JYU DISSERTATIONS 749

Zhonghua Chen

Multi-organ Medical Image Analysis, Modeling, and Segmentation Exploiting Pre-existing Knowledge

UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF INFORMATION
TECHNOLOGY

JYU DISSERTATIONS 749

Zhonghua Chen

Multi-organ Medical Image Analysis, Modeling, and Segmentation Exploiting Pre-existing Knowledge

Esitetään Jyväskylän yliopiston informaatioteknologian tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Agora-rakennuksen auditoriossa 3
helmikuun 16. päivänä 2024 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Information Technology of the University of Jyväskylä,
in building Agora, auditorium 3, on February 16, 2024 at 12 o’clock noon.

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2024

Editors
Marja-Leena Rantalainen
Faculty of Information Technology, University of Jyväskylä
Ville Korkiakangas
Open Science Centre, University of Jyväskylä

ISBN 978-951-39-9936-0 (PDF)
URN:ISBN:978-951-39-9936-0
ISSN 2489-9003

Permanent link to this publication: http://urn.fi/URN:ISBN:978-951-39-9936-0

ABSTRACT

Chen, Zhonghua
Multi-organ Medical Image Analysis, Modeling, and Segmentation Exploiting Pre-existing Knowledge
Jyväskylä: University of Jyväskylä, 2024, 96 p.
(JYU Dissertations
ISSN 2489-9003; 749)
ISBN 978-951-39-9936-0 (PDF)

In the field of medical image analysis (MIA), accurate segmentation of targets has encountered numerous challenges. Segmentation methods targeting a single specific object have emerged over the past few decades. But the implementation of segmentation methods for multiple targets is technically challenging. MIA is a laborious process with many stages. Current development platforms do not allow for multiple stages efficiently, which slows down the development process. The accurate segmentation of multiple targets simultaneously and the creation of an efficient development platform still require further research.

This thesis expands medical image segmentation (MIS) research in three ways. Firstly, it develops three-dimensional (3D) flexible and deformable multi-organ models, which are investigated using methods based on principal component analysis (PCA) and neural networks. The results expand the deformability of the multi-organ models but also address the general sparsity of samples in large dimensions. Secondly, existing knowledge of deformable models and human interaction is adapted for the simultaneous segmentation of multiple organs. The results demonstrate that such a technique can segment multiple organs quickly and accurately. The segmentation method is also stable or 'robust.' Thirdly, the thesis develops a software named AnatomySketch (AS) for implementing efficient medical imaging algorithms, enabling the aforementioned techniques. The software includes a flexible plug-in interface and a user-friendly graphical user interface (GUI), facilitating the creation and testing of rapid prototypes. The results show that the software bridges the gap between laboratory prototypes and clinical work and accelerates the development of imaging algorithms.

The thesis successfully demonstrates deformable organ model construction and performs simultaneous segmentation of multiple organs. Additionally, the AS software has proven to be effective in the development of medical imaging algorithms.

Keywords: Medical image analysis (MIA), statistical shape model (SSM), principal component analysis (PCA), stacked autoencoder (SAE), mouse micro-CT images, user interaction, deep learning (DL), AnatomySketch (AS)

TIIVISTELMÄ (ABSTRACT IN FINNISH)

Chen, Zhonghua
Monielimien lääketieteellisten kuvien analysointi, mallinnus ja segmentointi ennalta olevaa tietoa hyödyntäen
Jyväskylä: Jyväskylän yliopisto, 2024, 96 s.
(JYU Dissertations
ISSN 2489-9003; 749)
ISBN 978-951-39-9936-0 (PDF)

Lääketieteellisen kuva-analyysin alalla kohteiden tarkka segmentointi on vaativaa. Viime vuosikymmenien aikana kehitetyt segmentointimenetelmät ovat tarkoitettu pääasiassa yksittäisille kohteille. Useiden kohteiden segmentointiin kerralla ei edelleenkään ole sujuvia prosesseja. Lääketieteellinen kuva-analyysi sisältää monia vaiheita ja on työläs. Nykyiset kehitysalustat eivät mahdollista useiden vaiheiden tehokasta yhdistämistä, mikä hidastaa alan kehitystyötä. Tästä syystä useiden kohteiden samanaikainen tarkka segmentointi sekä tehokkaan kehitysalustan luominen tarvitaan lisätutkimusta.

Tämä väitöskirja laajentaa lääketieteellisen kuvan segmentoinnin tutkimusta kolmella tavalla. Ensinnäkin työssä on kehitetty kolmiulotteisia joustavia ja muunneltavia monielinmalleja, joita lähestyttiin nk. pääkomponenttianalyysin ja neuroverkkomenetelmien avulla. Tulokset laajentavat mahdollisuuksia monielinmallien muodonmuutoksille, mutta työssä myös käsitellään yleistä näytteistyksen harvuuden ongelmaa. Toiseksi työssä sovitetaan vallitsevaa ymmärrystä muunneltavista malleista ja ihmisen vuorovaikutuksesta useiden elinten samanaikaiseen segmentointiin. Tulokset osoittavat, että useita elimiä voidaan segmentoida nopeasti ja tarkasti. Segmentointimenetelmä on myös stabiili eli ”robusti”. Kolmanneksi työssä on kehitetty AnatomySketch nimetty ohjelmisto tehokkaiden lääketieteellisten kuvantamisalgoritmien toteuttamiseen perustuen edellä mainittuihin tekniikoihin. Ohjelmisto sisältää joustavan plug-in-liittymän ja käyttäjäystävällisen graafisen käyttöliittymän, mikä mahdollistaa nopeiden prototyyppien luomisen ja testaamisen. Tulokset osoittavat, että ohjelmisto kaventaa laboratorioprototyyppien ja kliinisen työn välistä kuilua ja nopeuttaa kuvantamisalgoritmien kehitystä.

Työ demonstroi muunneltavia elinmallien konstruoinnin sekä mallintaa ja segmentoitu onnistuneesti useampia elimiä yhdellä kertaa. Lisäksi AnatomySketch-ohjelmisto on osoittautunut toiminnallisuudeltaan tehokkaaksi lääketieteellisten kuvantamisalgoritmien kehityksessä.

Avainsanat: Lääketieteellinen kuva-analyysi, tilastollinen muotomalli, pääkomponenttianalyysi, pinottu autoenkooderi, hiiren mikro CT-kuvat, käyttäjävuorovaikutus, syväoppiminen, AnatomySketch.

Author
Zhonghua Chen
Faculty of Information Technology
University of Jyväskylä
Finland

Supervisors
Lauri Kettunen
Faculty of Information Technology
University of Jyväskylä
Finland

Hongkai Wang
School of Biomedical Engineering
Dalian University of Technology
China

Fengyu Cong
Faculty of Information Technology
University of Jyväskylä
Finland

Tapani Ristaniemi
Faculty of Information Technology
University of Jyväskylä
Finland

Reviewers
Jari Hyttinen
Faculty of Medicine and Health Technology
Tampere University
Finland

Jianxin Zhang
School of Computer Science and Engineering
Dalian Minzu University
China

Opponent
Gongning Luo
School of Computer Science and Technology
Harbin Institute of Technology
China

ACKNOWLEDGEMENTS

In the blink of an eye, four years have passed. My doctoral studies are coming to an end. I still remember my initial fear of communicating with others when I arrived in Finland, feeling lonely and out of place. However, over the past four years, I have made many friends and become more confident in communicating with others. Most importantly, I have become more independent and grown in so many ways. I am deeply grateful to many people who have selflessly helped me throughout my doctoral studies.

I would like to thank sincerely my supervisor, Professor Tapani Ristaniemi. He was a friendly and patient gentleman who frequently organized seminars to discuss the challenges students faced in their work. He always gave relevant suggestions for each issue. While motivating students, he never placed too much pressure on anyone. In addition, he was a very approachable professor who often found ways to help students solve problems in their lives. He treated everyone as an equal friend, which had left a profound impression on me.

I would also like to express my sincere gratitude to Professor Hongkai Wang. He has been my supervisor from the start of my master's studies until the end of my Ph.D. studies. He is a refined and modest gentleman who introduced me to the field of medical image processing and imparted his professional knowledge to me. Additionally, he provides guidance on scientific research and paper writing from various perspectives. For study, he always patiently assists me in solving research problems. He regularly organizes seminars to discuss and solve various issues students encounter. Furthermore, he impresses me with his unwavering patience and meticulousness in helping students revise their papers. In terms of life, he encourages students to approach life with optimism and proactivity. He also takes the initiative to help students overcome any difficulties they encounter. He treats students as equal friends. These valuable experiences will positively influence my future endeavors.

I also want to express my sincere gratitude to my supervisor, Lauri Kettunen. He is a very nice and knowledgeable gentleman, who has been incredibly supportive throughout my Ph.D. studies. We discuss weekly the work progress, the scientific research challenges, and difficulties I face. What really impresses me is his willingness to help find solutions to any issues I encounter. Like Professor Hongkai Wang, he is also very meticulous when helping me revise my papers, demonstrating his strong sense of responsibility. Professor Lauri has always treated everyone with equal respect and dignity. His personality charm has positively affected my life. He never imposes unreasonable demands on students, which helps to prevent unnecessary stress during scientific research. Furthermore, he consistently demonstrates a willingness to offer as much assistance as possible to students. His invaluable support and guidance throughout my Ph.D. studies are greatly appreciated.

My sincere thanks to Professor Fengyu Cong for providing me with the precious opportunity to pursue a Ph.D. degree in Finland. He is a highly ambitious professor who holds students to a high standard, both in academic research and in personal life. He is always selfless in providing help to students when they encounter difficulties. When I face confusion and depression, he encourages and inspires me to overcome every obstacle. He is a highly responsible teacher who truly cares for students.

I would like to express my gratitude to my reviewers, Professor Jari Hyttinen from the Tampere University and Professor Jianxin Zhang from Dalian Minzu University. I am grateful for the time they dedicated to provide valuable and detailed comments on my thesis. I would also like to thank Assistant Professor Gongning Luo from Harbin Institute of Technology for serving as my opponent during my doctoral defense.

I am also grateful for the assistance provided by my friends and colleagues Ye Han, Haoyu Zhai, Mingrui Zhuang, Xiaobang Sun, Jiaqi Zheng, Lili Tian, Liting Song, Dong Tang, Xiangyu Long, Xiulin Wang, Deqing Wang, Rui Yan, Guanghui Zhang, Xiaoshuang Wang, Huashuai Xu, Dongdong Zhou, and many others. They have been with me throughout my Ph.D. studies in Finland. I have had many unforgettable and wonderful experiences with them. Their companionship and support have been invaluable to me, and I am grateful for their friendship.

I also thank the University of Jyväskylä, the general program of the National Natural Science Fund of China (No. 81971693, 81401475, 61971445, and 61971089), the Science and Technology Innovation Fund of Dalian City (2018J12GX042), the Fundamental Research Fund Central Universities (No. DUT19JC01 and DUT20YG122), the National Key Research and Development Program (No. 2020YFB1711500, 2020YFB1711501, and 2020YFB1711503), the Liaoning Key Lab of IC & BME System, the Dalian Engineering Research Center for Artificial Intelligence, and the scholarships from China Scholarship Council (No. 201806060163) for providing financial support for my research.

Last but not least, I express my heartfelt appreciation to my parents and sister. Their unconditional love, unwavering support, and endless encouragement throughout my Ph.D. studies help me a lot. Their selfless love has been the driving force that has kept me moving forward.

Jyväskylä, Finland
March 28, 2023
Zhonghua Chen

ACRONYMS

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
AAM	Active appearance model
AID	Annotation-by-iterative-Deep-Learning
ALS	Alternative least squares
ANN	Artificial neural network
AS	AnatomySketch
ASD	Averaged surface distance
ASM	Active shape model
CNN	Convolutional neural network
CT	Computed tomography
DL	Deep learning
EVD	Eigenvalue decomposition
FCN	Fully convolutional network
FFD	Free-form deformation
FOV	Field of view
GAN	Generative adversarial network
GUI	Graphical user interface
HDLSS	High-dimensional-low-sample-size
MIA	Medical image analysis
MIS	Medical image segmentation
MRI	Magnetic resonance imaging
NMR	Nuclear magnetic resonance
PCA	Principal component analysis
PD	Proton density
PDM	Point distribution model
PET	Positron emission tomography
RF	Radiofrequency
ROI	Region of interest
SAE	Stacked autoencoder
SGD	Stochastic gradient descent
SSM	Statistical shape model
SVD	Singular value decomposition
TE	Echo time
TR	Repetition time
VHRBF	Variational Hermite Radial Basis Function

FIGURES

TABLES

CONTENTS

ORIGINAL PAPERS

LIST OF INCLUDED ARTICLES

I **Zhonghua Chen**, Tapani Ristaniemi, Fengyu Cong, and Hongkai Wang (2020). Multi-resolution statistical shape models for multi-organ shape modelling. In Advances in Neural Networks – ISNN 2020: 17th International Symposium on Neural Networks, pp. 74-84, Springer, Cham. https://doi.org/10.1007/978-3-030-64221-1_7.

II **Zhonghua Chen,** Hongkai Wang, Fengyu Cong, and Lauri Kettunen (2022). Construction of multi-resolution multi-organ shape model based on stacked autoencoder neural network. In 2022 14th International Conference on Advanced Computational Intelligence (ICACI), pp. 62-67. IEEE. https://doi.org/10.1109/ICACI55529.2022.9837706.

III **Zhonghua Chen,** Hongkai Wang, Fengyu Cong, and Lauri Kettunen (2022). Low-dose mouse micro-CT image segmentation based on multi-resolution multi-organ shape prior knowledge model. In 2022 International Conference on Computers, Information Processing and Advanced Education (CIPAE), pp. 349-353. IEEE. https://doi.org/10.1109/CIPAE55637.2022.00079.

IV Mingrui Zhuang, **Zhonghua Chen,** Hongkai Wang, Hong Tang, Jiang He, Bobo Qin, Yuxin Yang, Xiaoxian Jin, Mengzhu Yu, Baitao Jin, Taijing Li, and Lauri Kettunen (2022). AnatomySketch: An extensible open-source software platform for medical image analysis algorithm development. Journal of Digital Imaging, 35, 1623–1633, https://doi.org/10.1007/s10278-022-00660-5.

1 INTRODUCTION

This chapter begins by introducing the research background. Then it presents research motivation for the thesis. It concludes by outlining the structure of the thesis.

1.1 Research background

In real-world scenarios, the majority of medical images are unlabeled due to their inherent complexity. This presents a significant challenge in accurately identifying various organs or tissues in clinical applications. Precise segmentation offers reliable shape and volumetric information, playing a crucial role in disease diagnosis and quantitative analysis within clinical settings. We will provide a more detailed introduction related to MIS in the following three aspects.

1.1.1 The nature of medical image segmentation

Medical imaging plays a crucial role in various healthcare settings. The use of medical imaging is irreplaceable in identifying, assessing, and documenting the progression of different types of diseases, as well as the response to treatment. For example, computed tomography (CT) and magnetic resonance imaging (MRI) are widely used for lesion detection in internal organs, bones, and tissues in humans. However, the quality of medical images is often poor, resulting in blurred boundaries between different regions, such as different organs. As a result, it can be time-consuming and labor-intensive for doctors to read medical images for diagnosis. Consequently, researchers are dedicating efforts towards developing different MIS methods to segment different regions of interest (ROIs). Designing segmentation methods based on the boundaries of different ROIs is a common trend in MIS.

So, what is meant by the term "boundary"? Segmentation is closely related to the notion of a boundary. In the context of point-sets, the boundary of a set X, is denoted by ∂X. The boundary is what remains when the interior points of X are subtracted from its closure. Alternatively, one may define the set of boundary points by stating that a point x is on the boundary of the point set X, i.e., x belongs to ∂X, if all the open neighborhoods of x contain points both from X and its complement. It is important to note that both specifications of a boundary assume that X is a proper subset of some ambient point-set.

The concept of a boundary is intuitive in the context of point-sets but recognizing boundaries in digital medical images presents certain challenges. If we consider a subset of n-dimensional space, denoted as X, the dimension of its boundary is (n-1). This means that the boundary of any 3D organ or anatomical part is a 2-dimensional (2D) surface with no thickness. Consequently, medical images do not provide any explicit indication of the boundary. Moreover, every digital image is composed of a finite set of data points, rendering the definition of a boundary meaningless, at best. Therefore, a segmented image object can only be an informed approximation of the boundary.

The concept of a boundary in medical images can still be preserved by introducing a mapping from $\mathbb{R}^2$ to the image. For instance, consider a medical image resembling a standard photograph, which is a plane. By introducing a mapping from $\mathbb{R}^2$ to the image plane, every point in the image can be identified with a pair of coordinates $(x, y) \in \mathbb{R}^2$, as shown in Figure 1. Consequently, even though the image plane consists of a finite set of pixels, it can still be regarded as a point-set, and the definition of a boundary is well-defined in this context.

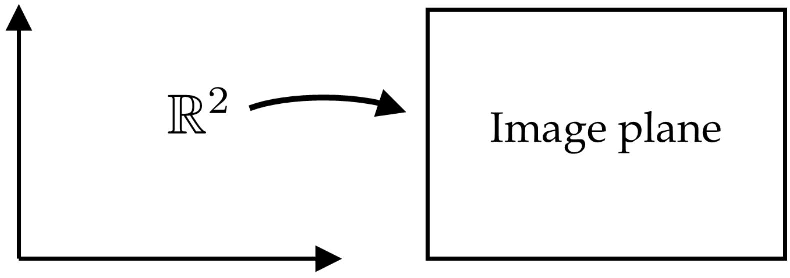


FIGURE 1 A map from $\mathbb{R}^2$ to an image plane parametrizing the image domain with pairs of real numbers

Formally such a construction is known in mathematics as a **manifold**. Such interpretation of medical images corresponds well with the clinical diagnostic practices. For instance, a purpose of segmenting medical images is to measure the size of objects, and this calls for some further structure. Image segmentation, i.e., finding the boundary of a ROI, is about **topology**. In other words, to find a boundary one does not need a ruler and to know of distances. But when it comes to measuring sizes of objects one needs a **metric** structure. Such metric can be introduced with the map

$$\mathbb{R}^2 \rightarrow \text{image plane},$$

by equipping $\mathbb{R}^2$ with an inner product – meaning given any pair of points, their distance is also known–that provides one with the information of distances.

