowing level to 350 and 40, respectively. In the PV and delayed phases, we set the window width and window level to 184 and 108, respectively.

Table 5.2 Information regarding CT images

Phase type	Number of cases	Number of slices
ART phase	11	1198
PV phase	9	951
Delayed phase	6	483

In the dynamic CT images, multiple phases of images were captured in multiple periods after the radiocontrast agent was injected. In our method, the liver region was manually segmented from the ART phase of the CT images. This operation eliminates the interference of organs, bones, or other textures in the slices, which may result in misdetection or mis-segmentation.

5.4.2 Evaluation Methods

Because our experiment involved two steps, i.e., tumor detection and tumor segmentation, we introduced two sets of evaluation methods to evaluate their performance. Tumor detection by the improved cascade R-CNN was evaluated via n-fold cross-validation. The training set used in the experiment was randomly segregated into n

sets, and for each set, $(n-1)$ sets were used for learning, whereas the remaining set was used for the test. Table 5.3 shows the number of slices associated with the existence of tumors in the ART phase images used in the experiment. In addition, the data were inflated by rotating the image to the left and right by 30° in each dataset.

Table 5.3 Details of experimental dataset for ART phase

Dataset	Number of slices containing tumors	Number of slices without tumors
1	96	203
2	68	230
3	90	211
4	71	229

In this study, the average precision (AP) was obtained via a four-fold cross-validation for each test dataset, and it was calculated to estimate the segmentation accuracy of the rectangular region containing tumors. The AP was obtained from the precision–recall curve, which indicates both the recall rate and precision rate simultaneously. A confusion matrix was used to summarize the classification of the proposed small bounding boxes by the network. The prediction results can be classified into four categories based on the true class and the class predicted by the model. For cases involving tumors, they are defined as true positive (TP) when they are correctly classified, and false negative (FN) when regarded as non-tumors by the model. Similarly, for the non-tumor cases, they are defined as true negative (TN) when classified as non-tumors, and false positive (FP) when classified as tumors.

Recall represents the percentage of correctly extracted tumors from true tumors. Precision represents the ratio of the correctly extracted tumors in all cases of the detected tumor. They can be calculated using formulas 5.6 and 5.7, as follows:

$$\text{recall} = \frac{TP}{TP + FN} \tag{5.6}$$

106

$$\text{precision} = \frac{TP}{TP + FP} \tag{5.7}$$

In this study, cascade R-CNNs were trained by combining the PV and delayed phase datasets into one owing to the small number of trained images. In addition, the numerical evaluation was limited to the ART phase dataset.

To evaluate the performance of tumor segmentation from the detected ROIs, the IoU, true positive rate (TPR), and false positive rate (FPR) were introduced. The IoU reflects the proportion of the correctly extracted area to the union of the true and extracted areas. The TPR reflects the proportion of the correctly extracted area to the true tumor area, and the FPR reflects the oversegmented area to the segmented area. In contrast to the tumor detection task in the first step, which is evaluated using class-level indicators, the result of tumor segmentation is evaluated at the pixel level. The formula for each evaluation is shown below (A is the area of the ground truth, and B is the area of segmentation):

$$\text{IoU} = \frac{A \cap B}{A \cup B} \tag{5.8}$$

$$\text{TPR} = \frac{A \cap B}{A} \tag{5.9}$$

$$\text{FPR} = \frac{B - A \cap B}{B} \tag{5.10}$$

5.4.3 Detection of Liver Tumors

The preprocessed CT images described in Section 5.4.1 were used as inputs for the faster R-CNN, cascade R-CNN, proposed network, and improved cascade R-CNN for 11 cases of the ART phase images. In the experiments, the models were tested using the configuration of Nvidia GeForce GTX 1070 with 16 GB of memory. The experimental results, including the AP and average recall of different IoU thresholds of each dataset, are shown in Tables 5.4 and 5.5. In Table 5.4, $\text{AP}_{T=0.5}$ represents the AP values obtained

by recognizing the region as TP when the IoU value of the output is greater than or equal to 0.5; the same applies for recall in Table 5.5.

Table 5.4 AP (%) of tumor detection results

Methods	$AP_{T=0.5}$	$AP_{T=0.6}$	$AP_{T=0.7}$	$AP_{T=0.8}$	$AP_{T=0.9}$
Faster R-CNN	73.2	65.3	51.1	24.1	0.973
Cascade R-CNN	74.9	67.6	56.3	27.8	0.992
Proposed method	76.8	72.0	61.3	36.8	2.11

Table 5.5 Recall (%) of tumor detection results

Methods	$Recall_{T=0.5}$	$Recall_{T=0.6}$	$Recall_{T=0.7}$	$Recall_{T=0.8}$	$Recall_{T=0.9}$
Faster R-CNN	82.5	76.6	62.1	36.5	3.97
Cascade R-CNN	83.6	77.8	68.3	41.1	5.71
Proposed method	84.4	79.5	68.6	44.0	7.13

5.4.4 Segmentation of Liver Tumors

Table 5.6 shows the IoU, TPR, and FPR of the tumor region segmentation results for three different methods, including the conventional method using the U-Net by Ronneberger et al. [61] and the proposed method with and without bilateral filtering. The inputs of the proposed methods were obtained from the output ROIs using the improved cascade R-CNN presented in Section 5.2.2. The kernel size w, distance variance σ_1, and variance value of the pixel σ_2, which are the parameters of the bilateral filter, were set to 9, 10, and 10, respectively. Examples of output from images containing contacted tumors are obtained using the conventional method [61] and the proposed method and presented in Figures 5.4 (a), (b) and (c), respectively. As shown in Table 5.6, the IoU of the proposed method was 72.7%, which is an improvement of 10.2% compared with the conventional method.

Table 5.6 Final Segmentation Results of Tumor Region for Different Methods

Methods	IoU	TPR	FPR
Conventional method [61]	62.5	80.3	20.2
Proposed method (without bilateral filtering)	66.9	70.8	2.35
Proposed method (with bilateral filtering)	72.7	76.2	4.75

(a)　　　　　　　　　(b)　　　　　　　　　(c)
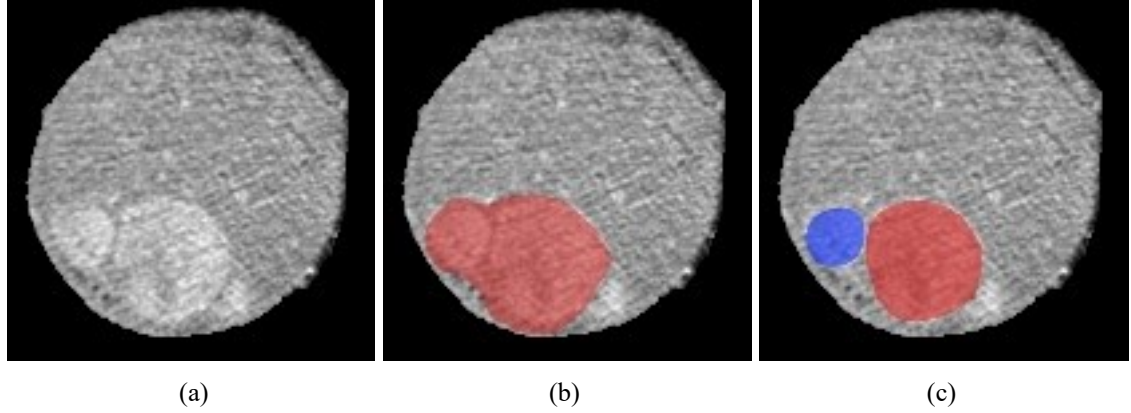
Figure 5.4 Comparison of segmentation results. In the original image (a), the region of tumors is closely adjacent. Final segmentation results of liver tumors by using (b) U-Net [61] and (c) proposed method are shown.

5.4.5 Discussion

Herein, we proposed a hybrid method to efficiently extract liver tumor regions from dynamic CT images. Specifically, an improved cascade R-CNN was used to obtain the initial ROIs of tumors in the first step, and a level-set framework involving phase-sensitive noise filtering was performed to obtain precise segmentation results of the liver tumor regions from the ROIs.

The results of tumor ROIs detected by the faster R-CNN, cascade R-CNN, and the proposed improved cascade R-CNN are shown in Tables 5.5 and 5.6. As shown, the improved cascade R-CNN yielded higher accuracies in terms of both the AP and recall compared with the other methods. In particular, when the threshold IoU was 0.8 and 0.9, both the AP and recall improved significantly owing to the addition of the head section

described in section 5.2.2. In addition, a highly accurate bounding box was obtained for a large tumor in the center. This is assumed to be associated with the number of anchor boxes. Many anchor boxes were applied to large objects in the RPN. Therefore, many boxes satisfying the IoU thresholds of 0.8 and 0.9 were retained, thereby effectively realizing learning in the added head section; consequently, a highly accurate bounding box for large objects was output.

For some of the cases, the tumor disappeared in the PV phase, despite its presence in the ART phase. This is attributable to differences in the concentration patterns for each lesion (cancer), such as the deep dyeing time and deep dyeing intensity, due to differences in the controlling blood vessels and blood flow. Consequently, only some of the tumors were stained in the PV phase. Therefore, a new image analysis method that considers these factors must be devised. Specifically, the class of tumors can be extended into multiple classes. In this study, the class of tumors is limited to only one; however, it can be further categorized into multiple classes, such as primary cancer, metastatic cancer, and intrahepatic bile duct cancer, to improve the problem of cancer disappearance. In particular, compared with primary metastatic cancer, metastatic carcinoma, intrahepatic cholangiocarcinoma, etc. are considered to be hypovascular and may not be stained sufficiently in the PV phase; hence, it would be meaningful to extract the ROIs of multiple classes of cancers [152].

The final segmentation of the tumor regions from the bounding box obtained from the improved cascade R-CNN is shown in Table 5.3. The highest TPR was 80.3% using the conventional method [61], and the highest FPR was 20.2%. In the proposed method without bilateral filtering, the FPR was the lowest at 2.35%, and the TPR reduced accordingly. By contrast, in the proposed method with bilateral filtering, the FPR was slightly higher than that of the method without bilateral filtering; however, the TPR improved accordingly. In general, the IoU of the proposed method was the highest (72.7%) compared with those of the other methods. The effectiveness of the bilateral filter was evident when used in conjunction with our proposed method, particularly for

smaller tumors.

When the distance between multiple tumor regions was small, as presented in Figures 5.4 (b) and 5 (c), multiple tumors were detected as one tumor, whereas they were detected separately when using the proposed method. This indicates that the proposed method can separate and extract tumors that were almost in contact with each other. Hence, this solves the problem of mis-segmentation of extremely close tumors present in the conventional method and validates the efficiency of our proposed tumor segmentation method, which benefits from two procedures, i.e., initial ROI detection and final segmentation of liver tumor regions.

Although the proposed method has conducted a better performance of tumor segmentation in the experiment, the detection and segmentation of liver tumors were performed based on manually delineated liver regions. In practical CAD systems, fully automatic process and analysis of medical images can reduce the burden of physicians. The proposed method is not limited to liver tumors, and it can also be applied to the other organs in abdominal or thoracic images after the multi-organ segmentation process.