Summing up, we may now conclude, MIS is formally about finding the boundaries of ROIs from images that are consider as maps

$$\mathbb{R}^n \rightarrow \text{image.}$$

1.1.2 Review of medical image segmentation

MIS is a vital technique for the delineation and extraction of ROIs in medical research and clinical applications. MIS aids computer-aided diagnosis, the study of anatomy, tissue detection, and surgical planning (Elnakib et al., 2011; Norouzi et al., 2014; Nosrati and Hamarneh, 2016; Pham et al., 2000). It frees experts and practitioners from the burden of repeated, tedious, and time-consuming image-reading tasks. Over the years, MIS has emerged as a growing subfield of MIA. Essentially, MIS is the process of partitioning a medical image, such as CT and MRI, into continuous sub-images with corresponding homogeneous characteristics. The objective of MIS is to extract various ROIs, each represented with a different label for different tasks. Typically, the implementation of MIS involves segmentation methods or the integration of segmentation methods with operating platforms.

MIS methods can be divided into two categories based on the types of research routines: conventional MIS methods and deep learning (DL)-based MIS methods. Conventional methods have been in development for several decades. There are many variants that accurately segment specific anatomical structures in different image modalities. These methods include thresholding, region-based, edge-based, clustering-based, atlas-based segmentation, and model-based segmentation (Kang et al., 2009; L. K. Lee et al., 2015; Pham et al., 2000; Ramesh et al., 2021).

Thresholding approaches involve partitioning the pixels of a medical image into different classes with binary values. The segmentation process implements by setting appropriate feature thresholds, such as grayscale. Typically, the number of threshold values corresponds to the groups of image pixels. During the thresholding segmentation process, selecting appropriate threshold values is crucial for achieving optimal results. As a result, numerous approaches have been proposed for selecting proper thresholding values over the past few decades. These approaches include analyzing grayscale distribution (Otsu, 1979), using image histograms (Li et al., 1997), constructing Gaussian mixture models for threshold selection (Fan et al., 2008; Z.-K. Huang and Chau, 2008), utilizing multilevel thresholding algorithms (Adollah et al., 2012; Maitra and Chatterjee, 2008), performing 3D adaptive thresholding (J. Zhang et al., 2010), and employing wavelet-based multiscale product thresholding (Vijay et al., 2012). According to the principles of selecting threshold values, thresholding approaches can generally be classified into three categories: global thresholding, local thresholding, and dynamic thresholding methods (Kang et al., 2009). The grayscale histogram distribution of an image has a significant impact on the segmentation accuracy of thresholding approaches. In practice, thresholding approaches demonstrate good segmentation efficacy on images with bimodal histograms. But they may fail to segment ROIs in a medical image without a bimodal histogram. Additionally, thresholding approaches do not utilize spatial information in an image. These two obvious limitations can lead to segmenting anatomical structures without meaningful anatomy, particularly in MRI images (Norouzi et al., 2014; Zanaty and Ghoniemy, 2016).

According to some given criteria (Haralick and Shapiro, 1985), region-based methods partition an image into different regions where the pixels have the same or similar characteristics, such as grayscale, texture, and homogeneity criteria (Kang et al., 2009; Zanaty and Ghoniemy, 2016). Region-based methods can be mainly divided into two subcategories: (a) region growing and (b) region splitting and merging. Region growing technique needs to set different seeds (i.e., pixels) to start region growing process (Ayman et al., 2013; Wu et al., 2008). Region-based methods has been developed for decades after Zucker proposed the basic region growing technique for image segmentation for the first time (Zucker, 1976). In the early stage of MIS, Heinonen et al. (Heinonen et al., 1998) applied another region growing technique for lesion segmentation in MRI images. Later, Boasch et al. (Bosch et al., 2002) and Scott et al. (Scott et al., 2003) expanded region growing methods into several medical image modalities, such as ultrasound images and X-ray images. Their work has made a significant influence on other following variants of region growing. Region growing methods based on seed selection criterion are often faster when the seeds are chosen properly (Zanaty and Ghoniemy, 2016). However, the ways for choosing a seed in an image are concerned with the segmented results greatly (Adams and Bischof, 1994; Ayman et al., 2010). As the name implies, the region splitting and merging technology is that users first divide an image into various disconnected regions. And then users merge these regions according to the given segmentation criteria to obtain the segmented results. The implementation of this kind of method depends on the principle of quad tree (Kelkar and Gupta, 2008; Malik et al., 2001). But this technique is susceptible to the number of segmentations and clearness of anatomical boundaries (del Fresno et al., 2009; Zanaty and Asaad, 2013). Consequently, region splitting and merging algorithms tend to consume significant computational resources and time due to their memory requirements.

An image is composed of various regions made up of different pixels, such as gray or colored pixels. Pixels are separated by boundaries where the intensity values change abruptly, such as texture variations and color discontinuities (Kang et al., 2009). Edge-based methods utilize distinct boundary discontinuities to detect all edges and connect them to form different regions (Farag, 1992). Gray histograms and gradient-based methods, as two basic and classic edge-based approaches, have laid the foundation for other variants of edge-based methods that followed (Kaur and Kaur, 2014). Edge detection, one of the most popular gradient-based methods, has given rise to several well-known algorithms such as Prewitt edge detection (Prewitt, 1970), Sobel edge detection (O. R. Vincent and Folorunso, 2009), Canny edge detection (Canny, 1986), and Roberts edge detection (Cherri and Karim, 1989). As edge detection has advantages for object recognition in high-contrast images, in 2006, Zhao et al. (Yu-Qian et al., 2006) proposed a new mathematical morphological edge detection algorithm for segmenting lungs in CT scans. This method showed better segmentation and denoising results. However, edge-based methods are not suitable for images with blurred boundaries and are vulnerable to the number of boundaries, especially for MRI images.

Another well-known image segmentation method is the watershed method, which has been attributed to both region-based segmentation methods in some studies (Hamarneh and Li, 2009; Kang et al., 2009), and edge-based segmentation methods in others (L. K. Lee et al., 2015). However, several review studies have classified watershed as a separate category (Kaur and Kaur, 2014; Pham et al., 2000; Zanaty and Ghoniemy, 2016). The watershed approach mimics the structure of a geographical watershed. It interprets low-intensity pixels as valleys and high-intensity pixels as mountain peaks. Thus, it enables segmentation of different regions within an image. This method is widely used in MIS because it can produce better division of an image. However, the disadvantage of the watershed method is that it is prone to generating false edges and over-segmenting an image.

Clustering-based methods are widely used in MIS because they can effectively separate pixels with similar characteristics into the same class while distinguishing pixels with dissimilar characteristics. These methods cluster pixels together based on their features, enabling them to identify and segment proper objects within an image. Because clustering-based methods do not require training data, they belong to the category of unsupervised learning methods. However, these methods need to iterate between segmentation and refining the characteristics of each cluster to improve their accuracy (Pham et al., 2000). The k-means algorithm (Coleman and Andrews, 1979; Dehariya et al., 2010), fuzzy C-mean algorithm (Dunn, 1973; Jain and Sharma, 2019), and expectation maximization algorithm (Ravindraiah and Tejaswini, 2013; Warfield et al., 2002) are the three classic clustering-based methods used in MIS. One clear advantage of clustering-based methods is their ability to process images quickly. However, they do not take advantage of spatial information within an image and are susceptible to homogeneous noise (Norouzi et al., 2014). Furthermore, while clustering does not require training data, the results can be highly dependent on the initial parameters fed into the algorithms. This can compromise the accuracy of the segmentation.

Deformable model-based methods are another common and popular approach in MIS. These methods partition images into different regions by minimizing energy functions that measure the variation within training data and the differences between training and testing data (Caselles et al., 1993; Kass et al., 1988; Terzopoulos and Szeliski, 1992). Deformable models can be trained by using prior knowledge, such as the shape and texture of an image, to optimize energy functions and achieve good deformability. In medical applications, deformable modeling is used to delineate, segment, and extract ROIs from medical images, such as the segmentation of organs and tissues, delineation of parametric curves or surfaces. This means that models can deform regularly under the influence of internal and external forces to generate object boundaries (Zhao and Xie, 2013). Statistical shape models (SSMs) have been proposed as a robust tool for segmenting and extracting ROIs in medical images, especially for 3D ROIs (Heimann and Meinzer, 2009). The two classic methods for statistical shape modeling are active shape models (ASMs) (Cootes et al., 1995) and active appearance models (AAMs) (Cootes et al., 2001), which have greatly influenced other variants of SSMs.

In conclusion, conventional MIS methods have been extensively developed and applied to the segmentation of various anatomical structures. For specific anatomical structures, there are corresponding segmentation methods, which significantly reduce the workload of doctors in clinical diagnoses. Figure 2 illustrates the general types of conventional MIS methods.

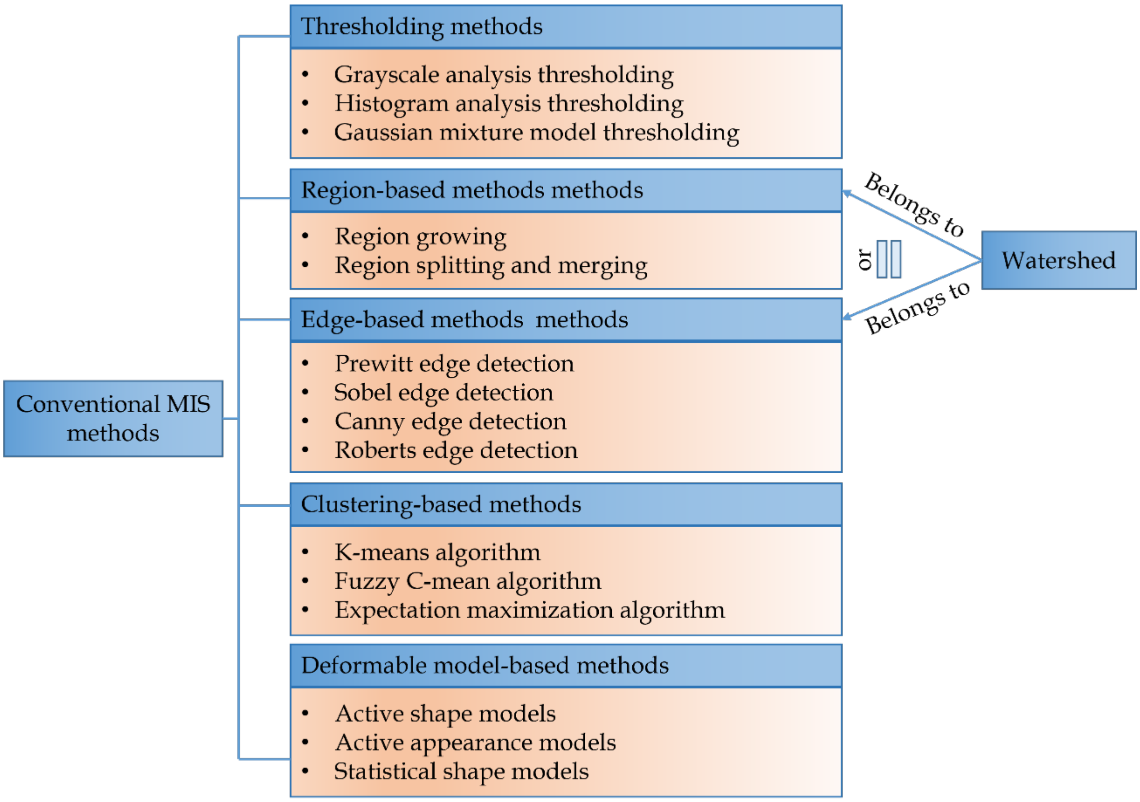


FIGURE 2 General types of conventional MIS methods

With the rapid development of hardware and software in computer-aided diagnosis, the quality of medical imaging has improved dramatically (Sahiner et al., 2019). The advances in these techniques have promoted the development of DL methods (LeCun et al., 2015) for MIS. As an emerging subfield of MIS, DL methods for MIS have increasingly become a research hotspot. Essentially, DL methods have evolved from binary classification in machine learning which belongs to artificial intelligence. But DL methods are much more complicated because they are presented with deep neural networks. Intuitively, DL methods are invented to simulate the learning process of humans. They extract useful information from various datasets such as, text, sound, images, videos, etc. in an unsupervised (C. Chen et al., 2020; J. Chen and Frey, 2020), semi-supervised (Shuai Chen et al., 2019; Nie et al., 2018), or supervised (Gu et al., 2018; R. Wang et al., 2022) manner. Deep neural networks, the concrete format of DL methods, are composed of multiple layers of neurons. Neurons are regarded as information processing units and connected in a certain way according to some mathematic rules. For MIS, DL methods such as fully convolutional network (FCN) (Bi et al., 2018; Roth, Oda, et al., 2018), U-net (Du et al., 2020; Siddique et al., 2021), and generative adversarial network (GAN) (Creswell et al., 2018; Goodfellow et al., 2020; Xun et al., 2022) in various studies are based on the convolutional neural network (CNN) (O'Shea and Nash, 2015). Research in recent years has shown that compared to conventional MIS methods, DL methods can segment ROIs in images faster and more accurately. Some examples include 3D MRI image segmentation based on FCN (Milletari et al., 2016), liver and spleen segmentation in CT images based on UNet 3+ (H. Huang et al., 2020), and tumor segmentation in MRI images based on GAN (Cirillo et al., 2021; Nema et al., 2020; Sun et al., 2020). Therefore, DL can further effectively help doctors obtain the desired results in clinical applications, thereby greatly reducing their workload. In addition, during the COVID-19 pandemic, DL methods have been widely used for quantitatively evaluating or extracting infected areas in patients' lungs from different image modalities (Han et al., 2022). This helps doctors make faster and more accurate diagnoses to save more lives

In conclusion, Zhao et al. pointed out that some segmentation methods can be also attribute to one of the following three categories: unsupervised (automatic) methods, semi-supervised (semi-automatic) methods, and supervised methods (Zhao and Xie, 2013). Figure 3 shows the classification of main segmentation methods in the domain of MIS.

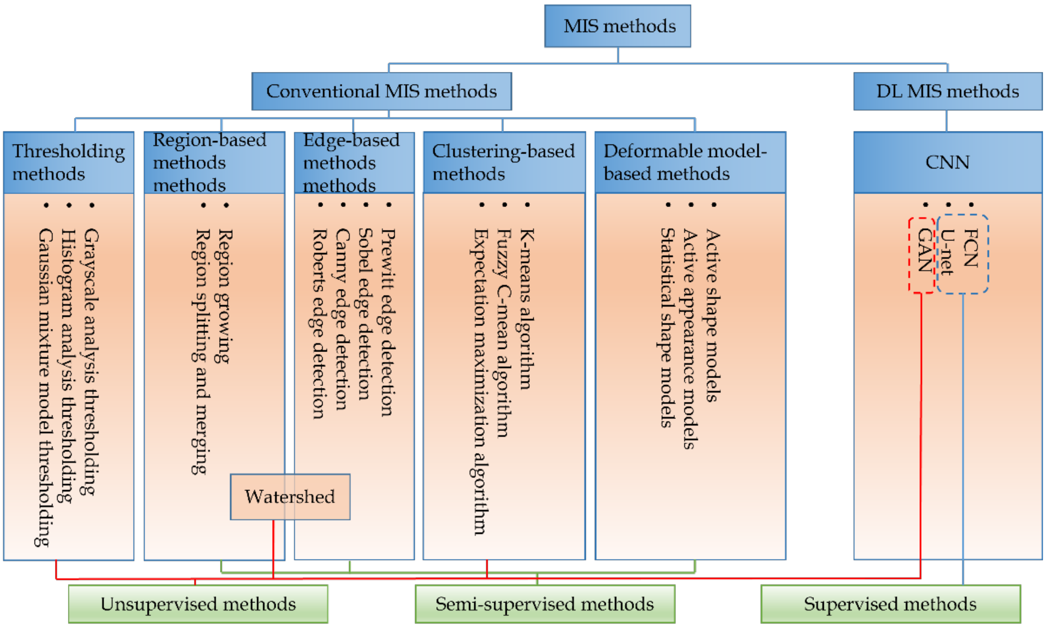


FIGURE 3 Classification of some popular segmentation methods in the domain of MIS

1.1.3 Prior knowledge

In practical application, various MIS methods are susceptible to different types of noise. These noises primarily originate from different image modalities, poor contrast in certain regions, irregular shapes of anatomical structures, and complex boundaries of organs. These factors can limit methods' performance in achieving accurate segmentation. However, previous studies have shown that incorporating prior knowledge into different MIS algorithms can result in more reliable and accurate results (Grau et al., 2004; Kim et al., 2005). Prior knowledge of medical images, such as CT or MRI images, must be further interpreted to enable effective segmentation. Medical images typically consist of various grey pixels, with intensity values that do not necessarily range from 0 to 255. Medical images contain multiple types of prior knowledge that can be applied in different segmentation methods, including texture, grayscale of pixels, volume, shape, and appearance (Xi et al., 2017). Additionally, there is associated geometric information for different regions, location and distance of anatomical structures, atlases of images, and deformable models in medical images (Elnakib et al., 2011; Heimann and Meinzer, 2009; G. Hu, 2009; Nosrati and Hamarneh, 2016; Pham et al., 2000).

In the past few decades, there have been numerous studies that thoroughly reviewed the development and applications of MIS. Combining prior knowledge with these studies has played a crucial role in improving the accuracy of MIS. For example, the use of deformable models in medical images has been extensively studied (Egmentation et al., 2000; Elnakib et al., 2011; Heimann and Meinzer, 2009; G. Hu, 2009; McInerney and Terzopoulos, 1996). Additionally, user interaction has been studied to improve the accuracy of segmentation (Olabarriaga and Smeulders, 2001). More recently, the use of deep atlas prior for liver and spleen segmentation has shown promising results (Huimin Huang et al., 2021). The incorporation of these prior knowledge approaches has contributed significantly to the advancement of MIS.

In 3D MIS, SSM, as a kind of shape prior knowledge, is often used to segment specific individual anatomical structure. Most of the previous studies used SSMs to segment

- Specific organs, e.g., the liver in the human body (Lamecker et al., 2004), brain regions (Shen et al., 2001), heart ventricles (Grosgeorge et al., 2013),
- Target lesions, e.g., tumors (Wilms et al., 2012),
- Bones, e.g., lower limbs trunk (Audenaert et al., 2019).

However, various organs are physically and functionally related to each other in an image. Segmentation of individual organs oversimplifies the complicated human anatomy. Such methods seldom consider the interaction between organs, which limits the segmentation performance. The complicated human anatomy therefore makes the development of segmentation methods targeting multiple organs difficult. Although there are fewer studies on multiple organs than single-organ studies, more accurate and comprehensive segmentation methods are constantly being developed. Some literature has pointed out that multi-objective modeling and analysis has attracted extensive attention of researchers (Leal-Taixé et al., 2017; Luo et al., 2021). The evidence is that the research on multi-organ segmentation has shown exponential growth (Cerrolaza et al., 2019), such as multi-resolution shape (Cerrolaza et al., 2011; Davatzikos et al., 2003), multi-organ model (Chaney et al., 2004; C. Lu et al., 2007; Pizer et al., 2003; Vera et al., 2011), DL model (Gibson et al., 2018; P. Hu et al., 2017; Roth, Shen, et al., 2018; Y. Wang et al., 2019; X. Zhou et al., 2017).

Recently, in some state-of-the-art studies, the edge information and region information of anatomical structures have been utilized to enhance the training of deep networks (Shaolong Chen et al., 2023). These shape prior knowledge can improve the segmentation efficiency of the network and achieve better segmentation results. Similarly, in some studies, organ shape prior knowledge has been employed to train shape-adversarial autoencoders for optimizing segmentation networks, thereby improving the segmentation results of the network (Lei et al., 2023). Furthermore, the fusion of boundary information and region information of lesions has also been shown to significantly improve the segmentation accuracy in certain research and application scenarios (Liu et al., 2022). To date, there have been very few studies on deep learning segmentation networks based on shape prior knowledge, and this research gap presents a potential hotspot for future investigations.

1.2 Research motivation

Although numerous MIS methods have been applied widely in various clinical applications and research, there is always room for developing better methods for MIS, especially with the incorporation of prior knowledge. However, conventional MIS methods focus on the segmentation of specific anatomical structures, neglecting inter-organ relations (Cerrolaza et al., 2019). Furthermore, medical images exhibit poor quality and complicated shapes of different anatomical structures. These limitations can result in obtaining poor or opposite ROIs by parametrizing each anatomical structure independently. Therefore, developing appropriate approaches to segment multiple ROIs simultaneously is a significant challenge for researchers. However, research shows that there are fewer studies focusing on multi-organ analysis compared to single-organ analysis (Cerrolaza et al., 2019). Some famous methods were proposed to handle specific problems of single-organ analysis without prior knowledge of inter-structure (Cootes et al., 1995; Gerig et al., 2001; Leventon et al., 2002; Pizer et al., 1999). In contrast, early multi-organ segmentation methods were relatively simple but computationally intensive (Cates et al., 2008; Frangi et al., 2002; Pizer et al., 2003).

Recently, more studies have used DL methods for multi-organ segmentation, which have further improved the accuracy of segmentation (Fu et al., 2021; Gibson et al., 2018). However, due to the complexity of multi-organ segmentation, research in this area needs more exploration. Ideally, multi-organ segmentation can greatly improve the efficiency of obtaining multiple targets simultaneously in medical images, making further research in this area necessary. Based on this, we identify four research questions:

1. How can we create a sophisticated deformable model for multiple organs, such as livers, left kidneys, right kidneys, and spleens, based on an improved PCA algorithm for low-contrast mouse micro-CT images? How can we address computational burden for a large-scale data during the process of model construction?
2. Can artificial neural networks (ANNs), such as stacked autoencoder (SAE) networks, be used to build multi-resolution multi-organ deformable models with greater deformability, incorporating more non-linear deformation components to generate additional deformable modes?
3. How can we leverage the shape prior knowledge of the model to achieve more accurate segmentation of multiple organs in medical images? Should we integrate an algorithm into the segmentation process for improved accuracy, or is human interaction a viable approach for achieving better performance?
4. Given the complexity of MIA, including multi-organ segmentation, software that integrates various medical image processing methods plays a crucial role. Should we develop an extensible open-source platform with a user-friendly GUI that can integrate multiple methods to facilitate medical image processing?

So, the motivation behind this dissertation is to provide solutions to these problems.

1.3 Structure of dissertation

The rest of this dissertation is organized as follows. Chapter 2 introduces the experimental data used in the published articles, together with methods used in these different articles. Chapter 3 introduces the corresponding evaluation methods. Chapter 4 summarizes the four articles included in this thesis. Chapter 5 discusses the experimental data, methodologies, limitations and future directions. Chapter 6 concludes the whole thesis.

2 MATERIALS AND RELATED WORK

This chapter commences by introducing the concepts of common medical image modalities in Section 2.1 and elucidating the image data employed in the study in Section 2.2. Subsequently, the research pertaining to PCA and ANN for modeling is comprehensively reviewed in Sections 2.3 and 2.4, respectively. Furthermore, Section 2.5 presents the methodology of multi-organ segmentation based on a constructed model and Variational Hermite Radial Basis Function (VHRBF) (Brazil et al., 2010). Finally, Section 2.6 provides an in-depth description of the comprehensive software developed for MIA. The aim of this chapter is to provide the reader with a more comprehensive understanding of common medical image modalities and the methods employed in our study.

2.1 Common medical imaging modalities

The role of medical images in health monitoring and disease diagnosis has become increasingly significant. Medical images are generated by exposing mammal bodies to various media, such as X-rays, ultrasonic waves, and electromagnetic fields, which interact with the tissues, organs, bones inside the body and produce images (Webb, 2022).

However, the quality of such medical images are not comparable to those of ordinary photographs, which are generated from reflected light measurements. Medical images rely on more advanced techniques and there are technical reasons why the medical images can’t be as sharp or clear as ordinary photos. Accordingly, the segmentation methods developed for photographing are not directly applicable to medical imaging. For this reason, medical images call for specific and tailor-made segmentation techniques.

Next, we will take a brief view on the technical challenges of medical imaging techniques that explain the imaging quality.

2.1.1 X-ray imaging

X-rays, discovered by W. Röntgen in 1895, exhibit varying absorption properties in different materials based on their density (Panchbhai, 2015). X-ray imaging is based on the differential absorption of X-ray energy by various materials. As X-rays pass through substances, the radiation attenuates. This is detected in the contrast of the resulting images. Different materials absorb X-ray energy in varying degrees, leading to variations in brightness and contrast in X-ray images. For instance, bones absorb more X-ray energy compared to paper and water, resulting in distinct contrast between bones and soft tissues (Maier et al., 2018; Miwa and Otsuka, 2017).

X-rays are typically generated with X-ray tubes. They consist of a vacuum tube with a cathode and a metal anode. When the cathode is heated by a filament, thermal electrons are emitted and accelerated. This process involves deceleration and deflection, leading to the production of X-rays. Two types of X-ray radiation are generated: characteristic X-ray radiation, produced when electrons interact with inner-shell electrons of the target material, and bremsstrahlung radiation, caused by interactions with the target atomic nuclei (Seibert, 2004; KT Wilson et al., 2020; Kieran Wilson and Schaub, 2019). Modern X-ray detection systems convert transmitted X-rays into light and then into electronic signals, replacing older X-ray films (Maier et al., 2018).

It is unavoidable X-ray images are somewhat blurry. Firstly, it is impossible to technically construct an ideal point source for X-rays. Consequently, as the X-rays do not come from a single point, an X-ray image can never be completely sharp. Secondly, when X-rays interact with matter, they interact with the electrons and nuclei of the atom and with the associated electric field. This leads not only to absorption, but also to (elastic and inelastic) scattering causing noise and blurriness in the images. Furthermore, the effects are highly dependent on the energy of the X-rays (Boone, 2000; Maier et al., 2018).

Figure 4 displays an X-ray image of a patient's chest[1] demonstrating the slight inherent blurriness of X-ray images. As depicted in this figure, the rib bones in the chest cavity overlap with each other significantly. The distinction of individual bones within the highlighted region is not discernible to human eyes. This inherent low imaging quality of X-ray images poses considerable challenges in MIA.

---

[1] https://unsplash.com/s/photos/x-ray

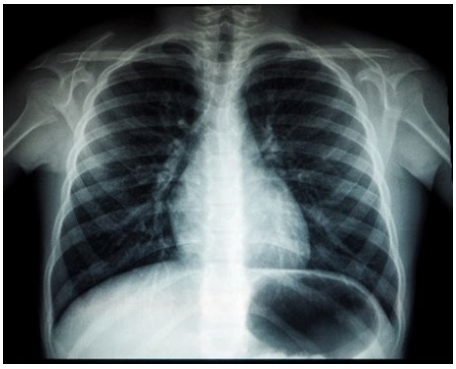


FIGURE 4 An example of a chest X-ray

2.1.2 Computed tomography (CT) imaging

Based on X-ray imaging, CT technology was developed to overcome the limitations of conventional X-ray imaging. The 2D images obtained from X-rays do not preserve all spatial information of the object as they are projections from 3D space to a plane (Maier et al., 2018; Shah et al., 2014). By taking multiple projections from different angles, a 3D reconstruction can be constructed, providing one with a sectional view of the object. This concept was the original motivation behind the development of CT scanning.

The first CT system was constructed in 1971 by G. N. Hounsfield and A. McLeod Cormack (Cierniak and Cierniak, 2011). Subsequently, the introduction of spiral CT by W. Kalender et al. marked a significant advancement, offering faster data acquisition, shorter reconstruction times, as well as higher spatial resolution and quantization depth (Kalender et al., 1990).

In recent years, significant progress has been made in CT technology regarding temporal and spatial resolution (Leng et al., 2015; J. Wang and Fleischmann, 2018). Among them, the development of dual-source CT in 2005 represents a significant milestone (Brodoefel et al., 2008). It incorporates two X-ray sources and detectors within a single scanner, leading to a substantial increase in scanning speed. Additionally, the number of slices acquired in parallel has also been increased, expanding the imaging field of view and reducing motion artifacts in the imaged object.

As CT relies on X-ray images, evidently, it shares the same technical issues that result in blurry images. However, taking several projections makes it possible to significantly improve the image quality. That is, each projection, i.e., each X-ray image can be consider as a sample of the imaged object. A sufficiently large number of projections makes it possible to solve (with a reasonable accuracy) an inverse problem. The inverse problem arises because the acquired X-ray projections do not directly provide information about the internal structure in a straightforward manner (G. T. Herman, 2009). The goal of the inverse problem is to determine the characteristics and structure of the object that has generated these samples. In principle, in the context of CT imaging, the inverse problem can be established by assuming a non-ideal X-ray source. This means that the X-ray source need not be a perfect point source improving the image quality. However, noise caused by scattering can further degrade the quality of the acquired data. Various mathematical techniques and algorithms (G. T. Herman, 2009; Kak and Slaney, 2001; Natterer, 2001), such as filtered backprojection and iterative reconstruction methods (Kak and Slaney, 2001; Natterer, 2001), are employed to address the inverse problem and mitigate the impact of noise and imperfections. The process typically involves finding a reconstruction algorithm that can sufficiently accurately recover the internal structure of the object based on the acquired projections.

The mathematical principles underlying the reconstruction of CT images primarily involve the Radon transform and the Fourier slice theorem (Jiang Hsieh, 2003; Kak and Slaney, 2001; Natterer and Natterer, 1986). The Radon transform states that any integrable function can be uniquely represented by integrating over all lines in its domain. The function integrable function can be reconstructed from these projections. The Fourier slice theorem establishes an equivalence relationship between the Fourier transform of a projection and the original function. It allows an estimation of the original function by computing the Fourier transform of the projections. Based on these principles, two main reconstruction methods are employed in CT imaging. The first one utilizes the Radon transform. It employs the Beer-Lambert law, which converts X-ray projections into line integrals for CT reconstruction (Seeram, 2018). The second method is filtered backprojection, which exploits the Fourier slice theorem to directly invert the projection process for CT reconstruction, without requiring calculations in the frequency domain (Pontana et al., 2011).

Building upon these principles, the introduction of multi-row detectors has greatly facilitated the advancement of CT technology (Fischbach et al., 2003; Wormanns et al., 2004). This improved CT technology enables multi-slice CT scanning, significantly enhancing imaging speed. Additionally, cone-beam CT employs image intensifiers or the latest flat-panel detectors to capture objects with a large field of view in a single rotation (Abramovitch and Rice, 2014; Scarfe and Farman, 2008). Cone-beam CT finds wide applications in interventional imaging and other fields. However, it is important to note that object reconstruction algorithms rely on a rotational plane. Consequently, deviations of the beam from this plane tend to introduce uncertainties and artifacts in the reconstruction process.

In practical usage, several factors need to be considered when constructing CT images (Davis and Vachhani, 2017; Maier et al., 2018):

- Spatial resolution: Spatial resolution refers to the CT system's ability to distinguish small details in the image. It is influenced by factors such as detector pixel size, focal spot size, and reconstruction algorithms. Higher spatial resolution enables the visualization of fine structures.
- Noise: CT images inherently contain noise due to the statistical characteristics of X-ray detection. Noise can degrade image quality and reduce diagnostic confidence. Techniques such as tube current modulation and iterative reconstruction algorithms are employed to reduce noise while preserving image quality (Vardhanabhuti et al., 2013).

- Image artifacts: Artifacts are unintended features or distortions present in CT images that do not accurately represent underlying anatomical structures. Various factors can cause artifacts, including patient motion, metal objects, beam hardening, and patient size. Artifacts can affect image interpretation and diagnosis.
- Beam hardening: Beam hardening occurs when the X-ray beam passing through an object becomes harder (more penetrating) as it traverses the material. This can lead to cupping artifacts and inaccurate Hounsfield unit measurements, especially in dense objects like bones (Boas and Fleischmann, 2012). Correction algorithms are used to minimize the impact of beam hardening artifacts.
- Scatter artifacts: Scatter artifacts are caused by X-ray photons deviating from their original path due to interactions with patient tissues. These scattered photons can reduce contrast and result in streak artifacts, thereby reducing image quality. Scatter correction techniques are applied to mitigate these artifacts (Baer and Kachelrieß, 2012).
- Partial volume effect: The partial volume effect occurs when a voxel (3D pixel) contains a mixture of different tissue types, leading to inaccurate CT numbers and blurred tissue boundaries. High-resolution imaging, thin-slice acquisition, and advanced reconstruction algorithms can help reduce the impact of partial volume effect (Pelc, 2014).
- Metal artifacts: Metal implants or objects within the patient's body can cause severe artifacts in CT images. These artifacts manifest as streaks, dark or bright bands, and geometric distortions, making it challenging to assess surrounding anatomical structures. Techniques to minimize metal artifacts are employed to mitigate their impact.
- Motion artifacts: Patient motion during the image acquisition process can result in motion artifacts, leading to structural blurring or ghosting. It can be caused by involuntary patient motion, respiration, or cardiac motion. Motion compensation techniques, such as gating or respiratory-triggered imaging, are used to mitigate these artifacts (Goo and Allmendinger, 2017).
- Truncation artifacts: Truncation artifacts occur when the imaged patient or object extends beyond the field of view (FOV) of the CT scanner. This leads to data truncation, resulting in ring-like or dark bands in the reconstructed image. Increasing the FOV or applying advanced reconstruction algorithms helps minimize truncation artifacts (J Hsieh et al., 2004; Y. Huang et al., 2020).

These factors highlight that CT image segmentation faces numerous research challenges of significant value. Achieving accurate and reliable segmentation of ROIs has been an ongoing research objective. From the principles and imaging conditions of CT, it is evident that variations in image quality, such as noise, artifacts, and contrast, have a substantial impact on the delineation accuracy of target boundaries. Limited scanner resolution leads to partial volume effects that result in blurring and incorrect labeling. Moreover, the presence of heterogeneous tissue characteristics and overlapping intensities makes tissue differentiation challenging. Additionally, complex anatomical variations pose challenges in handling inter-patient and intra-patient differences using algorithms. Furthermore, the time and computational complexity associated with volumetric data necessitate the continuous development of effective real-time segmentation algorithms. These factors collectively present challenges for the segmentation task. Addressing these challenges requires the development of improved segmentation methods, which is also the primary focus of this thesis.

Figure 5 shows a cross-sectional CT image of a human chest[2]. Compared to X-ray images, CT imaging offers better contrast between different organs within the chest cavity, particularly the boundaries separating the lungs from adjacent organs. This is because that CT imaging utilizes a combination of X-ray attenuation measurements from multiple angles to reconstruct cross-sectional images. The multi-angle approach allows for the differentiation of tissues with similar attenuation coefficients. By capturing X-ray projections from various angles and using sophisticated mathematical algorithms, CT imaging can generate enhanced contrast images between different soft tissues within the chest, including the lungs and adjacent organs.

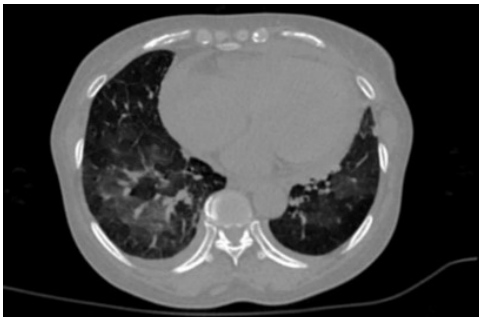


FIGURE 5 A cross-sectional CT image of a human chest

2.1.3 Ultrasonic imaging

In addition to the aforementioned common imaging modalities, ultrasound imaging is also widely used in medical diagnosis. Sound waves are generated by the periodic compression of molecules or atoms (Sarvazyan et al., 2013). These waves beyond a frequency of 20 kHz are referred to as ultrasound waves (Cravotto and Cintas, 2009). Ultrasound can be used for echolocation. Organisms like bats determine the location of objects by measuring the time between emitting ultrasound waves and receiving the reflected echoes. Based on this principle, researchers have developed sonar technology for underwater distance measurement and ultrasound devices for assisting clinical diagnosis.