5.5 Conclusion

Herein, we proposed a hybrid statistical method for the efficient segmentation of the liver tumor region using dynamic CT images. Specifically, an improved cascade region-based convolutional network was employed to automatically extract the region of interest of tumors, and the tumor regions were segmented using a level-set framework combined with bilateral filtering. The proposed method efficiently improved the segmentation accuracy of the surrounding liver tumors in multiphase CT images. The additional cascade structure of the head sections in the proposed network was particularly beneficial when the tumors were considerably near to each other. The accurately segmented region of tumors facilitated the subsequent diagnosis of liver cancer. However, in the proposed method, the liver region used for tumor detection is segmented manually, which incurs an

additional human–computer interaction cost.

In future studies, the quantity of training data can be increased, particularly in cases where the tumor regions are close to each other, to increase the accuracy for similar cases. In addition, the final segmentation result of tumors can be integrated across different phases, which may reduce the effects of different appearances of tumors from different phases. In addition, accurate automatic segmentation of liver regions can be considered while ensuring the integrity of the liver regions and excluding interference from non-liver regions.

Chapter 6

Summary and Outlook

In this thesis, we proposed five statistic-based approaches that are designed to solve some of the current problems in the field of human organ or small tissues segmentation from medical images, which forms an improved scheme of computer-aided diagnosis systems for cancer diagnosis. The improvement of the proposed method from the experiment results indicates their capability of these methods in the application of segmentation problems for the CAD systems.

6.1 Conclusions and Remarks

(1) Landmarks correspondence for single organ SSMs by *k*-means clustering

In the first method proposed in section 3.3, we proposed an automatic landmark correspondence method for the generation of 3-D SSMs. In a previously proposed SSMs building method [99-100], the reference surface for landmarks corresponding is generated by a quadric-based simplification. To select more representative landmarks and obtain a relatively evenly distributed set of landmarks, we introduce a *k*-mean method to cluster the vertices into groups. The vertices in the same cluster share similar spatial features and their centroid is regarded as the selected landmark.

The quality of SSMs built by using the proposed landmarks correspondence method is higher than that of the SSMs built with a previous method in terms of generalization ability and specificity while maintaining the same compactness. The improvement of the model quality indicates that the prototype of landmarks generated by the proposed method can increase the ability to learn deformation patterns and produce new shapes.

(2) Multi-organ SSM building by non-rigid ICP registration

In the second method proposed in section 3.4, a series of surface registration strategies combining NICP registration based on G-RBF are applied to the landmarks corresponding problem of multi-organ SSMs. The Demons registration used in section 3.3 relies on a single spherical surface, which is invalid in jointed shape containing surface meshes of multiple organs. The registration between the surfaces in the training set is conducted by the NICP method instead. We developed five registration strategies of registration involving the different combinations of single or multiple organs in the detailed matching process.

These strategies are tested in the experiment of building multi-organ SSMs from a training set containing 30 cases of surfaces which are consist of four abdominal organs. The result of registration and SSMs building shows that the proposed NICP registration strategy can improve the accuracy of registration and increase the quality of SSMs built by the corresponding landmarks from the registration. The single organ matching process is effective for the misregistration problems between different organs.

(3) Localization and coarse segmentation of multi-organ by SSMs

In the third method proposed in section 4.1, an automatic multi-organ segmentation method from CT image is proposed. In the segmentation tasks by using SSMs, the position of the organs in the medical images is required so that the mean shape of the SSM can be started at a relatively correct position, due to the search process of model deformation depends on the intensity appearance of its surroundings voxels. If the initial position is far from the actual position of the organs, the deformation of the model may be led to a wrong direction and result in bad segmentation. To solve this problem, a RF regression model is trained to find the candidate position and initialize the SMMs before the searching and deformation process in the segmentation task. The segmentation accuracy is improved in the experiment, which indicates the capability of the RF based multi-organ segmentation approach proposed.

(4) Fine segmentation of multiple organs by supervoxel classification

In section 4.2, a novel method for automatic abdominal multi-organ segmentation by introducing spatial information in the process of supervoxels classification. The supervoxels with boundaries adjacent to anatomical edges are separated from the image by using the simple linear iterative clustering (SLIC) from the images. The images are clustered into small units of supervoxels with similar intensities and positions. The RF classifier is built to predict the labels of the supervoxels according to their spatial and intensity features. Compared with the previous method, the process of model building, initialization, and searching is not required. The segmentation result is directly obtained from the classification of the supervoxels. Thirty abdominal CT images are used in the experiment of segmentation task for the spleen, right kidney, left kidney, and liver region. The experiment result shows that the proposed method achieves a higher accuracy of segmentation compared to our previous model-based method.

(5) Liver Tumor Detection and Segmentation by a Hybrid Method

In chapter 5, we proposed a hybrid method to efficiently extract liver tumor regions from dynamic CT images. In the task of liver tumor segmentation, the small tumors are difficult to distinguish especially when they appear in adjacent regions in the liver. This problem is more difficult when the segmentation is carried in the multiphase CT image, as the appearance of the same tumor may be different due to their physical features. To solve this problem, an improved cascade R-CNN was used to obtain the initial ROIs of tumors in the first step, and a level-set framework involving phase-sensitive noise filtering was performed to obtain precise segmentation results of the liver tumor regions from the ROIs. The adjacent tumors are successfully distinguished by the cascade R-CNN due to the additional head sections in the proposed network. The accurately segmented region of tumors facilitated the subsequent diagnosis of liver cancer.

6.2 Future Works

In the following section, we introduce some of the ideas and planning the future work.

- The robustness is one of the advantages of SSMs methods, and it is an important factor in the application of medicine. The main architecture can be retained, and the detailed process can be refined by more effective feature-guided methodologies of deep learning techniques. The landmark correspondence can be improved by introducing more effective registration approaches. The statistical analysis can also be extended to non-linear space to extract a more representative pattern of shape deformation. The active appearance model of higher dimensions can be extended to the inner voxel data of the organs instead of their surfaces.

- The proposed methods are mainly tested in the thorax and upper abdominal CT images. Many segmentation tasks focus on the other human structures or modalities of images that may share similar applicability. The proposed framework of CAD system can be extended to the application of other tissues, especially the images containing multiple adjacent objects.