When sound waves propagate through an object, they scatter (Maier et al., 2018) and experience attenuation (Powles et al., 2018). Small material inhomogeneities or rough surfaces can cause sound wave scattering. Internal

---

[2] https://wiki.cancerimagingarchive.net/display/Public/CT+Images+in+COVID-19

scattering within a material can affect sound wave imaging, especially in the presence of rough surfaces. As the wavelength decreases and the surface roughness increases, the angle of reflection also increases. When sound waves pass through obstacles, objects, or openings, diffraction occurs, causing a change in the direction of the waves (Maier et al., 2018). The extent of diffraction depends on the wavelength, with greater diffraction effects observed as the wavelength increases (Piechowicz, 2011). When the wavelength is smaller than the size of the obstacle, object, or opening, diffraction can be considered negligible (Piechowicz, 2011). Attenuation refers to the gradual reduction in sound wave intensity as it propagates through a medium. Attenuation follows an exponential decay law, and the attenuation coefficient, measured in decibels (dB), depends on tissue type and ultrasound frequency (Shriki, 2014). The attenuation coefficient comprises absorption and scattering components (Shriki, 2014).

Due to the presence of boundaries between different anatomical structures in the human body, sound waves undergo partial reflection and transmission at these boundaries. The intensity of sound wave reflection is quantified by the acoustic impedance of the medium. When two media have equal impedances, no reflection occurs (Maier et al., 2018). However, significant reflection occurs when there is a large impedance mismatch between two media. Therefore, medical ultrasound is not effective for examining organs containing air, such as the lungs, due to the low impedance of air. This causes ultrasound waves to be reflected, scattered, and attenuated when passing through air-containing media, resulting in decreased image quality (Patey and Corcoran, 2021).

In ultrasound imaging, there are different types of spatial resolution, particularly axial and lateral resolution (Alomari et al., 2015; Carovac et al., 2011). Axial resolution refers to the ability to differentiate structures located behind each other in ultrasound imaging. Higher axial resolution means the ability to distinguish smaller distances between two structures. Axial resolution is highly dependent on the ultrasound wave frequency. Lateral resolution involves the ability to differentiate adjacent structures within the same lateral distance. Lateral resolution is typically not as good as axial resolution. Additionally, ultrasound offers various imaging modes, including A-mode, B-mode, and M-mode (Carovac et al., 2011). A-mode and M-mode generate 1D images, while B-mode can be used to acquire 2D or even 3D images (Maier et al., 2018). To obtain 2D images inside the body, ultrasound devices sample different directions of 2D planes in 3D space.

Compared to other imaging techniques, medical ultrasound has the advantages of low cost and non-ionizing radiation (Protection, 2017; Shung, 2011). Other advantages include non-invasiveness, minimal discomfort, rapid image acquisition, and wide application in visualizing the motion and function of structures, organs, and blood vessels (Maier et al., 2018; Ortiz et al., 2012). Medical ultrasound is considered one of the safest imaging modalities available today and can even be used during pregnancy. However, ultrasound waves do have potential effects on the body, including mild heating or cavitation to some extent (Yoshizawa et al., 2017). Nonetheless, medical ultrasound has become the preferred modality for clinical imaging and finds increasing applications in various diagnostic and therapeutic scenarios.

However, like other imaging modalities, ultrasound imaging techniques also face technical challenges that affect the overall image quality. The wavelength of ultrasound waves and the inherent limitations of ultrasound transducers result in lower image resolution compared to CT and MRI (Hoskins et al., 2019). Factors such as tissue interfaces, beamforming errors, and scattering effects can cause image artifacts during the imaging process (Patey and Corcoran, 2021; Szabo, 2004). Additionally, as ultrasound waves pass through different tissues, they experience attenuation, leading to signal loss. Signal loss can result in decreased image quality and limited penetration depth. Structures or tissues located behind highly attenuating or reflecting objects can cause shadowing or acoustic blocking, leading to blurred or distorted regions in the ultrasound image (Hoskins et al., 2019). All of these make it challenging to accurately visualize certain anatomical structures.

Figure 6 displays an ultrasound image of a fetus (Ahmed, 2017). The full-body photograph of the baby in the image is very blurry. The organs and tissues inside the baby cannot be distinguished. Compared to the CT image mentioned earlier, the image quality of medical ultrasound is poorer. This may be because that ultrasound waves have longer wavelengths than X-rays used in CT imaging. The longer wavelength results in lower spatial resolution, making it challenging to differentiate small structures or details within the body. And sound waves are prone to scattering and attenuation as they pass through different tissues. When there is a large impedance mismatch between two tissues, significant reflection occurs. This can cause ultrasound waves to be reflected, scattered, and attenuated, leading to decreased image quality. CT imaging, on the other hand, is not affected by acoustic impedance mismatch and can produce higher quality images even in the presence of air or gas-filled structures.

FIGURE 6 A fetal ultrasound image

2.1.4 Magnetic resonance imaging (MRI)

Another commonly used medical imaging technique is MRI imaging. MRI systems utilize magnetic resonance to achieve non-ionizing radiation imaging of the human body's internal structures. Under the influence of a strong magnetic field, the spin axes tend to align along the direction of the magnetic field, forming a net magnetization M. Nuclear spins precess around the axis of the main magnetic field (Mlynárik, 2017). The precession frequency is determined by the Larmor frequency, which depends on the gyromagnetic ratio and the strength of the main magnetic field (Mlynárik, 2017). With a relatively weak radiofrequency (RF) field (Mlynárik, 2017), the direction of M can be manipulated, enabling excitation and emission of RF waves. After excitation, M gradually returns to its equilibrium position through relaxation processes while emitting signals.

The relaxation rate depends on the tissue, as the interactions of protons with hindered water molecule motion are limited in large molecules or dense tissues (Maier et al., 2018). This means that different tissues, such as water and fat, will reach the end of relaxation with different speed. This results in varying receiving signal intensities during relaxation. Relaxation forms the basis of contrast between different tissues in MRI images. By analyzing the recovery and decay processes of longitudinal and transverse magnetization, different tissue types can be recognized. For more details, see (Maier et al., 2018; Serai, 2022; Vuong et al., 2017).

MRI offers excellent soft tissue contrast and provides a variety of possibilities for functional imaging, such as visualizing blood flow (Markl et al., 2012), tissue perfusion (E. Lee et al., 2015), and diffusion processes (Le Bihan and Johansen-Berg, 2012). The physical basis of MRI involves nuclear magnetic resonance, where different contrast images are generated by adjusting the magnetic field and RF pulse sequences.

In addition, the contrast of MRI images can be controlled through a variety of parameters, including sequence types and settings for each sequence (Maier et al., 2018). When selecting the contrast or weighting of specific images, the key is to ensure that the contrast primarily reflects changes related to one of the inherent characteristics of these tissues. The three main types of contrast are called by the names T_1-weighted, T_2-weighted, and proton density (PD)-weighted imaging (Maier et al., 2018; Serai, 2022).

MRI offers unique advantages over CT, particularly in visualizing soft tissues and providing functional information. MRI is valuable in areas such as neurological imaging, musculoskeletal imaging, cardiac imaging, and evaluating soft tissue abnormalities. However, CT imaging utilizes X-rays, which have high attenuation in dense structures such as bones, resulting in strong contrast between different tissues. MRI relies on the magnetic properties and relaxation times of hydrogen protons in the body's tissues. The contrast in MRI images is primarily based on the differences in relaxation times between tissues. While MRI can distinguish between tissues with varying water content, it may not provide the same level of contrast as CT. Some factors, including the strength of the magnetic field, the size of the imaging voxel, and the imaging sequence can highly affect the MRI image quality. These factors may lead to severe blurriness in MRI images. Moreover, MRI is highly sensitive to motion artifacts, as even slight movements can cause blurring in the images. MRI scans typically take some time to acquire images. Prolonged scan times increase the likelihood of motion artifacts, particularly in patients who have difficulty remaining still (Katti et al., 2011).

Figure 7 displays an MRI image that includes the cervical and lumbar spine (Y. Zhou et al., 2017). From this figure, it can be observed that MRI imaging can effectively visualize the soft tissues and organs inside the human body. But it yields poor image quality for bones compared to CT imaging. This is due to the differences in relaxation times of tissues, specifically the longitudinal (T_1) and transverse (T_2) relaxation times. Soft tissues, such as muscles, organs, and fluids, have different relaxation times, allowing for good differentiation and contrast in MRI images. On the other hand, bones have a relatively short T_1 and T_2 relaxation time, resulting in poor contrast in MRI. The high mineral content of bones, mainly consisting of calcium hydroxyapatite, leads to short T_1 relaxation times, reducing the contrast between bone and surrounding tissues. Additionally, the fast signal decay of bone due to short T_2 relaxation time limits the ability of MRI to clearly distinguish bone structures. These factors ultimately lead to a decrease in the quality of bone imaging in MRI.

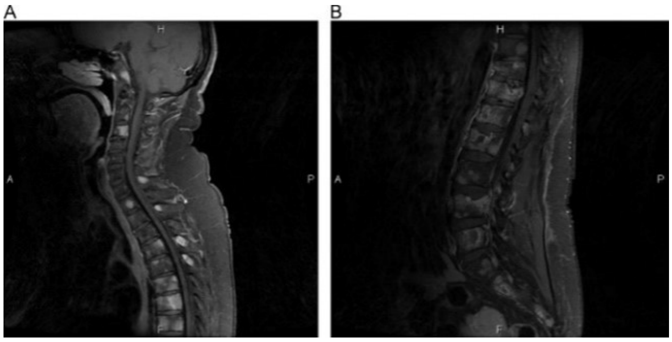


FIGURE 7 A cervical and lumbar spine MRI image

2.1.5 Motivation for choosing CT images in our research

To sum up, we use CT images as the subject of our research for the following several reasons:

- High contrast resolution: CT images provide high contrast resolution, allowing for clear differentiation between different tissues and structures (J. Hu et al., 2004; Pauwels et al., 2013). This makes it easier to identify and segment specific ROIs within the image.
- 3D imaging: CT scans provide volumetric data, allowing for 3D visualization of the scanned area (Salvolini et al., 2000). This is particularly beneficial for complex anatomical structures or regions where depth information is crucial for accurate segmentation.
- Soft tissue visualization: While MRI technology has its advantages, CT imaging is still well-suited for visualizing soft tissues, such as organs, muscles, and blood vessels (Pauwels et al., 2013). It provides detailed information about the internal structures, making it valuable for

segmentation tasks that involve differentiating and delineating these tissues.
- Wide availability: CT imaging is a widely used modality in medical imaging, and there is a substantial amount of publicly available CT image datasets. This availability makes it convenient for researchers to access and work with CT images for developing and evaluating segmentation algorithms.
- Clinical relevance: CT scans are commonly used in clinical practice for diagnosing and monitoring various medical conditions (Hussain et al., 2022). By using CT images for segmentation research, the findings and techniques developed can be more readily translated into practical applications and potentially improve patient care.
- Multimodal fusion: CT images can be combined with other imaging modalities, such as MRI or positron emission tomography (PET), to enhance the segmentation accuracy (Zaidi et al., 2008; Y.-D. Zhang et al., 2020). The fusion of multiple modalities can provide complementary information and improve the delineation of structures or abnormalities.

Overall, CT images offer valuable features for MIS, including high contrast resolution, 3D imaging capability, and the ability to visualize soft tissues. These characteristics make CT a popular choice for researchers working on segmentation algorithms and applications in the medical field.

2.2 Dataset used in the research

In Articles I-III, we use 98 mouse micro-CT images for constructing multi-resolution multi-organ models. These images are provided by the Molecular Imaging Centre of the University of California, Los Angeles (Stout et al., 2005; Suckow and Stout, 2008; H. Wang et al., 2011). The mice selected for the imaging experiment weigh between 15 and 30 grams. To obtain clear abdominal organs of mice, the university researchers inject the Fenestra LC contrast agent (ART, Montreal, QC, Canada) into the mice during the imaging process (Suckow and Stout, 2008; Willekens et al., 2009). Fenestra LC liver contrast agent is designed for micro-CT imaging studies in small rodent animals, particularly mice. It overcomes the limitation of poor soft tissue contrast in micro-CT imaging (Holdsworth and Thornton, 2002). Specifically, for liver imaging in small animals, Fenestra LC liver contrast agent exhibits superior imaging capabilities. It is an iodine-based lipid compound in which the iodine atoms efficiently absorb X-rays, resulting in higher attenuation and producing enhanced soft tissue contrast (S. Herman, 2004). In micro-CT images without the use of contrast enhancement, the contrast of soft tissues is significantly poor (H. Wang et al., 2019). Thus, researchers are unable to visually discern the boundaries between different tissues and organs. As shown in Figure 8, the left image represents a contrast-enhanced micro-CT image, while the right image represents a non-enhanced micro-CT image (H. Wang et al., 2019). It is evident from the images that the non-enhanced tissues appear blurry, while the contrast-enhanced image greatly improves discernibility. The frequent use of contrast agents to enhance soft tissue contrast adds complexity and higher costs to experimental procedures (Yan et al., 2017). Consequently, to balance radiation dose and CT imaging quality simultaneously, most image analysis studies utilize low-contrast CT images as research data.

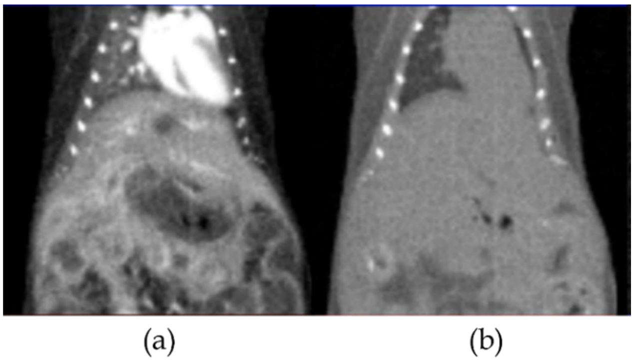


FIGURE 8 An exemplar comparison of contrast-enhanced (a) and low-contrast (b) mouse images

In the process of selecting qualified mouse micro-CT images, it is necessary to eliminate some mouse images that contain motion artifacts, blurred edges of organs, and body deviations from normal postures. Wang et al. describe the imaging system, imaging equipment parameters, and imaging parameters of mice in detail in their study (H. Wang et al., 2019). Although the process of imaging is strictly standardized, when modeling mouse abdominal images, the collected image data should also be preprocessed.

After collecting mouse micro-CT image data, we then preprocess the data. In the first step, we invite small animal imaging experts to manually segment the livers, spleens, left kidneys and right kidneys in mouse micro-CT images using MITK (Wolf et al., 2005) software. At the same time, imaging experts label the segmented organs. These labeled segmentation results serve as the gold standards for experimental analysis. In the second step, we use the moving cube algorithm (Lorensen and Cline, 1987) to convert all gold standards into 3D surfaces represented by points. The third step is to randomly select one of the 98 3D surfaces as a reference template. The template is then registered to the other 97 surfaces by using a point cloud registration method (Marani et al., 2016; Park and Lim, 2014). The purpose of this step is to make all 3D surfaces contain the same number of points, and the same point on different surfaces corresponds to the same anatomical location. Through the above three steps, we finally get the training data for model building. Figure 9 shows the preprocessing process of the training data.

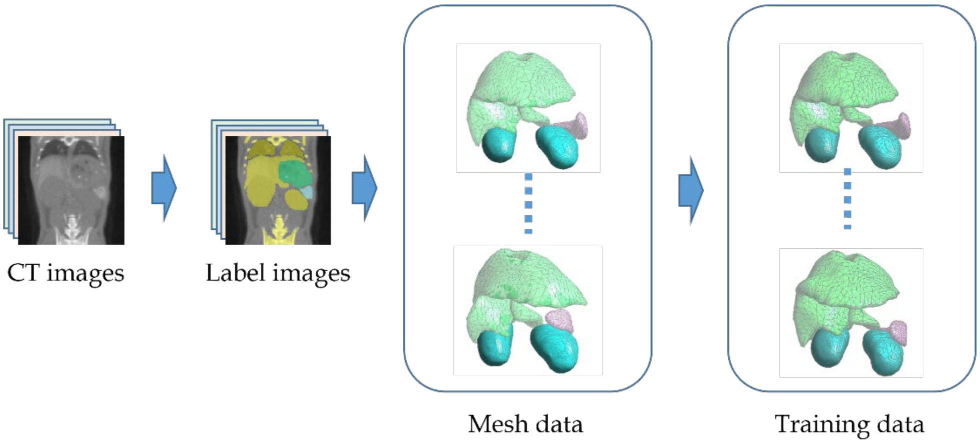


FIGURE 9 Preprocessing process of training data

By integrating the model developed in Article I with the VHRBF function (Brazil et al., 2010), Article III successfully achieves multi-organ segmentation of the abdomen in low-contrast mouse micro-CT images. The raw mouse micro-CT images used in the study are produced by injecting contrast agents. Therefore, in Article III, we add some random noise and adjust the image contrast manually. Figure 10 displays examples of the images generated, ranging from high-contrast raw data to low-contrast experimental data required for segmentation. These experimental data are used to conduct the segmentation experiment.

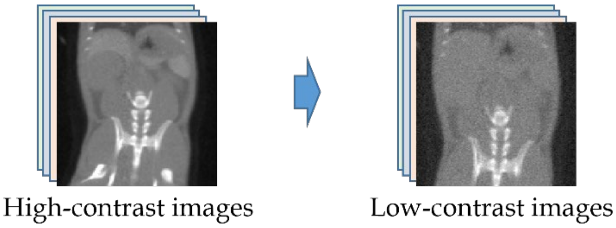


FIGURE 10 Examples of high-contrast images and low-contrast images

2.3 Medical image modeling based on PCA

In this section, we use the improved PCA method to solve the first problem raised in Section 1.2. PCA is a statistical method for data analysis. It uses mathematics to transform various correlated variables into a few linearly independent variables. The transformed variables are called principal components that can represent all original variables. In the process of using PCA, the information contained in the experimental data will be projected into the orthogonal subspace.The subspace is generated based on the data set for dimensionality reduction and storage. Therefore, in digital image processing, the main purpose of PCA is to reduce the large dimension of image information to the small dimension of the independent variables. Shlens point out that PCA is an essentially mathematical process of extracting interesting information from massive data (Shlens, 2014). Later, PCA has been developed into one of the most successful data compression and storage methods.

In the process of extracting principal components, PCA needs to perform singular value decomposition (SVD) on the data matrix or perform eigenvalue decomposition (EVD) on the covariance matrix of the data. The extracted principal components can represent the original data information almost without loss. The storage space required by these principal components is greatly reduced. Therefore, PCA plays a very important role in removing data redundancy, extracting main features, and compressing data. At the same time, each new variable obtained from PCA transformation is a linear combination of all original variables. So, PCA is a method of linear dimensionality reduction or modeling. In MIS research, PCA is widely used in conventional image segmentation methods. Its model construction for various medical images corresponds to the image segmentation methods based on deformable models introduced in Section 1.1.2.

For MIS, conventional image segmentation methods based on deformable models usually need to model the target regions first. For example, in the modeling process, 2D or 3D deformable models are constructed using statistics-based techniques. Then the models are registered to medical images for ROI segmentation (Cootes et al., 1994; Hill et al., 1993). During model building, a preprocessed training set of medical images is analyzed using PCA to obtain a mean shape model with principal deformation modes. The mean shape model and the principal deformation modes are linearly summed to generate new shapes. In addition, Tsai et al. (Tsai et al., 2003) use PCA to extract the signal representation of the training data to construct a segmented curve parametric model of implicit representation. This model can be used to segment organs in MRI images. But distance functions based on PCA to represent shapes have limitations. It can cause problems with inconsistent shape modeling frameworks.

As mentioned in Section 1.1.2, deformable models belong to conventional segmentation methods. Before image segmentation, specific organs or tissues need to be modeled. But most of these models are unable to model multiple organs simultaneously. Furthermore, most models do not adequately simulate shape changes. And during the model building process, it is possible to encounter the high-dimensional-low-sample-size (HDLSS) problem (Aoshima et al., 2018; Hall et al., 2005).

Therefore, to solve the HDLSS problem, we use a down-sampling-and-interpolation strategy to down sample the points of the training data and interpolate the deformation components (principal components) obtained by PCA respectively. To build a multi-organ deformable model that can deform from global to local, we extend the SSM modeling method of Wilms et al. (Wilms et al., 2017) from 2D to 3D modeling in mouse micro-CT images.

The steps in Article I to address the above issues include:

Step 1 Down sample the preprocessed N 3D image data given in Section 2.2 to obtain training dataset. Each training dataset consists of M points.

Step 2 Calculate the mean value $\vec{\boldsymbol{\mu}}$ of the training dataset $\{\boldsymbol{S}_i\}_{i=1}^{N}$:

$$\vec{\boldsymbol{\mu}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{S}_i \tag{1}$$

Step 3 Subtract the mean value from each training data to get the difference matrix, then compute a covariance as follows:

$$C = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{S}_i - \vec{\boldsymbol{\mu}})(\boldsymbol{S}_i - \vec{\boldsymbol{\mu}})^T \tag{2}$$

Step 4 Compute the point-to-point Euclidean distance $d_{(X_i,X_j)}$on the mean value model. The distances are represented by a distance matrix D. Then, select an appropriate distance threshold τ. The values in matrix D less than or equal to the threshold are converted to 1, and the values greater than the threshold are converted to 0. Finally, the mask matrix R is obtained:

$$D = \begin{bmatrix} d_{(X_1,X_1)} & \cdots & d_{(X_1,X_M)} \\ \vdots & \ddots & \vdots \\ d_{(X_M,X_1)} & \cdots & d_{(X_M,X_M)} \end{bmatrix} \tag{3}$$

$$\rho_{i,j} = \begin{cases} 1, & d_{(X_i,X_j)} \leq \tau \\ 0, & d_{(X_i,X_j)} > \tau \end{cases} \tag{4}$$

$$R = \begin{bmatrix} \rho_{1,1} & \cdots & \rho_{1,3M} \\ \vdots & \ddots & \vdots \\ \rho_{3M,1} & \cdots & \rho_{3M,3M} \end{bmatrix} \tag{5}$$

Step 5 The values in the covariance matrix in step 3 are dot multiplied by the values in the corresponding positions in the mask matrix. The improved covariance matrix C' can be obtained:

$$C' = C \cdot R \tag{6}$$

Step 6 Perform eigenvalue decomposition on the covariance matrix C' to obtain eigenvalues and eigenvectors. Arrange the eigenvalues from large to small, and their corresponding eigenvectors are also rearranged, where eigenvalues are $[\lambda_1, \lambda_2, \ldots, \lambda_{3M}]$, and eigenvectors are $[V_1, V_2, \ldots, V_{3M}]$.

Step 7 The vector $\boldsymbol{S}_i$ representing one training data can be formed by a linear combination of eigenvectors and the mean value model:

$$\boldsymbol{S}_i = \vec{\boldsymbol{\mu}} + \alpha_1 V_1 + \alpha_2 V_2 + \cdots + \alpha_{3M} V_{3M} \tag{7}$$

Step 8 Commonly, the eigenvectors corresponding to the first few eigenvalues can represent the training models with different deformable modes to the greatest extent:

$$S = \vec{\boldsymbol{\mu}} + \alpha_1 V_1 + \alpha_2 V_2 + \cdots + \alpha_l V_l, \quad l < 3M \tag{8}$$

Step 9 Finally, we use Laplace iteration diffusion algorithm to obtain all deformation vectors (eigenvectors) on the image data before downsampling:

$$\vec{x}_{i_{(n+1)}} = \vec{x}_{i_n} + \frac{\theta}{K}\sum\nolimits_{j=0}^{K}\left(\vec{x}_{j_n} - \vec{x}_{i_n}\right) \tag{9}$$

Here n indicates the number of iterations, i indicates the vertex index, $\vec{x}_{i_n}$ indicates a deformation vector of i-th vertex at n-th iteration, $j = [0, 1, \ldots, K]$ indicates $K + 1$ indices of neighbor vertices around vertex i, and θ is the smooth intensity coefficient.

2.4 Medical image modeling based on SAE

In Section 2.3 of this chapter, the improved PCA method is able to extract the main features of the images and compress medical images. A small number of linear combinations of main features (main eigenvectors) can construct a variety of deformable models with different deformation modes. Therefore, the improved method based on PCA is a linear modeling method. So besides building a linear model, is there any other way to mine the nonlinear deformation modes of a deformable model? In view of this, we introduce a SAE neural network in the research to further explore the nonlinear deformation capability of the multi-organ deformable model. we built a SAE neural network to solve the second problem raised in Section 1.2.

As mentioned in Section 1.1.2, deformable model-based methods are widely used in medical image modeling and segmentation. Most PCA-based methods, as one of the deformable model-based methods, are used for specific organ modeling and segmentation. The improved PCA method in Section 2.3 can construct a multi-organ deformable model with deformations of various range levels. Compared with the globally deformable model constructed by traditional PCA, this method has more linear deformation modes. But these deformation modes are limited to the linear combination of different deformation components. During the PCA modeling process, if the input training data has outliers, the deformation performance of the constructed model will drop dramatically. Therefore, most PCA methods are sensitive to outliers. In addition, if the training data is too large and the computer memory usage increases exponentially, PCA modeling may fail. The deformation components of the model cannot be obtained (Heng Huang et al., 2008; Wilms et al., 2017). An autoencoder network has a simple structure, which means researchers can adjust hyperparameters flexibly (Andonie, 2019; Bischl et al., 2021; Feurer and Hutter, 2019; L. Yang and Shami, 2020). Moreover, it can learn different types of nonlinear deformation modes from training samples. For example, Litany et al. use a variational autoencoder to learn the latent deformation space of non-rigid deformable objects(Litany et al., 2018). This can predict and reconstruct occluded and missing parts of human bodies and faces (Litany et al., 2018). At the same time, Ranjan et al. use a convolutional grid autoencoder to capture the nonlinear changes of the face to reconstruct a complete face model (Ranjan et al., 2018). The drawback of this approach, however, is that the meshes that make up the model need to share the same topology.

To explore more nonlinear deformation modes, we stack autoencoders to obtain the SAE neural network. Autoencoders are a special type of unsupervised neural network (Baldi and Hornik, 1989; Rumelhart et al., 1986). Autoencoders play a huge role in automatically acquiring nonlinear features (M. Chen et al., 2012). Its main purpose is to compress and encode the input data into meaningful features, and then decode them so that the reconstructed output is as similar as possible to the original input (Bank et al., 2020). As shown in Figure 11, a basic autoencoder consists of three parts which are input layer, hidden layer and output layer (G. Zhang et al., 2020). Assume that the number of neurons in the input layer and the hidden layer are m, n, ($m > n$), respectively. It can be seen from the symmetrical structure of the autoencoder that the number of neurons in the output layer is also m. The encoder consists of an input layer and a hidden layer. The decoder consists of a hidden layer and an output layer. The encoder uses a function f_1 to transform the high-dimensional input data $x = \{x_1, x_2, \dots, x_m\}^T$ into a low-dimensional hidden representation $h = \{h_1, h_2, \dots, h_n\}$. The function f_1 is defined as:

$$f_1(x) = h = \sigma(Wx + b) \tag{10}$$

where σ represents the encoder's activation function. W is a $n \times m$ weight matrix, and $b \in R^n$ is a bias vector of the hidden layer. The weight matrix W and the bias vector b parameterize the decoder.

In the decoding stage, the decoder maps the hidden representation h back to a reconstruction $x' = \{x_1', x_2', \dots, x_m'\}^T$ through a function f_2. The function f_2 can be written as:

$$f_2(x) = x' = \sigma'(W'h + b') \tag{11}$$

where σ' is the activation function of the decoder. W' is a $m \times n$ weight matrix, and $b' \in R^m$ is a bias vector for the output layer. Likewise, the weight matrix W' and the bias vector b' parameterize the decoder.

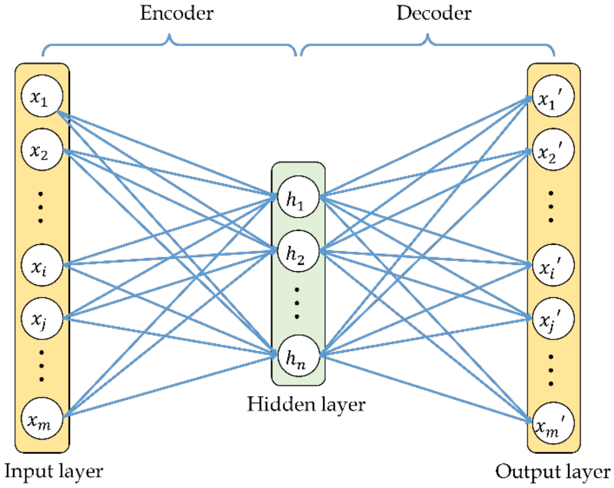


FIGURE 11 The structure of a basic autoencoder

In fact, autoencoders have close relationship with PCA-based methods (Ghojogh et al., 2021; Lempitsky, 2019). If an activation function used on each layer of an autoencoder is linear, the final encoding vector obtained by the encoder will correspond to the principal components created by PCA. Therefore, to obtain more useful potential nonlinear features (Hinton and Salakhutdinov, 2006; Japkowicz et al., 2000), the activation functions σ and σ' are usually nonlinear, such as, typical rectified linear function (ReLU), Leaky ReLU, hyperbolic tangent function, and sigmoid function (Z. Yang et al., 2022; S. Zhang et al., 2017). Compared with PCA, nonlinear activation functions can help autoencoders reconstruct better results. They can also build dimensionality reduction models for more complex data. Interested readers can read the literatures (Hinton and Salakhutdinov, 2006; W. Wang et al., 2014; Y. Wang et al., 2016) to learn more about the application of autoencoders in dimensionality reduction.

After the construction of an autoencoder, the training data needs to be input into the autoencoder for training. During training, the network hyperparameters are constantly adjusted such that the autoencoder minimizes the error between x and x'. Two methods of square error and cross-entropy are usually used to compute the reconstruction error (G. Zhang et al., 2020). Their formulas are defined as follows:

Square error:

$$E(x, x') = \|x - x'\|^2 \tag{12}$$

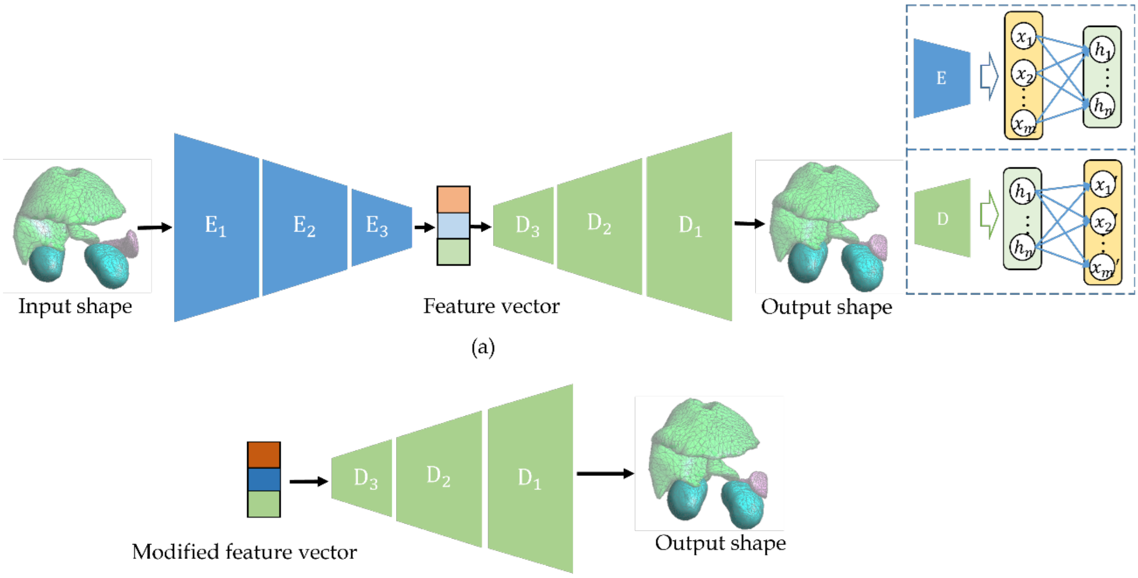
Cross-entropy:

$$E(x, x') = -\sum_{i}^{m} (x_i \log x_i' + (1 - x_i) \log(1 - x_i')) \tag{13}$$

Then add a regularized term RT and the adjustment factor λ to the error function to construct the loss function of the autoencoder. Autoencoders typically optimize the loss function using alternative least squares (ALS) (Takane et al., 1977) or stochastic gradient descent (SGD) (Bertsekas and Tsitsiklis, 2000). The loss function is defined as:

$$Loss(x, x') = \left(\sum_{x \epsilon R^m} E(x, x')\right) + \lambda \cdot RT \tag{14}$$

The research in Article II concatenates the encoding parts and decoding parts of multiple autoencoders respectively. The purpose of this article is to maximize the compression of the input data while retaining the input data information. After continuous testing and adjustment, we finally build a three-level SAE. As shown in Figure 12, a three-level SAE consists of three encoding layers and three decoding layers. In the figure, E represents an encoder and D represents a decoder. The corresponding D and E form an autoencoder. The training data is compressed in the three-layer encoding network to obtain a matrix $M_F = \{\overrightarrow{f_1}, \overrightarrow{f_2}, \dots \overrightarrow{f_i}, \dots, \overrightarrow{f_n}\}$ containing n feature vectors, and M_F can reconstruct the original input data through the three-layer decoding network. During the decoding process of feature vectors, modification for the values of feature vectors can change the reconstructed data. Changes in the reconstructed data are reflected in the deformation of 3D models. Therefore, we can obtain different multi-organ deformable models by changing feature vectors.



(a)

(b)

FIGURE 12 The architecture of our proposed three-level SAE

2.5 Multi-organ segmentation based on shape prior knowledge

The multi-resolution multi-organ SSM constructed in Section 2.3 can be used as a shape prior knowledge for MIS. In this section, we mainly introduce how Article III combines shape prior knowledge and a curve interpolation method to achieve multiple organ segmentation in mouse images. Article III attempts to address the third problem raised in Section 1.2.

Practically, mice are one of the most used animals for imaging. Realizing automatic segmentation of mouse CT images provides important reference value for computer-aided clinical analysis and diagnosis. Moreover, the accurate segmentation of mouse organs is of great significance for the study of volume reconstruction (Pitiot et al., 2006), pharmacokinetics (Corot et al., 2006), organ anatomy (H. Wang et al., 2011), and lesion detection (Rudyanto et al., 2013). In fact, the CT imaging process can be affected by noise, spatial resolution, contrast, etc. (Zarb et al., 2010). These factors eventually lead to blurred mouse images. As illustrated in Figure 13, a typical mouse CT image usually cannot clearly display the boundaries of various organs. The boundaries between the lower part of the liver, the upper part of the spleen, and the upper parts of the kidneys cannot even be accurately identified by human eyes. One way to obtain high contrast images is to use contrast agents. However, the use of contrast agents will greatly increase the cost and complexity of mouse imaging (Lee et al., 2013), thus affecting the experimental results. Some contrast agents can even be toxic to animals (Hasebroock and Serkova, 2009; Rogosnitzky and Branch, 2016). Therefore, to image large numbers of mice, contrast agents are usually not used. These factors pose great challenges to the segmentation of tissues and organs in images.

FIGURE 13 A typical blurred mouse abdominal micro-CT image

On the other hand, low-contrast images may cause the failure of segmentation for fully automatic segmentation methods. To obtain multiple organs, it is a feasible approach by incorporating human interaction and shape prior knowledge into the segmentation process. In Article III, we use the SSM constructed in Article I as shape prior knowledge for image segmentation. SSMs are a powerful tool for MIS. The early and well-known version of this method was proposed by Cootes et al., including ASMs (Cootes et al., 1995) and AAMs (Cootes et al., 2001). Over the past three decades, many variants have been developed on this basis. In many research, the point distribution model (PDM) is used to represent different model shapes (Cootes and Taylor, 1995; Hill and Taylor, 1994). Based on this, Stegmann et al. introduced how to use traditional PCA to construct 2D SSM in detail in their research (Stegmann and Gomez, 2002). Heimann et al. points out that in general, more sufficient shape information in the training data helps to build a more sophisticated SSM (Heimann and Meinzer, 2009). The results obtained by segmentation are also more accurate. For more detailed segmentation methods based on SSMs, readers can refer to (Heimann and Meinzer, 2009).