- Adequate training data is important for the SSMs training. The quality of the model can be improved by adding more training data, which can improve the deformation ability of the model.

- In future studies of tumor segmentation, the quantity of training data can be increased, particularly in cases where the tumor regions are close to each other, to increase the accuracy for similar cases.

- In addition, the final segmentation result of tumors can be integrated across different phases, which may reduce the effects of different appearances of tumors from different phases.

- In the final output of our CAD system, the contours of tumors are delineated.

However, it is important to classify the tumors as malignant or benign for the diagnosis of cancers. A further classification of the tumor will provide a more useful reference for the physicians.

Bibliography

[1] J. H. Scatliff, S. Johnston, "Andreas Vesalius and Thomas Willis: their anatomic brain illustrations and illustrators," *American Journal of Neuroradiology*, vol. 35, no. 1, pp. 19-22, 2014.

[2] B. K. Menon, C. D. d'Esterre, et al., "Multiphase CT Angiography: A New Tool for the Imaging Triage of Patients with Acute Ischemic Stroke," *Radiology*, vol. 75, no. 2, pp. 510-520, 2015.

[3] S. Roy, Y. Chi, et al., "Three-Dimensional Spatiotemporal Features for Fast Content-Based Retrieval of Focal Liver Lesions," *IEEE Transactions on Biomedical Engineering*, vol. 61, pp. 2768-2778, 2014,

[4] H. Ji, J. D. McTavish, et al., "Hepatic Imaging with Multidetector CT," *Radiographics*, vol. 21, no. 1, pp. S71-S80, 2001.

[5] I. R. Kamel, M. A. Choti, el al., "Surgically Staged Focal Liver Lesions: Accuracy and Reproducibility of Dual-phase Helical CT for Detection and Characterization," *Radiology*, vol. 227, no. 3, pp. 752- 757, 2003.

[6] H. J. Kim, A. Y. Kim, et al., "Transient Hepatic Attenuation Differences in Focal Hepatic Lesions: Dynamic CT Features," *American Journal of Roentgenology*, vol. 184, no. 1, pp. 83-90, 2005.

[7] P. J. Navin, S. K. Venkatesh, "Hepatocellular carcinoma: state of the art imaging and recent advances," *Journal of clinical and translational hepatology*, vol. 7, no. 1, pp. 72, 2019.

[8] J. Yanase, E. Triantaphyllou, "A systematic survey of computer-aided diagnosis in medicine: Past and present developments," *Expert Systems with Applications*, vol. 138, pp. 112821, 2019.

[9] H. Fujita, Y. Uchiyama, et al., "Computer-aided diagnosis: The emerging of three CAD systems induced by Japanese health care needs," *Computer methods and programs in biomedicine*, vol. 92, no. 3, pp. 238-248, 2008.

[10] M. Firmino, G. Angelo, et al., "Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy," *Biomedical engineering online*, vol. 15, no. 1, pp. 1-17, 2016.

[11] P. H. Myers, C. M. Nice, H. C. Becker, "Automated computer analysis of radiographic images," *Radiology*, vol. 83, no. 6, pp. 1029-1034, 1964.

[12] G. S. Lodwick, C. L. Haun, et al., "Computer diagnosis of primary bone tumors: a preliminary report," *Radiology*, vol. 80, no. 2, pp. 273-275, 1963.

[13] H. C. Becker, W. J. Nettleton, et al., "Digital computer determination of a medical diagnostic index directly from chest X-ray images," *IEEE Transactions on Biomedical Engineering*, no. 3, pp. 67-72, 1964.

[14] C. Balleyguier, K. Kinkel, et al., "Computer-aided detection (CAD) in mammography: does it help the junior or the senior radiologist?" *European journal of radiology*, vol. 54, no. 1, pp. 90-96, 2005.

[15] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems*, Man, and Cybernetics, vol. 9, no. 1, pp. 62-66, 1979.

[16] H. Ng, "Automatic thresholding for defect detection," *Pattern Recognition Letters*, vol. 27, no. 14, pp. 1644-1649, 2006.

[17] F. Argenti, L. Alparone, et al., "Fast algorithms for texture analysis using co-occurrence matrices," *IEE Proceedings F Radar and Signal Processing*, vol. 137, no. 6, pp. 443-448, 1990.

[18] B. B. Chaudhuri, N. Sarkar, "Texture segmentation using fractal dimension," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 1, pp, 72-77, 1995.

[19] R. Adams, L. Bischof, "Seeded region growing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 641-647, 1994.

[20] C. Wang, C. Zhao, et al., "Multiple cues region growing segmentation on tongue image," *International Journal of Computers in Healthcare*, vol. 2, no. 2, pp. 84-97, 2015.

[21] L. Vincent, P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.13, no. 6, pp. 583-598, 1991.

[22] H. P. Ng, S. H. Ong, et al., "Medical image segmentation using k-means clustering and improved watershed algorithm. Image Analysis and Interpretation," *IEEE southwest symposium on image analysis and interpretation*, pp. 61-65, 2006.

[23] N. Senthilkumaran, R. Rajesh, "Edge detection techniques for image segmentation–a survey of soft computing approaches," *International journal of recent trends in engineering*, vol. 1, no. 2, pp. 250-254, 2009.

[24] P. Gibbs, D. L. Buckley, et al., "Tumour volume determination from MR images by morphological segmentation," *Physics in medicine and biology*, vol. 41, no. 11, pp. 2437, 1996.

[25] A. Banumathi, R. K. Devi, V. A. Kumar V A, "Performance analysis of matched filter techniques for automated detection of blood vessels in retinal images," *TENCON 2003. Conference on Convergent Technologies for the Asia-Pacific Region*, vol. 2, pp. 543-546, 2003.