The VHRBF function used in Article III guides the deformation of SSM to achieve efficient segmentation results. The steps to obtain multiple organs in the mouse abdomen include:

Step 1 Delineate some discrete curves (landmarks) around the rough boundaries of target organs in low-contrast images. The number of landmarks $\{\boldsymbol{X}^{\boldsymbol{1}}, \boldsymbol{X}^{\boldsymbol{2}}, \dots, \boldsymbol{X}^{\boldsymbol{N}}\} \epsilon \mathbb{R}^3$ is N. Other points forming a closed curve with the landmarks are calculated by the interpolation function f:

$$f(\boldsymbol{X}) = \sum_{j=1}^{N}\left\{\alpha_j\left\|\boldsymbol{X}-\boldsymbol{X^j}\right\|^3 - 3\langle\boldsymbol{\beta^j},\boldsymbol{X}-\boldsymbol{X^j}\rangle\left\|\boldsymbol{X}-\boldsymbol{X^j}\right\|\right\} + \langle\boldsymbol{a},\boldsymbol{X}\rangle + b \tag{15}$$

Where $\boldsymbol{X} \epsilon \mathbb{R}^3$ is a point in a 3D space. $[\alpha_1, \alpha_2, \dots, \alpha_j, \dots, \alpha_N, b] \epsilon \mathbb{R}$ is a set of 1D interpolation coefficients. $[\boldsymbol{\beta^1}, \boldsymbol{\beta^2}, \dots, \boldsymbol{\beta^j}, \dots, \boldsymbol{\beta^N}, \boldsymbol{a}] \epsilon \mathbb{R}^3$ is a set of 3D interpolation coefficients.

Step 2 Give the normals $\{\boldsymbol{n^1}, \boldsymbol{n^2}, \dots \boldsymbol{n^j}, \dots, \boldsymbol{n^N}\} \epsilon \mathbb{R}^3$ of all landmarks. Then solve the coefficients as follows to get the interpolation function f:

$$\sum_{j=1}^{N}\left\{\alpha_j\left\|\boldsymbol{X}-\boldsymbol{X^j}\right\|^3 - 3\langle\boldsymbol{\beta^j},\boldsymbol{X}-\boldsymbol{X^j}\rangle\left\|\boldsymbol{X}-\boldsymbol{X^j}\right\|\right\} + \langle\boldsymbol{a},\boldsymbol{X}\rangle + b = 0 \tag{16}$$

$$\nabla f(\boldsymbol{X^j}) = \boldsymbol{n^j} \tag{17}$$

$$\sum_{j=1}^{N}\alpha_j = 0 \tag{18}$$

$$\sum_{j=1}^{N}\{\alpha_j\boldsymbol{X^j} + \boldsymbol{\beta^j}\} = 0 \tag{19}$$

Step 3 Scale and translate multi-organ deformable models. The center of the transformed deformable model coincides with the center of the target organs. The SSM transformation can be written as:

$$x_c = \frac{x_a - x_{a_0}}{l_{x_a}} l_{x_b} - x_{b_0} \tag{20}$$

$$y_c = \frac{y_a - y_{a_0}}{l_{y_a}} l_{y_b} - y_{b_0} \tag{21}$$

$$z_c = \frac{z_a - z_{a_0}}{l_{z_a}} l_{z_b} - z_{b_0} \tag{22}$$

where $\boldsymbol{X_a} = (x_a, y_a, z_a)$ represents any point on SSM before transformation. $\boldsymbol{X_c} = (x_c, y_c, z_c)$ represents any point on SSM after transformation. $\boldsymbol{X_{a_0}} = (x_{a_0}, y_{a_0}, z_{a_0})$ is the center of SSM before transformation. $\boldsymbol{X_{b_0}} = (x_{b_0}, y_{b_0}, z_{b_0})$ is the center of target organs. l_{x_a}, l_{y_a}, l_{z_a} represent lengths of SSM before transformation in x, y, z directions respectively. l_{x_b}, l_{y_b}, l_{z_b} represent lengths of target organs in x, y, z directions respectively.

Step 4 Input $\boldsymbol{X_c}$ into the interpolation function f for obtaining function's values $f(\boldsymbol{X_c})$. The calculation process uses a multi-thread mechanism for acceleration.

Step 5 At the same time, Input $\boldsymbol{X_c}$ into gradient function $\nabla f(\boldsymbol{X_c})$ to obtain the gradient of $\boldsymbol{X_c}$. The multi-thread mechanism is used to speed up the calculation process as well.

Step 6 Equation (16) shows that $f(\boldsymbol{X}) = 0$ corresponds to the delineated landmarks on the target organs. During the interpolation process, $f(\boldsymbol{X_c}) > 0$ means that the boundary of the SSM exceeds the delineated curves (landmarks). The SSM needs to deform inwards close to the boundary of target organs. $f(\boldsymbol{X_c}) < 0$ means that the boundary of the

SSM is lower than the delineated curves. SSM needs to deform outward close to the boundary of target organs. The gradient of the $\boldsymbol{X_c}$ represents the direction and speed of each movement of the SSM. The iterative deformation process of SSM can be written as:

$$\boldsymbol{X} = \varphi(\boldsymbol{X_1} + \boldsymbol{G}) + (1 - \varphi)\boldsymbol{X_2} \tag{23}$$

where $\boldsymbol{X_1}$ represents any point on the SSM before the next iteration. $\boldsymbol{G}$ represents any gradient corresponding to the point. $\boldsymbol{X_2}$ represents any point on the SSM after fitting the SSM to points $[\boldsymbol{X_1} + \boldsymbol{G}]$**.** This process represents one complete deformation of SSM. $\boldsymbol{X}$ represents any point on the SSM after one deformation. φ is a weight factor. Empirically, it is set as 0.8.

Combining steps 4, 5, and 6, the SSM is constantly moving and deforming. The SSM adjusts to the delineated curves and tries to get as close to the curves as possible. Iterate steps 4, 5, and 6 to compute the values of $f(\boldsymbol{X})$ and the gradients of points $[\boldsymbol{X}]$ on the SSM. These values modify the move direction and rate of SSM. When the value of f is 0 or close to 0, the iteration stops. Finally, the SSM can segment the target organs in a mouse micro-CT image.

Step 7 If the results of the SSM segmentation are not very accurate, human interaction correction is required. The operator again delineates the boundaries of the target organs to obtain new landmarks. Repeat steps 1 to 6 until the SSM segments more accurate organs.

After various tests for these steps, the method provided in Article III saves a lot of time for the segmentation of multiple organs. Figure 14 shows the whole process of VHRBF-guided multi-resolution multi-organ SSM to segment multiple organs in a mouse abdomen.

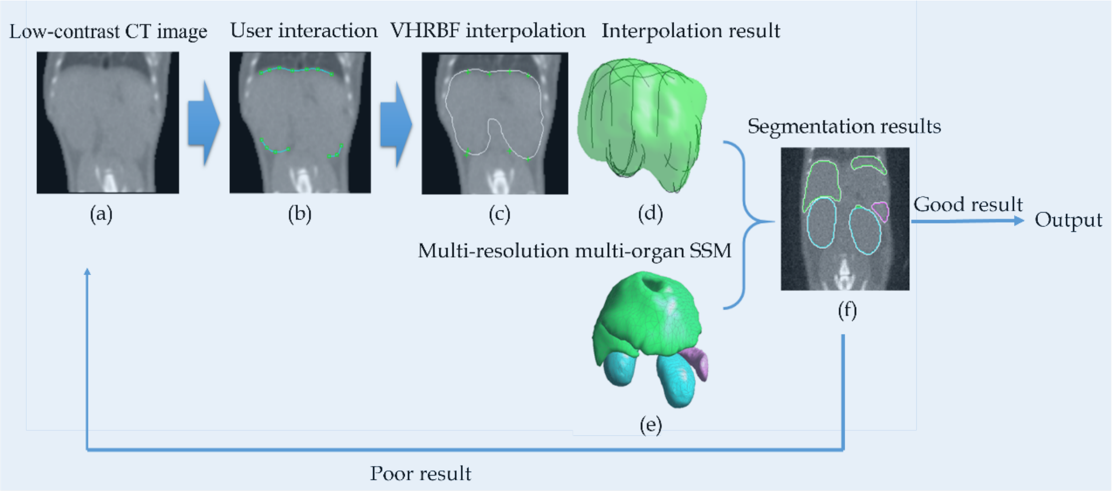


FIGURE 14 VHRBF-guided SSM segmentation process for multiple organs in a mouse micro-CT image

2.6 An extensible open-source software development for medical image analysis

Sections 2.3 and 2.4 focus on medical image modeling, while Section 2.5 centers around MIS. These techniques are all integral components of the complex process of developing algorithms for MIA. In this section, we primarily discuss the approach taken by Article IV to develop software that can be extended to support general MIA. The objective of Article IV is to address the fourth problem raised in Section 1.2.

Computer-aided MIA is a critical component of clinical diagnosis and treatment. The development of MIA algorithms is a complex process. It typically involves algorithm design, model training, software development, and performance testing. As a result, a suitable software platform is required to assist developers with each step of algorithm development. An effective development platform should feature a user-friendly graphical interface and a plug-in interface for algorithm integration. Over the past few decades, toolboxes and software programs have been developed to address these specific needs. Classic MIA toolboxes include itk (McCormick et al., 2014; Yoo et al., 2002), vtk (Schroeder et al., 2006), elastix (Klein et al., 2009; Shamonin et al., 2014), ANTS (Avants et al., 2009), the library pyraiaomics (Van Griethuysen et al., 2017) for radiomics research, and the DL framework monai[3]. Mainstream development software includes ITK-SNAP (Yushkevich et al., 2016), MITK (Wolf et al., 2005), Tur-gleSeg (Poon et al., 2007), Seg3D for image segmentation, DicomAnnotator (Dong et al., 2020) for data annotation, and SpheroidJ for analyzing imaging patterns (Lacalle et al., 2021), as well as MNI SISCOM (Moreau et al., 2021) and OIPAV (L. Zhang et al., 2019). However, these tools can be challenging to use for extending user-developed algorithms. In contrast, the popular 3D Slicer (Pieper et al., 2004) features a rich library of extension modules (Fedorov et al., 2012; Pinter et al., 2015), providing strong extension capabilities. Nevertheless, its interaction process and programming mechanism may not be user-friendly.

Conventional MIA methods have limitations, prompting the use of DL algorithms to aid in computer-aided diagnosis. However, training these models requires large amounts of annotated data. This is time-consuming and labor-intensive for human experts. Annotation-by-Iterative-Deep-Learning (AID) can alleviate these challenges. The AID process involves the following steps:

- Annotating a small amount of training data to train the network,
- Having the trained network automatically annotate more data,
- Experts proofreading the annotated data and adding it to the training set to retrain the network.

By continuously performing these AID steps, the network's annotating ability improves, and the required human effort is gradually reduced. However, implementing the AID process requires software platform support with

---

[3] https://monai.io/

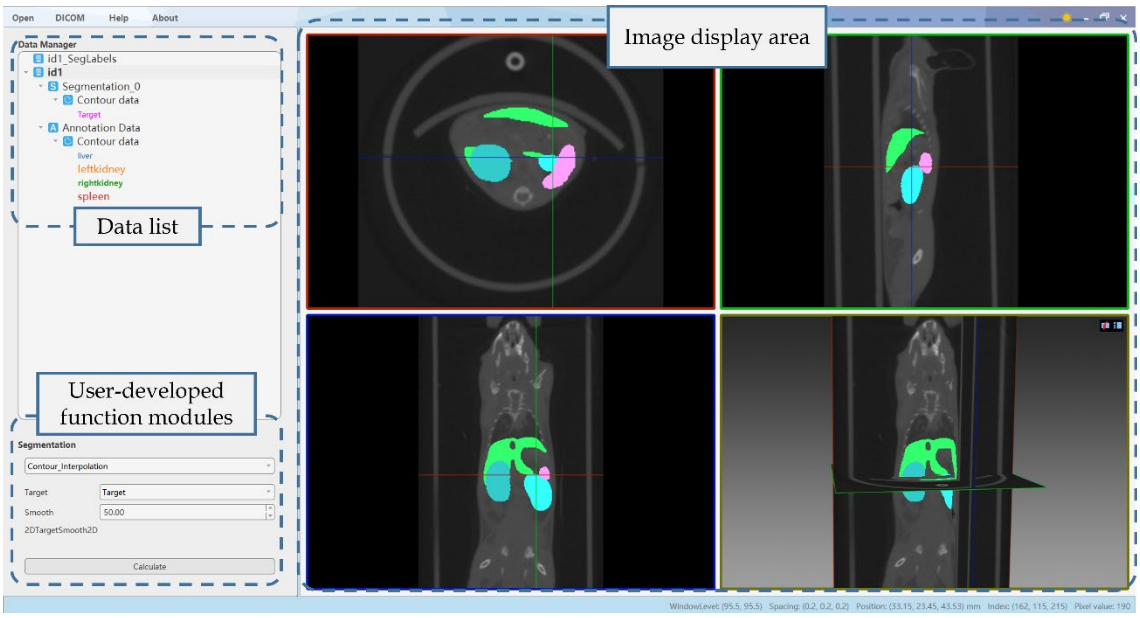
annotation functions and plug-in interfaces. The existing software with AID annotation function, RIL-Contour (Philbrick et al., 2019), cannot expand the plug-in interface. Thus, it cannot be a general auxiliary tool for MIA algorithm development. To facilitate DL research, the network model and AID process need to be integrated into one software. However, currently, no software exists with this feature. This forces researchers to switch between different auxiliary tools, complicating MIA algorithm development. The lack of a GUI means that some algorithms can only be released in source code. This makes it challenging for ordinary users to understand. All these factors contribute to the complexity of MIA algorithm development.

We have developed AS to address the challenges mentioned earlier. AS allows for data annotation, algorithm integration, and image visualization. It has a simple DL module interface and supports multi-touch and stylus inputs for image annotation. Its primary functions are a user-friendly GUI for data visualization and human-computer interaction, and a flexible plug-in interface for expanding user-developed algorithms.

In this study, the development of the AS software is comprised of four modules: software design, interactive annotation and proofreading, extension modules, and DL support.

Software design

The AS GUI consists of a menu bar, data management list, user development module panel, and image display area (see Figure 15). It can be controlled via mouse/keyboard in desktop mode or stylus/touch screen in tablet mode.



Target: Target
Smooth: 50.00

WindowLevel: (95.5, 95.5) Spacing: (0.2, 0.2, 0.2) Position: (33.15, 23.45, 43.53) mm Index: (162, 115, 215) Pixel value: 190

FIGURE 15 AnatomySketch interface

In desktop mode, the layout is like MITK, ITK-SNAP, and 3D Slicer, but with the ability to select algorithm modules from different developers in the panel. This makes AS highly extensible. It can also be integrated into user programs for data visualization and analysis of user interaction and data flow. The GUI allows overlaying of variables for image registration/segmentation accuracy verification. Tablet mode adds multi-touch screen and stylus operations for annotating ROIs. Figure 16 shows the three-layer AS software architecture (interaction, function, and data). The function layer coordinates other modules for data flow and user interaction, with blue arrows indicating data flow and red dashed arrows indicating command flow.

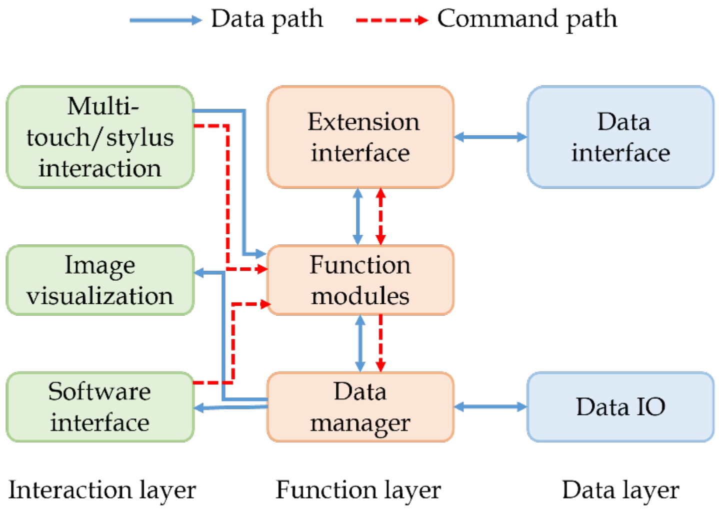


FIGURE 16 The basic architecture of AS

Interactive annotation and proofreading

AS allows users to annotate ROIs in images using a mouse or stylus. These annotations can be saved as separate files or utilized as input data for user-developed plug-in modules. Furthermore, AS includes a boundary correction tool constructed with free-form deformation (FFD) (Sederberg and Parry, 1986), which can be used to proofread AS segmentation results. With this tool, users can move the contour line of the 2D or 3D target area to refine the accuracy of segmented results. Figure 17 demonstrates the process, where the FFD method creates a 6x6 grid around the starting point of the contour line. The cubic B-spline inverse interpolation function (Tayebi et al., 2022) is then utilized to transfer the movement vector to the 6x6 grid vertices. Finally, the cubic B-spline interpolation function is applied to interpolate the deformation vector of the contour points, with interpolation defined as follows:

$$P = \sum_{i=0}^{3} \binom{3}{i} (1-m)^{3-i} m^{i} \left(\sum_{j=0}^{3} \binom{3}{j} (1-n)^{3-j} n^{j} p_{i,j} \right) \tag{24}$$

where the interpolated deformation vector of a contour vertex, P, is defined as a function of the normalized local coordinate of the dragging start point, (m, n), and the deformation vector of grid node (i, j), $p_{i,j}$. For more details on this method, please refer to (Sederberg and Parry, 1986).