[26] M. Reuter, S. Biasotti, et al., "Discrete Laplace–Beltrami operators for shape analysis and segmentation," *Computers & Graphics*, vol. 33, no. 3, pp. 381-390, 2009.

[27] H. B. Kekre, S. M. Gharge, "Image segmentation using extended edge operator for mammographic images," *International journal on computer science and Engineering*, vol. 2, no. 4, pp. 1086-1091, 2010.

[28] A. J. Vyavahare, "Canny based DRLSE Algorithm for Segmentation," *International Journal of Computer Applications*, vol. 102, no. 7, pp. 1-5, 2014.

[29] A. F. Frangi, W. J. Niessen, et al., "Multiscale vessel enhancement filtering," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 130-137, 1998.

[30] J. N. Kaftan, A. P. Kiraly, et al., "Fuzzy pulmonary vessel segmentation using optimized vessel enhancement filtering," *First International Workshop on Pulmonary Image Analysis*, pp. 233-242, 2008.

[31] D. L. Wilson, J. A, Noble, "An adaptive segmentation algorithm for time-of-flight MRA data," *IEEE Transactions on Medical Imaging*, vol. 18, no. 10, pp. 938-945, 1999.

[32] M. S. Hassouna, A. A. Farag, et al., "Cerebrovascular segmentation from TOF using stochastic models," *Medical Image Analysis*, vol. 10, no. 1, pp. 2-18, 2006.

[33] K. Held, E. R. Kops, et al., "Markov random field segmentation of brain MR images," *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp. 878-886, 1997.

[34] M. Cabezas, A. Oliver, et al., "A review of atlas-based segmentation for magnetic resonance brain images," *Computer methods and programs in biomedicine*, vol. 104, no. 3, pp. e158-e177, 2011.

[35] M. Lorenzo-Valdés, G. I. Sanchez-Ortiz, et al., "Atlas-based segmentation and tracking of 3D cardiac MR images using non-rigid registration," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 642-650, 2002.

[36] I. Isgum, M. Staring, et al., "Multi-atlas-based segmentation with local decision fusion—Application to cardiac and aortic segmentation in CT scans," *IEEE transactions on medical imaging*, vol. 28, no. 7, pp. 1000-1010, 2009.

[37] M. Kass, A. Witkin, D. Terzopoulos, "Snakes: Active contour models," *International journal of computer vision*, vol. 1, no. 4, pp. 321-331, 1988.

[38] C. Xu, J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Transactions on image processing*, vol. 7, no. 3, pp. 359-369, 1998.

[39] L. D. Cohen, "On active contour models and balloons," *CVGIP: Image understanding*, vol. 53, no. 2, pp. 211-218, 1991.

[40] S. Osher, J. A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations," *Journal of Computational Physics*, vol. 79, no. 1, pp. 12-49, 1988.

[41] A. Tsai, A. Yezzi, et al., "A shape-based approach to the segmentation of medical imagery using level sets," *IEEE transactions on medical imaging*, vol. 22, no. 2, pp. 137-154, 2003.

[42] D. Cremers, M. Rousson , R. Deriche, "A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape," *International journal of computer vision*, vol. 72, no. 2, pp. 195-215, 2007.

[43] B. N. Li, C. K. Chui, et al., "Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation," *Computers in biology and medicine*, vol. 41, no. 1, pp. 1-10, 2011.

[44] T. F. Cootes, C. J. Taylor, et al., "Training models of shape from sets of examples," *Proceedings of the British Machine Vision Conference*, pp. 9-18, 1992.

[45] T. F. Cootes, C. J. Taylor, et al., "Active shape models—their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38-59, 1995.

[46] T. F. Cootes, G. J. Edwards, et al., "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681-685, 2001.

[47] R. H. Davies, C. J. Twining, et al., "3D Statistical Shape Models Using Direct Optimisation of Description Length," *European conference on computer vision*, pp. 3-20, 2002.

[48] B. Van Ginneken, A. F. Frangi, et al., "Active shape model segmentation with optimal features," *IEEE transactions on medical imaging*, vol. 21, no. 8, pp. 924-933, 2002.

[49] B. Leibe, A. Leonardis, B. Schiele, "Combined object categorization and segmentation with an implicit shape model," *Workshop on statistical learning in computer vision*, vol.2, no. 5, pp. 7, 2004.

[50] T. Heimann, I. Wolf, et al., "3D active shape models using gradient descent optimization of description length," *Biennial International Conference on Information Processing in Medical Imaging*, pp. 566-577, 2005.

[51] T. Heimann, I, Wolf, H. P. Meinzer, "Active shape models for a fully automated 3D segmentation of the liver–an evaluation on clinical data," *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pp. 41-48, 2006.

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

[53] G. Li, "Study on the Method of Constructing a Statistical Shape Model and Its Application to the Segmentation of Internal Organs in Medical Images," Kitakyushu: Kyushu Institute of Technology, 2013.

[54] C. Li, J. Zhang, Q. Feng, "Liver CT image segmentation using statistical shape model based on statistical and specific information," *Journal of Southern Medical University*, vol. 32, no. 1, pp. 23-27, 2012.

[55] T. Shen, H. Li H, et al., "Active Volume Models for Medical Image Segmentation," *IEEE Transactions on Medical Imaging*, vol. 30, no. 3, pp. 774-791, 2011.

[56] L. Wang, M. Kohnen, et al., "Fast automated segmentation of femoral heads in fluoroscopic X-ray images," *International symposium on biomedical imaging*, pp. 984-988, 2011.

[57] W. E. Reddick, J. O. Glass, et al., "Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks," *IEEE Transactions on medical imaging*, vol. 16, no. 6, pp. 911-918, 1997.

[58] D. E. Rumelhart, G. E. Hinton, et al., "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533-536, 1988.

[59] G. E. Hinton, R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504-507, 2006.

[60] H. C. Shin, H. R. Roth, et al., "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.

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

[62] K. Kamnitsas, C. Ledig, et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical image analysis*, vol. 36, pp. 61-78, 2017.

[63] M. Ozkan, B. M. Dawant, R. J. Maciunas, "Neural-network-based segmentation of multi-modal medical images: a comparative and prospective study," *IEEE transactions on Medical Imaging*, vol. 12, no. 3, pp. 534-544, 1993.

[64] T. Mcinerney, D. Terzopoulos, "Deformable models in medical image analysis: a survey," *Medical Image Analysis*, vol. no. 2, pp. 91-108, 1996.

[65] G. Szekely, A. Kelemen, et al., "Segmentation of 2D and 3D objects from MRI volume data using constrained elastic deformations of flexible Fourier contour and surface models," *Medical Image Analysis*, vol. 1, no. 1, pp. 19-34, 1996.

[66] H. Bium, "A transformation for extracting Fnew descriptions of shape," Symposium on Modeis for the Perception of Speech and Visua1 Form, 1964.

[67] P. J. Besl, H. D. Mckay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239-256, 1992.

[68] T. Okada, M. G. Linguraru, et al., "Abdominal multi-organ segmentation from CT images using conditional shape–location and unsupervised intensity priors," *Medical image analysis*, vol, 26, no. 1, pp. 1-18, 2015.

[69] V. Zografos, A. Valentinitsch, et al., "Hierarchical multi-organ segmentation without registration in 3D abdominal CT images," *International MICCAI Workshop on Medical Computer Vision*, pp. 37-46, 2015.

[70] J. Yao, R. M. Summers, "Statistical location model for abdominal organ localization," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 9-17, 2009.

[71] X. Liu, M. G. Linguraru, et al, "Organ pose distribution model and an MAP framework for automated abdominal multi-organ localization," *International Workshop on Medical Imaging and Virtual Reality*, pp. 393-402, 2010.

[72] A. Criminisi, J. Shotton J, et al., "Regression forests for efficient anatomy detection and localization in CT studies," *International MICCAI Workshop on Medical Computer Vision*, pp. 106-117, 2010.

[73] O. Pauly, B. Glocker, et al., "Fast multiple organ detection and localization in whole-body MR Dixon sequences," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 239-247, 2011.

[74] R. Gauriau, R. Cuingnet, et al., "Multi-organ localization with cascaded global-to-local regression and shape prior," *Medical image analysis*, vol. 23, no. 1, pp. 70-83, 2015.

[75] P. N. Samarakoon, E. Promayon, and Fouard C, "Light random regression forests for automatic multi-organ localization in CT images," *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 371-374, 2017.

[76] M. G. Linguraru, J. A. Pura, et al., "Statistical 4D graphs for multi-organ abdominal segmentation from multiphase CT," *Medical image analysis*, vol. 16, no. 4, pp. 904-914, 2012.

[77] P. Aljabar, R. A. Heckemann, et al., "Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy," *Neuroimage*, vol. 46, no. 3, pp. 726-738, 2009.

[78] A. Montillo, J. Shotton, et al., "Entangled decision forests and their application for semantic segmentation of CT images," *Biennial International Conference on Information Processing in Medical Imaging*, pp. 184-196, 2011.

[79] M. G. Uzunbas, O. Soldea, et al., "Coupled nonparametric shape and moment-based intershape pose priors for multiple basal ganglia structure segmentation," *IEEE transactions on medical imaging*, vol. 29, no. 12, pp. 1959-1978, 2010.

[80] A. Shimizu, R. Ohno R, et al., "Segmentation of multiple organs in non-contrast 3D abdominal CT images," *International journal of computer assisted radiology and surgery*, vol. 2, no. 3, pp. 135-142, 2007.

[81] R. Wolz, C. Chu, et al., "Automated abdominal multi-organ segmentation with subject-specific atlas generation," *IEEE transactions on medical imaging*, vol. 32, no. 9, pp. 1723-1730, 2013.

[82] C. Chu, N. Oda, et al., "Multi-organ segmentation based on spatially-divided probabilistic atlas from 3D abdominal CT images," *International conference on medical image computing and computer-assisted intervention*, pp. 165-172, 2013.

[83] N. Lay, N. Birkbeck, et al., "Rapid multi-organ segmentation using context integration and discriminative models," *International Conference on Information Processing in Medical Imaging*, pp. 450-462, 2013.

[84] T. Takaoka, Y. Mochizuki, H. Ishikawa, "Multiple-organ segmentation by graph cuts with supervoxel nodes," *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, pp. 424-427, 2017.

[85] X. Zhou, R. Takayama, et al., "Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method," *Medical physics*, vol. 44, no. 10, pp. 5221-5233, 2017.

[86] H. R. Roth, H. Oda, et al., "An application of cascaded 3D fully convolutional networks for medical image segmentation," *Computerized Medical Imaging and Graphics*, vol. 66, pp. 90-99, 2018.

[87] E. Gibson, F. Giganti, et al., "Automatic multi-organ segmentation on abdominal CT with dense V-networks," *IEEE transactions on medical imaging*, vol. 37, no. 8, pp. 1822-1834, 2018.

[88] Y. Wang, Y. Zhou, et al., "Abdominal multi-organ segmentation with organ-attention networks and statistical fusion," *Medical image analysis*, vol. 55, pp. 88-102, 2019.

[89] Y. Wang, Y. Zhou, et al., "Training multi-organ segmentation networks with sample selection by relaxed upper confident bound," *International conference on medical image computing and computer-assisted intervention*, pp. 434-442, 2018.

[90] M. F. Bobo, S. Bao, et al., "Fully convolutional neural networks improve abdominal organ segmentation," *Medical Imaging 2018: Image Processing. International Society for Optics and Photonics*, vol. 10574, pp. 105742V, 2018.