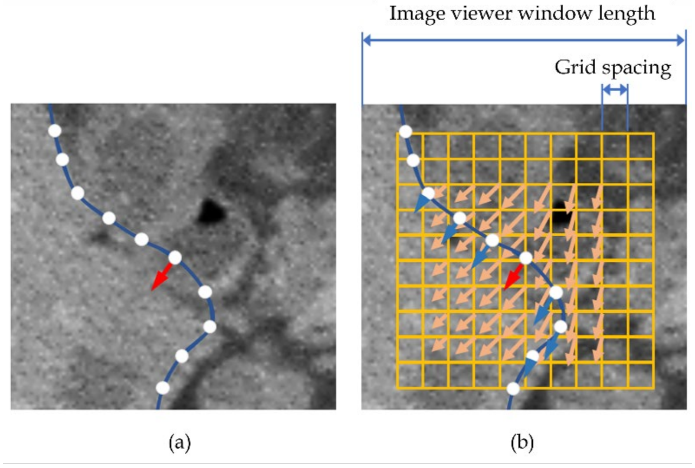


FIGURE 17 Interactive annotation and a demonstration of proofreading principles

Extension module

AS has a flexible plug-in interface, as shown in Figure 18. The interface comprises configuration and program files. Configuration files define the extension module's type, input and output parameters, and GUI settings in a text format file. AS uses the configuration file to implement the extended interface. Program files are executable files or dynamic link libraries that serve as the core component of the algorithm, written by users. The configuration file generates the module interface, which calls the program file to execute the core algorithm. Finally, the processed results is displayed in AS image display area.

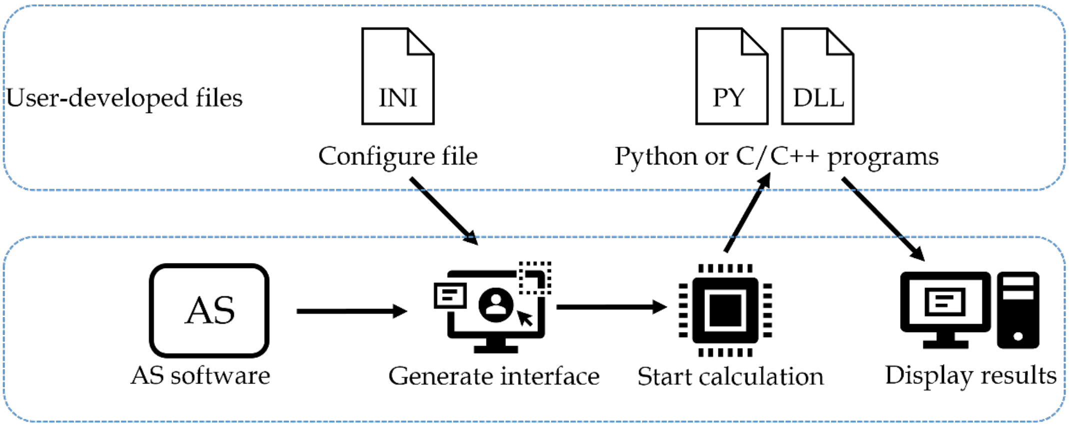


FIGURE 18 The workflow of the extension module

DL support

AS can integrate DL networks, which can be developed using Pytorch, Tensorflow, or Keras. These networks can be compiled into executable files. Users can use their own Python program to train the annotation data in AS by calling these executable files. When AS tackles DL problems, it employs the core idea of using annotation, proofreading, and plug-in interface to enhance the algorithm development efficiency.

3 EVALUATION METHODS

This chapter presents the evaluation methods used for the experimental results in the included articles. The main work of Articles I and II is to construct multi-resolution multi-organ deformable models for mouse abdominal micro-CT images. To evaluate the accuracy of model construction, we introduce two objective evaluation metrics: generalization ability and specificity (Davies et al., 2009). The main work of Article III is to use the constructed deformable model in Article I to segment livers, spleens, left kidneys and right kidneys in mouse abdominal images. We apply two spatial overlap-based metrics, the Dice coefficient (Taha and Hanbury, 2015) and averaged surface distance (ASD) (H. Wang et al., 2012) to evaluate the segmentation accuracy. The main work of Article IV is to develop an extensible open-source software platform with a friendly GUI and a flexible plug-in interface for integrating user-developed algorithm modules. We assess the software performance qualitatively from the aspects of convenience, time, and efficiency.

3.1 Evaluation metrics for shape models

The generalization ability of a deformable model measures the ability of the model to represent unknown samples other than the training samples. If a model is overfitting on training samples, it cannot represent unknown samples. In Articles I and II, we use a leave-one-out strategy[4] to measure the generalization ability of the model. One sample is selected from the dataset as the generalization target, and other samples are used as model training data. The constructed model is then fitted to the chosen sample. This process is repeated until every sample in the dataset has been selected. Therefore, the generalization ability G_M of the model can be defined as:

---

[4] https://www.baeldung.com/cs/cross-validation-k-fold-loo

$$G_M = \sqrt[2]{\frac{1}{n}\sum\nolimits_{i=1}^{n} \left\| x_M^i - x_i \right\|^2} \tag{25}$$

where n is the number of all samples (equal to shapes). x_M^i is the constructed deformable model with all samples excluding sample x_i.

The specificity of the model is used to measure the ability of the deformable model to represent the samples in the training set. A model with high specificity should only generate samples in the training set, or generate instances close to the samples in the training set. Therefore, the level of specificity is highly related to the success of model construction. According to the probability density defined by the model, we randomly generate k samples using a deformable model. Each sample is compared to members in the training set. Therefore, the specificity S_M of the model can be defined as:

$$S_M = \sqrt[2]{\frac{1}{k}\sum\nolimits_{j=1}^{k} \left\| x_M^j - x_j' \right\|^2} \tag{26}$$

where x_M^j is a shape instance generated by the constructed model. x_j' is the corresponding nearest sample in the training set. From the definition of generalization and specificity of the model, we know that the smaller the values calculated by equations (25) and (26) are, the better the model constructed is.

3.2 Evaluation metrics for segmentation

In 3D MIS, the accuracy of segmentation is usually represented by the spatial overlap ratio between the segmented results and the gold standard (expert segmented results). One of the concrete representations of the overlap ratio is the Dice coefficient (Dice, 1945). It is used most widely to measure the coincidence degree of 3D volume. The dice coefficient is defined as follows:

$$\text{Dice} = \frac{2|R_1 \cap R_2|}{|R_1| + |R_2|} \tag{27}$$

where R_1 and R_2 represent segmented organs and gold standard respectively. $|\cdot|$ represents the number of voxels. $\cap$ represents the overlapping regions between R_1 and R_2. According to the equation, the Dice coefficient is between 0 and 1. If Dice coefficient is 0, it means no overlapping between these two regions. If Dice coefficient is 1, it means complete overlapping between these two regions. The higher the value is, the better the segmented result is.

Another one of the concrete representations of the overlap ratio is ASD. The ASD is used to measure the distance from the surface of the segmented regions to the surface of the gold standard. The formula of ASD is written as:

$$\mathrm{ASD} = \frac{1}{2}\left(\frac{1}{n_1}\sum\nolimits_{i=0}^{n_1} d_i + \frac{1}{n_2}\sum\nolimits_{j=0}^{n_2} d_j\right) \tag{28}$$

where n_1 and n_2 are the numbers of vertices on the segmented organs and gold standard respectively. Calculate the distances from the i th vertex on the segmented organs to n_2 vertices on gold standard. The minimum distance is denoted by d_i. Likewise, calculate the distances from the jth vertex on gold standard to n_1 vertices on segmented organs. The minimum distance is denoted by d_j. Based on the equation, the smaller the value is, the better the segmented result is.

3.3 Evaluation metrics for the development of the extensible open-source software

The performance of the AS software is evaluated across three key aspects: software prototyping, data annotation, and human-computer interaction. For users, convenience, fast response times, and software effectiveness stand as the most critical evaluation metrics. During software prototyping, AS offers a user-friendly and flexible GUI, integrating developed DL models. This effectively bridges the gap between algorithm development and clinical applications, particularly enhancing efficiency for specific image modalities. Besides this, medical image annotation presents a significant challenge for DL developers. AS offers flexible deployment options on computers and tablets, enabling users to conveniently annotate images in less time. Thus data annotation facilitates DL model training, especially for specific tasks.

While fully automated methods may yield faster results, they often lack robustness in complex clinical scenarios. Semi-automated methods, which rely on human-computer interaction, prove crucial in practical applications. AS enables users to correct annotations in real-time during DL training, leading to significant accuracy improvements. The concrete performance is assessed across four distinct applications: customized MRI image diagnosis, interactive lung lobe segmentation, human-AI collaborated spine disc segmentation, and AID for DL model training.

4 OVERVIEW OF INCLUDED ARTICLES

This chapter first provides a brief structure of the research. Then it provides a concise overview of each article in the research, including objective, method, results, and the authors' contributions.

4.1 Structure of the research

The thesis examines the construction of multi-resolution multi-organ models using PCA-based methods and autoencoder networks, the segmentation of multiple organs through model deformation, and the development of more efficient MIA software. The entire thesis can be divided into three parts.

The first part of this thesis focuses on exploring the underlying mechanisms of multi-organ deformation models. We use the improved PCA method and SAE neural network to construct deformable models for mouse abdominal organs. For the purpose of research, we represent the organs in the image using a PDM (L. Lu et al., 2006). Firstly, during the modeling process using the PCA method, we calculate the Euclidean distances between different points to study the influence of point correlation on model deformation. We control modeling by selecting different distance thresholds, resulting in three levels of model deformation components ranging from global to local. Additionally, the large number of points required to build the model cause the HDLSS problem (Aoshima et al., 2018; Hall et al., 2005). We apply down-sampling-and-interpolation strategy to solve this problem. This study is important in revealing how organ correlation affects the deformation of each organ. Secondly, during the modeling process using the SAE neural network, we explore the potential non-linear deformation capabilities of the constructed model. PCA-based modeling has limited deformation capabilities and focuses on linear deformation modes. To exploit the non-linear deformation capabilities of the deformable model, we use the SAE neural network to extract features such as distance and shape of organs. Finally, we select the best deformation components for representing the model's deformation. These two studies for model construction further reveals that the correlation and independence of organs have a significant impact on the deformation of the constructed model. We test these conclusions on the dataset of mouse micro-CT images, which lays the foundation for the next step of image segmentation research.

The second part of this research focuses on the implementation of multi-organ segmentation in mouse images. To simultaneously realize the segmentation of liver, spleen, left kidney and right kidney in a mouse micro-CT image, we introduce the multi-resolution multi-organ deformation model constructed by the improved PCA method in the first part. Then we register the model to the image by iteration. At the same time, to better guide the model deformation, we use VHRBF interpolation function (Brazil et al., 2010) to roughly delineate the boundaries of the organs in the image during the registration process. For achieving a more accurate segmentation results, the above steps need to be iterated repeatedly. This part of the research plays an important role for exploring how to segment multiple organs quickly and accurately in medical images.

The third part of this study focuses on the development of MIA software. MIA is a complex process that includes a series of steps including data processing, data modeling, data visualization, and human-computer interaction interface design. Each step consumes a lot of researcher's energy. Therefore, a software that can help researchers quickly perform MIA, such as MIS, is essential. In this way, researchers can focus more on their core research work. In this part of our research, we develop an extensible open-source software platform, AnatomySketch (AS). It can flexibly integrate medical image processing algorithm modules developed by different users and assist users to quickly annotate and analyze data. Compared with previous development software, such as MITK (Wolf et al., 2005), AS can complete the algorithm development of MIA faster under the premise of ensuring the accuracy of the obtained results. The conclusions have been verified under four different exemplar applications, which means this software can provide more alternative algorithm development solutions for future research.

To sum up, Articles I and II contribute to the first part of the whole study. Article I first uses an improved PCA method to build a multi-resolution multi-organ model. Then we interpolate all deformation components for the constructed model. Compared with Article I, Article II mainly uses the SAE neural network to build a multi-layer deep neural network structure to train the organs represented by points. Features that control the deformation of different parts of the model are acquired during training. The parameters of the features are tuned to obtain multi-resolution models with various nonlinear deformations.

Article III contributes to the second part of the whole study. Based on Article I, Article III incorporates VHRBF function into the multi-resolution multi-organ deformation model. At the same time, human interaction is continuously imposed to iteratively register the liver, spleen, left kidney, and right kidney in the abdomen of a mouse to achieve simultaneous segmentation of multiple organs.

Article IV contributes to the third part of the whole study. To assist MIA algorithm development, Article IV develops a universal open-source software mostly by using C++ and python languages. The performance of the software is verified in different applications such as image annotation, integration of extensible algorithm modules and acceleration of DL training. In addition, it further clarifies contribution in a potential sharing and development of a user community.

4.2 Article I: Multi-resolution statistical shape model for multi-organ shape modelling

Zhonghua Chen, Tapani Ristaniemi, Fengyu Cong, and Hongkai Wang. (2020). Multi-resolution statistical shape models for multi-organ shape modelling. In Advances in Neural Networks – ISNN 2020: 17th International Symposium on Neural Networks, pp. 74-84. Springer, Cham.

Objective

PCA is widely used to build SSMs in medical images. However, the conventional modeling methods based on PCA suffer from the HDLSS problem. In addition, most PCA-based modeling methods aim at modeling a single specific organ or tissue. And the constructed models have limited deformation modes. Therefore, in this study, we aim to construct a 3D multi-organ statistical shape model with global-to-local deformation modes while addressing the HDLSS problem.

Method

In this study, the preprocessed 98 mouse abdominal multi-organ models in Section 2.2 were used as pre-training data. The pre-training data consists of 3D points and point connections:

- We first down sampled the pre-training data to obtain the training set. The training set also contains 98 shape samples.
- The mean shape of the shape samples was then calculated.
- We calculated the difference matrix for each shape sample to the mean shape as well as the corresponding covariance matrix.
- Then we computed the Euclidean distance between any two points on the mean shape to obtain a distance matrix and chose the maximum distance.
- Based on the maximum distance, we selected an appropriate threshold and set the values greater than the threshold in the distance matrix to 0. Likewise, we set the values smaller than the threshold in the distance matrix to 1. Thus, the mask matrix was then obtained. Its size is the same as the covariance matrix.

- We multiplied the corresponding elements of covariance matrix and the mask matrix and performed an eigenvalue decomposition for their results. The eigenvectors were obtained.
- Next, we carried out interpolation for the eigenvectors. The deformation components of the mean shape before down sampling were obtained. These components were added to the mean shape before down sampling to obtain a SSM which can deform from globally to locally.
- Finally, we evaluated the deformation performance from the generalization ability and specificity of the constructed SSM.

Results

We first conducted qualitative analysis on the constructed multi-resolution multi-organ SSM. Figure 19 shows 9 representative deformation results of the SSM. As shown in Figure 19, the constructed SSM mainly contains three levels of deformation effects: global level, single organ level and local organ level. We selected the first three components from three different levels respectively to demonstrate the deformation ability of the constructed SSM. PC_{g1}, PC_{g2}, and PC_{g3} are the selected deformation components for global level. Their corresponding eigenvalues are $\lambda_{g1}, \lambda_{g2}$, and λ_{g3}. α_{g1}, α_{g2}, and α_{g3} are the corresponding shape coefficients. Likewise, PC_{s1}, PC_{s2}, and PC_{s3} are the selected deformation components for single organ level. Their corresponding eigenvalues are $\lambda_{s1}, \lambda_{s2}$, and λ_{s3}. α_{s1}, α_{s2}, and α_{s3} are the corresponding shape coefficients. And for the right shapes in Figure 19, PC_{l1}, PC_{l2}, and PC_{l3} are the selected deformation components for local organ level. Their corresponding eigenvalues are $\lambda_{l1}, \lambda_{l2}$, and λ_{l3}. α_{l1}, α_{l2}, and α_{l3} are the corresponding shape coefficients. In Figure 19(a), the deformation components of the global level mainly control the deformation of the whole mean shape, including multiple organs simultaneously. In Figure 19(b), the deformation components at the single organ level mainly control the deformation of one organ in the mean shape, such as individual liver, spleen, and kidney. However, the deformation components at the local organ level in Figure 19(c) mainly control the small local deformation of one organ, such as the local deformation of the kidney and spleen.