[91] Y. Zhou, Z. Li, et al., "Prior-aware neural network for partially-supervised multi-organ segmentation," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10672-10681, 2019.

[92] P. H. Conze, A. E. Kavur, et al., "Abdominal multi-organ segmentation with cascaded convolutional and adversarial deep networks," *Artificial Intelligence in Medicine*, vol. 117, pp. 102109, 2021.

122

[93] X. Fang, P. Yan, "Multi-organ segmentation over partially labeled datasets with multi-scale feature abstraction," *IEEE Transactions on Medical Imaging*, vol. 39, no. 11, pp. 3619-3629, 2020.

[94] W. E. Lorensen, H. E. Cline. "Marching cubes: A high resolution 3D surface construction algorithm," *ACM siggraph computer graphics*, vol. 21, no. 4, pp. 163-169, 1987.

[95] J. C. Gower, "Generalized procrustes analysis," *Psychometrika*, vol. 40, no. 1, pp. 33-51, 1975.

[96] I. T. Jolliffe, "Choosing a subset of principal components or variables," *Principal component analysis*, pp. 111-149, 2002.

[97] R. H. Davies, "Learning shape: optimal models for analysing natural variability," The University of Manchester (United Kingdom), 2002.

[98] M. A. Styner, K. T. Rajamani, et al., "Evaluation of 3D correspondence methods for model building." *In Biennial International Conference on Information Processing in Medical Imaging*, pp. 63-75, 2003.

[99] J. Wu, G. Li, et al., "Statistical Shape Model Generation Using Diffeomorphic Surface Registration," *In Proceedings of the 2nd International Conference on Biomedical Signal and Image Processing*, pp. 37-41, 2017.

[100] G. Li, J. Wu, et al., "Statistical shape model building method using surface registration and model prototype," *Optics & Laser Technology*, vol. 110, pp. 234-238, 2019.

[101] J. A, Hartigan, M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100-108.

[102] X. Gu, Y. Wang, et al., "Genus zero surface conformal mapping and its application to brain surface mapping," *IEEE transactions on medical imaging*, vol. 23, no. 8, pp. 949-958, 2004.

[103] B. T. T. Yeo, M. Sabuncu, et al., "Spherical demons: Fast surface registration," *In International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 745-753, 2008.

[104] S. G. Armato III, G. McLennan, et al., "The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans", *Medical physics*, vol. 38, no. 2, pp. 915-931, 2011.

[105] P. Vanden Berghe, J. Demol, "Virtual anatomical reconstruction of large acetabular bone defects using a statistical shape model," *Computer methods in biomechanics and biomedical engineering*, vol. 20, pp. 577-586, 2017.

[106] O. V. Kaick, H. Zhang, et al., "A survey on shape correspondence," *Computer graphics forum*, vol. 30, pp. 1681-1707, 2011.

[107] Y. Sahillioğlu, "Recent advances in shape correspondence," *The Visual Computer*, vol. 36, pp. 1705–1721, 2020.

[108] G. K. L. Tam, Z. Cheng, et al., "Registration of 3D Point Clouds and Meshes: A Survey from Rigid to Nonrigid," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, pp. 1199 - 1217, 2013

[109] R. H. Davies, C. J. Twining, et al., "Building 3-D statistical shape models by direct optimization," *IEEE Transactions on Medical Imaging*, vol. 29, pp. 961-981, 2009.

[110] N. Ravikumar, A. Gooya, et al., "Group-wise similarity registration of point sets using Student's t-mixture model for statistical shape models," *Medical image analysis*, vol. 44, pp. 961-981, 2018.

[111] J. J. Cerrolaza, M. L. Picazo, et al., "Computational anatomy for multi-organ analysis in medical imaging: A review," *Medical image analysis*, vol. 56, pp. 44-67, 2019.

[112] E. A. Audenaert, J. V. Houcke, et al., "Cascaded statistical shape model based segmentation of the full lower limb in CT," *Computer methods in biomechanics and biomedical engineering*, vol. 22, pp. 644-657, 2019.

[113] T. F. Cootes, A. Hill, et al., "Use of active shape models for locating structures in medical images," *Image and vision computing*, vol. 12, no. 6, pp. 355-365, 1994.

[114] P. Samarakoon, "Random Regression Forests for Fully Automatic Multi-Organ Localization in CT Images," Université Grenoble Alpes, 2016.

[115] https://www.synapse.org/#!Synapse:syn3193805/wiki/217789. Accessed 30 November 2021.

[116] https://ww2.mathworks.cn/help/images/ref/imregister.html. Accessed 30 November 2021.

[117] H. R. Roth, C. Shen, et al., "A multi-scale pyramid of 3D fully convolutional networks for abdominal multi-organ segmentation," *In Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 417-425, 2018.

[118] M. Soltaninejad, G. Yang, et al., "Supervised learning based multimodal MRI brain tumour segmentation using texture features from supervoxels," *Computer methods and programs in biomedicine*, vol. 157, pp. 69-84, 2018.

[119] R. Achanta, A. Shaji, et al., "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, pp. 2274-2282, 2012.

[120] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5-32, 2001.

[121] H. Sung, J. Ferlay, et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 71, no. 3, pp. 209-249. 2021.

[122] B. K. Menon, C. D. d'Esterre, et al., "Multiphase CT Angiography: A New Tool for the Imaging Triage of Patients with Acute Ischemic Stroke," *Radiology*, vol. 75, no. 2, pp. 510-520, 2015.

[123] S. Roy, Y. Chi, et al., "Three-Dimensional Spatiotemporal Features for Fast Content-Based Retrieval of Focal Liver Lesions," *IEEE Transactions on Biomedical Engineering*, vol. 61, pp. 2768-2778, 2014,

[124] H. Ji, J. D. McTavish, et al., "Hepatic Imaging with Multidetector CT," *Radiographics*, vol. 21, no. 1, pp. S71-S80, 2001.

[125] I. R. Kamel, M. A. Choti, el al., "Surgically Staged Focal Liver Lesions: Accuracy and Reproducibility of Dual-phase Helical CT for Detection and Characterization," *Radiology*, vol. 227, no. 3, pp. 752- 757, 2003.

[126] H. J. Kim, A. Y. Kim, et al., "Transient Hepatic Attenuation Differences in Focal Hepatic Lesions: Dynamic CT Features," *American Journal of Roentgenology*, vol. 184, no. 1, pp. 83-90, 2005.

[127] P. H. Conze, V. Noblet, "Scale-adaptive supervoxel-based random forests for liver tumor segmentation in dynamic contrast-enhanced CT scans," *International journal of computer assisted radiology and surgery*, vol. 12, no. 2, pp. 223-233, 2017.

[128] C. Sun, S. Guo, et al., "Automatic Segmentation of Liver Tumors from Multiphase Contrast Enhanced CT Images based on FCNs," *Artificial Intelligence in Medicine*, vol. 83, pp. 58-66, 2017.

[129] F. Ouhmich, V. Agnus, "Liver Tissue Segmentation in Multiphase CT Scans using Cascaded Convolutional Neural Network," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, pp. 1275-1284, 2019.

[130] Y. Wu, Q. Zhou, "Hepatic Lesion Segmentation by Combining Plain and Contrast Enhanced CT Images with Modality Weighted U-Net," *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 255-259, 2019.

[131] Y. Xu, M Cai, et al., "PA-ResSeg: A Phase Attention Residual Network for Liver Tumor Segmentation from Multiphase CT Images," *Medical Physics*, 2021.

[132] N. H. Abdel-massieh, M. M. Hadhoud, K. M. Amin K M, "Fully Automatic Liver Tumor Segmentation from Abdominal CT Scans," *The 2010 International Conference on Computer Engineering & Systems*, pp.197-202, 2010.

[133] Z. Al Sadeque, T. I. Khan, et al., "Automated Detection and Classification of Liver Cancer from CT Images using Hog-SVM Model," *2019 5th International Conference on Advances in Electrical Engineering*, pp. 21-26, 2019.

[134] C. C. Chang, H. H. Chen, et al., "Computer-aided diagnosis of liver tumors on computed tomography images," *Computer methods and programs in biomedicine*, vol. 145, pp. 45-51, 2017.

[135] A. M. Anter, A. E. Hassenian, "CT Liver Tumor Segmentation Hybrid Approach using Neutrosophic Sets, Fast Fuzzy C-means and Adaptive Watershed Algorithm," *Artificial Intelligence in Medicine*, vol. 97, pp. 105-117, 2019.

[136] A. Das, P. Das, et al., "Detection of liver cancer using modified fuzzy clustering and decision tree classifier in CT images," *Pattern Recognition and Image Analysis*, vol. 29, 2o. 2, pp. 201-211, 2019.

[137] R. Ranjbarzadeh, S. B. Saadi, "Automated Liver and Tumor Segmentation based on Concave and Convex Points using Fuzzy C-means and Mean Shift Clustering," *Measurement*, vol. 150, 2020.

[138] P. F. Christ, F Ettlinger, et al., "Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks," ArXiv preprint, arXiv:1702.05970, 2017.

[139] X. Li, H. Chen, et al., "H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volume," *IEEE Transactions on Medical Imaging*, vol. 37, pp. 2663-2674, 2018.

[140] K. Yan, M. Bagheri, R. M. Summers, "3D Context Enhanced Region-based Convolutional Neural Network for End-to-End Lesion Detection," *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pp. 511-519, 2018.

[141] H. Seo, C. Huang, el al., "Modified U-Net (mU-Net) with Incorporation of ObjectDependent High Level Features for Improved Liver and Liver-Tumor Segmentation in CT Images," *IEEE Transactions on Medical Imaging*, vol. 39, pp. 1316-1325, 2019.

[142] X. Wang, S. Han, et al., "Volumetric Attention for 3D Medical Image Segmentation and Detection," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 175-184, 2019.

[143] Z. Bai, H. Jiang, et al., "Liver tumor segmentation based on multi-scale candidate generation and fractal residual network," *IEEE Access*, vol. 7, pp. 82122-82133, 2019.

[144] L. Chen, H. Song, et al., "Liver tumor segmentation in CT volumes using an adversarial densely connected network," *BMC bioinformatics*, vol. 20, no. 16, pp. 1-13, 2019.

[145] Y. A. Ayalew, K. A. Fante, M. A. Mohammed, "Modified U-Net for liver cancer segmentation from computed tomography images with a new class balancing method," *BMC Biomedical Engineering*, vol. 2, no. 1, pp. 1-13, 2021.

[146] D. T. Kushnure, S. N. Talbar, "MS-UNet: A multi-scale UNet with feature recalibration approach for automatic liver and tumor segmentation in CT images," *Computerized Medical Imaging and Graphics*, vol. 89, 2021.

[147] J. Li, X. Ou, et al., "Study on strategy of CT image sequence segmentation for liver and tumor based on U-Net and Bi-ConvLSTM," *Expert Systems with Applications*, vol. 180, 2021.

[148] R. Girshick, J. Donahue, et al., "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580-587, 2014.

[149] R. Girshick, "Fast R-CNN," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440-1448, 2015.

[150] S. Ren, K. He, et al., "Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.1, pp. 91-99, 2015.

[151] Z. Cai, N. Vasconcelos, "Cascade R-CNN: Delving Into High Quality Object Detection," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6154-6162, 2018.

[152] P. Silverman, "Multislice CT in Imaging the Liver," *Cancer Imaging*, vol. 3, pp. 149-154, 2003.