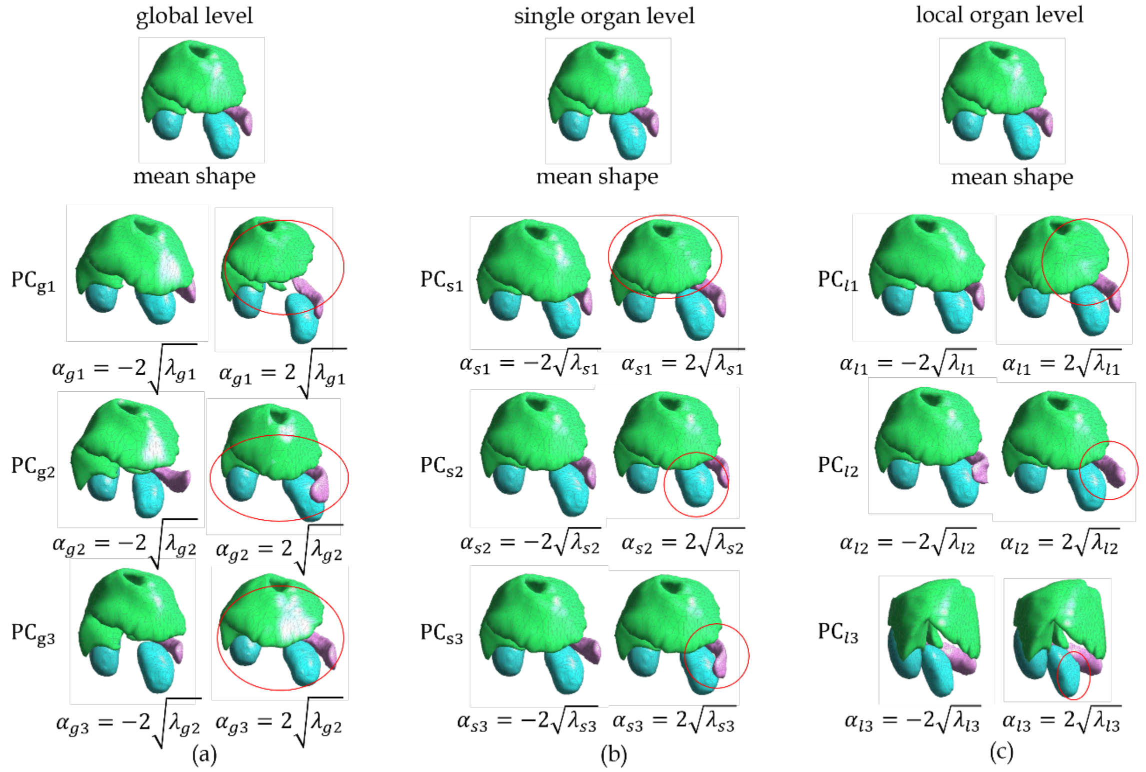


FIGURE 19 Demonstration of representative deformation results of three levels in the constructed SSM

We also quantitatively analyzed the constructed SSM from model's generalization ability and specificity. In this study, deformations at the global level, single organ level, and local organ level constitute a multi-resolution deformable model. The deformation modes at the global level corresponds to the traditional SSM deformation modes. They were used as the control group in this study. We computed the averaged generalization errors and specificity errors of the constructed SSM at three different levels, respectively. As shown in Figure 20, the averaged generalization errors, and averaged specificity errors at the three levels are all within 1.00 mm. From the global level to the local organ level, the generalization errors and specificity errors of the model are gradually decreased. This shows that multi-organ SSM with multi-resolution is more accurate for organ boundary registration. At the same time, it verified that the constructed SSM has better generalization ability and specificity than the traditional SSM. In addition, for the constructed SSM, the generalization errors and specificity errors at the local organ level are less than or equal to 0.3 mm. These errors are much lower than the generalization errors and specificity errors of the traditional SSM (global organ level). And they are close to the pixel resolution of 0.2 mm in mouse micro-CT images.

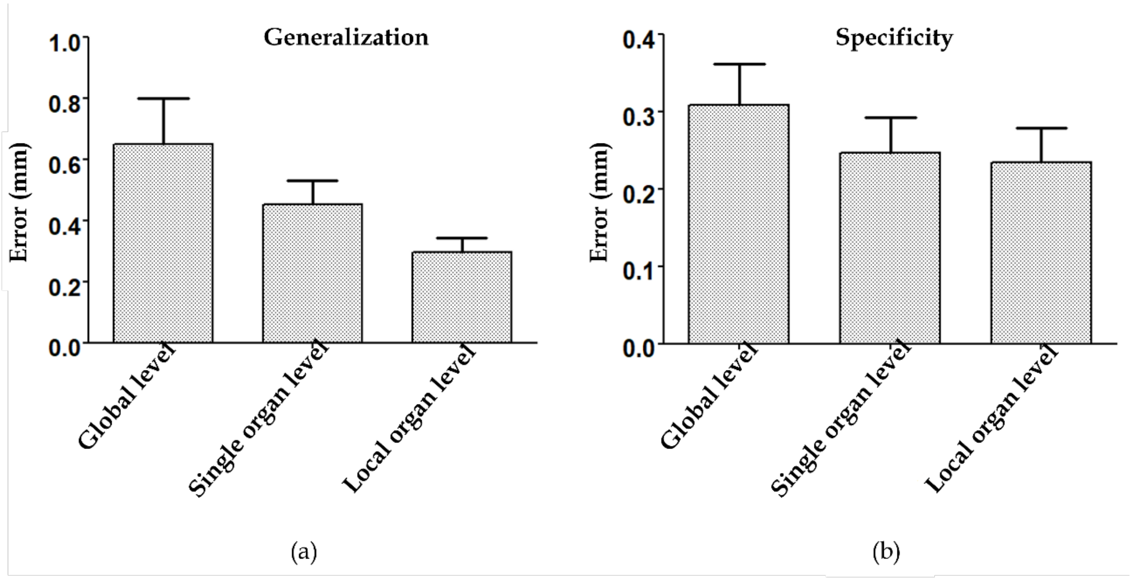


FIGURE 20 Generalization ability and specificity of the constructed SSM

Errors in Article I

In Figure 3 of Article I, there is an ambiguity regarding the deformation components, eigenvalues, and deformation coefficients representing the constructed SSM. Therefore, a correction is made in this paper. Errors in Article I are given in Table 1.

TABLE 1 Corrections to variables in Figure 3 of Article I

Type of variables	Current variables	Correct variables	Figure
Deformation component	PC_1, PC_2, and PC_3	PC_{g1}, PC_{g2}, and PC_{g3}.	3(a)
Deformation component	PC_1, PC_2, and PC_3	PC_{s1}, PC_{s2}, and PC_{s3}.	3(b)
Deformation component	PC_1, PC_2, and PC_3	PC_{l1}, PC_{l2}, and PC_{l3}.	3(c)
Eigenvalue	λ_1, λ_2, and λ_3	$\lambda_{g1}, \lambda_{g2}$, and λ_{g3}.	3(a)
Eigenvalue	λ_1, λ_2, and λ_3	$\lambda_{s1}, \lambda_{s2}$, and λ_{s3}.	3(b)
Eigenvalue	λ_1, λ_2, and λ_3	$\lambda_{l1}, \lambda_{l2}$, and λ_{l3}.	3(c)
Shape coefficient	α_1, α_2, and α_3	α_{g1}, α_{g2}, and α_{g3}.	3(a)
Shape coefficient	α_1, α_2, and α_3	α_{s1}, α_{s2}, and α_{s3}.	3(b)
Shape coefficient	α_1, α_2, and α_3	α_{l1}, α_{l2}, and α_{l3}.	3(c)

Authors contributions in this article

Zhonghua Chen and **Hongkai Wang** proposed the method. **Zhonghua Chen** conducted the data analysis and wrote the original manuscript. **Hongkai Wang** and **Tapani Ristaniemi** revised and supervised the manuscript. **Fengyu Cong** supervised the manuscript.

4.3 Article II: Construction of multi-resolution multi-organ shape model based on stacked autoencoder neural network

Zhonghua Chen, Hongkai Wang, Fengyu Cong, and Lauri Kettunen. (2022) Construction of multi-resolution multi-organ shape model based on stacked autoencoder neural network. In 2022 14th International Conference on Advanced Computational Intelligence (ICACI), pp. 62-67. IEEE.

Objective

Deformable model-based MIS methods contribute to the segmentation of anatomical structures dramatically. Most of the conventional modeling methods are realized based on PCA. However, most of PCA-based methods are linear. This means that the deformation modes of a deformable model are generated linearly. The combination of linear deformation components generates different shapes. But linear modeling methods oversimplify the complex of human anatomy. They may not adequately represent the true deformation patterns of organs and tissues. Therefore, in this study we aim to explore more nonlinear deformation modes of deformable models by using a SAE neural network.

Method

Autoencoders efficiently compress data and extract nonlinear features of data during encoding. During the decoding process, autoencoders show good performance for data recovery and image reconstruction. Based on this, in this study, we still used 98 abdominal mouse micro-CT images preprocessed in Section 2.2 as training data. Each training sample consists of 752 3D points. For constructing a multi-level SAE, we concatenated the encoding and decoding layers of three autoencoders respectively. Then, the concatenated encoding and decoding layers were combined. After numerous tests and adjustments, we finally chose a three-level SAE with three autoencoders. The number of neurons in each layer in encoding part is 2256, 1000, 200, 30. Likewise, the number of each layer in decoding part is 30, 200, 1000, 2256. The decoding part and the encoding part are symmetrically in the three-level SAE. 98 training data were input to train the built SAE network. We continuously optimized the parameters of the SAE network to obtain a better SAE network qualitatively. Next, the mean shape of the training data was input into the SAE network to extract 30-dimensional features. The 30-dimensional features were continuously modified and decoded simultaneously. Finally, a multi-organ model with multiple nonlinear deformation modes was obtained.

Results

We fed the mean model of the training data into the trained three-level SAE network. The model was encoded by three encoding layers to obtain a 30-dimensional feature vector $\vec{F} = \{F_1, F_2, \dots, F_{30}\}$. We chose 0.1 as an offset unit. Then we selected one value from $\vec{F}$ for numerical modification and kept the other values unmodified. The modified feature vectors are reconstructed into different multi-resolution multi-organ models through decoding layers. We found during decoding that modifications in some values controlled similar shape changes. Therefore, we picked out some representative modified values to demonstrate the constructed model. Figure 21 shows 17 representative organ models reconstructed with modified values. Columns 1, 3, and 5 in Figure 21 are mean models. Columns 2, 4, and 6 are reconstructed models. Deformed regions of the reconstructed models are circled. From the figure, the modifications of values F_4, F_7, F_{11}, and F_{15} correspond to the deformation of a liver. Modifications of values F_2, F_3, F_6, and F_{13} correspond to the deformation of a spleen. Modifications of values F_7, F_{15}, F_{30} correspond to deformation of the right kidney. Modifications of values F_{13}, F_{14}, F_{25}, and F_{30} correspond to deformation of the left kidney. However, modifications of values F_8, F_{26} can control the relative positions of organs in models. As a comparison, we also show the deformation effect of the multi-resolution multi-organ SSM constructed in Article I. As shown in Figure 22, columns 1 and 4 are mean models $\boldsymbol{\mu}$. $\{PC_1, PC_2, \ldots, PC_{10}\}$ are deformation components. $\{\lambda_1, \lambda_2, \ldots, \lambda_{10}\}$ are the corresponding eigenvalues. $\{\alpha_1, \alpha_2, \ldots, \alpha_{10}\}$ are the deformation coefficients. The deformation modes of the multi-organ model in Figure 22 are mainly manifested on the liver and spleen. But these deformation modes can also be almost all achieved by three-level SAE. We qualitatively analyzed Figures 21 and 22 and found that three-level SAE can be constructed with more different kinds of nonlinear deformation modes. These modes include deformations at the global level, the single organ level, the local organ level, and the organ relative position level, etc. The organ model thus constructed in this study has richer deformation effects.

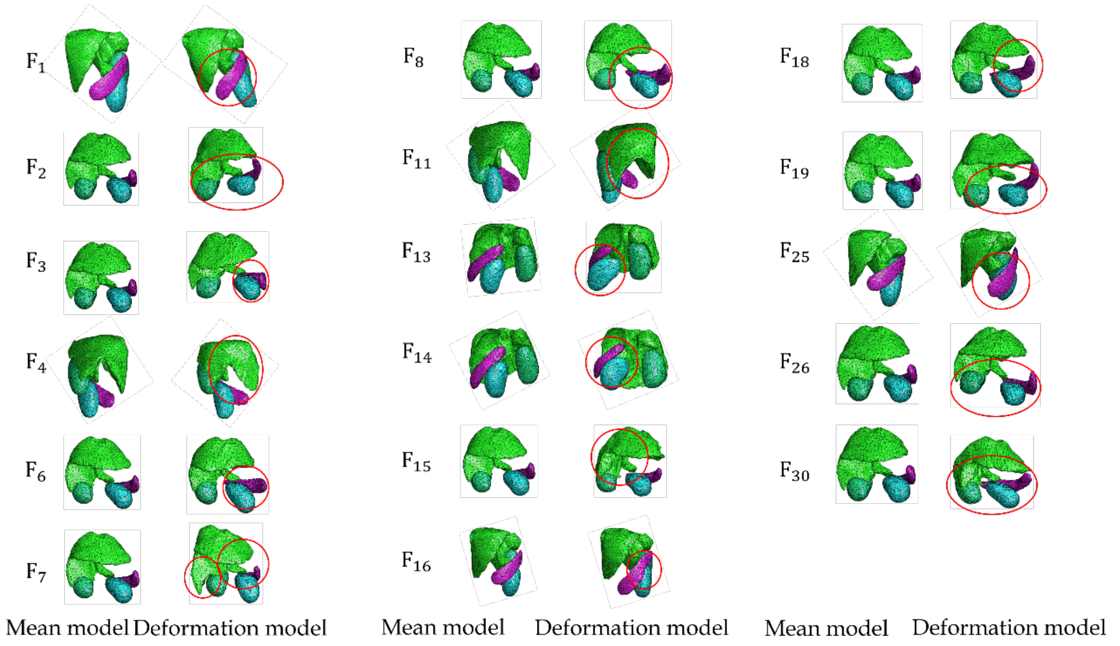


FIGURE 21 Different deformations of the constructed model corresponding to different 30-dimensional features

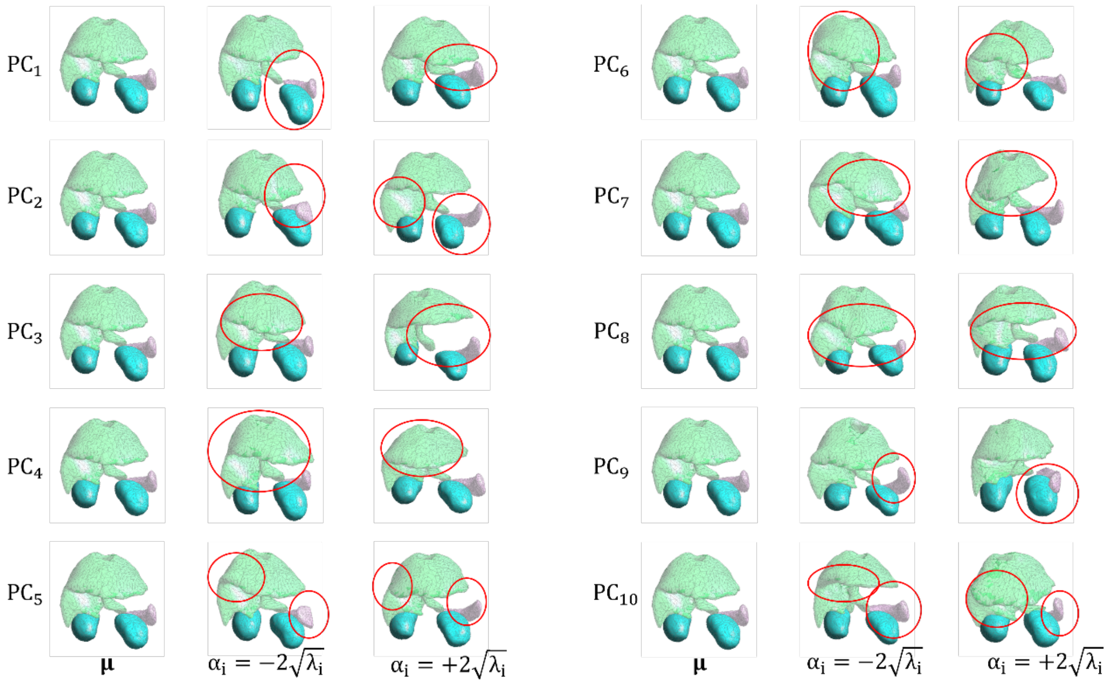
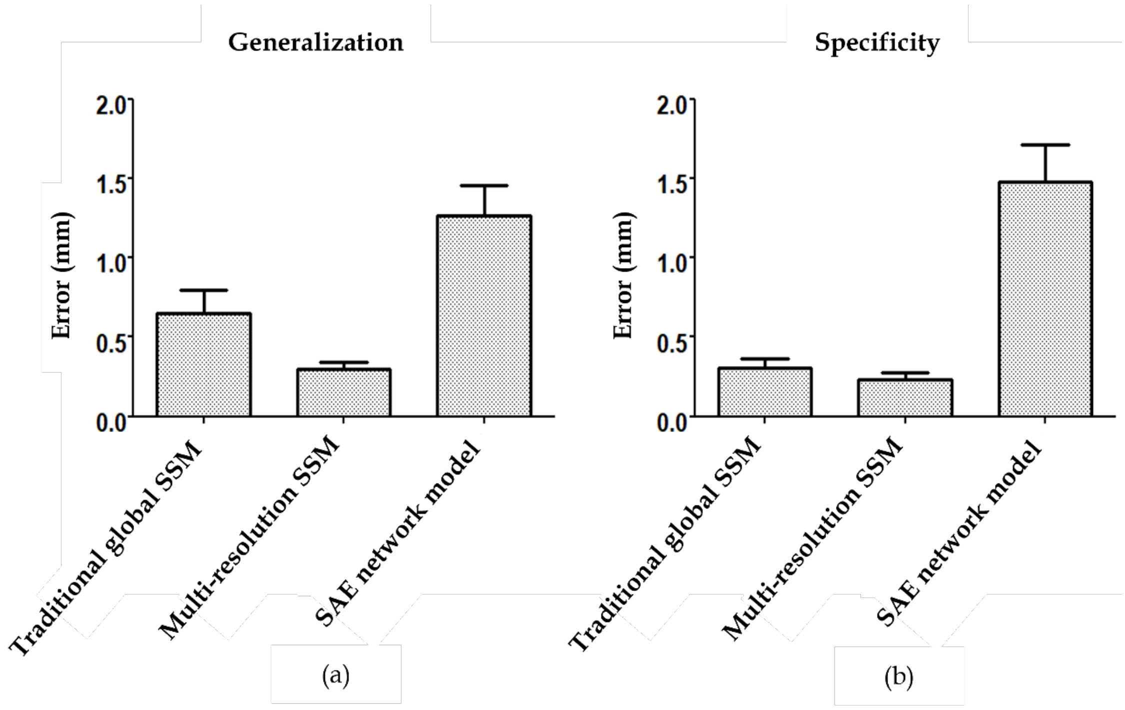


FIGURE 22 Different deformations of a multi-resolution multi-organ SSM constructed in Article I

We also quantitatively analyzed the constructed model by using generalization ability and specificity. Figure 23 compares the three-level SAE modeling effect and the SSM modeling effect. Figure 23(a) shows that in terms of generalization error, the modeling errors of traditional global multi-organ SSM and multi-resolution SSM are respectively 0.5 and 0.3 mm. But the SAE modeling error is about 1.5 mm. Figure 23(b) shows that, in terms of specificity error, the modeling errors of traditional global multi-organ SSM and multi-resolution SSM are around 0.3 mm. The modeling error of SAE is still around 1.5 mm. From this analysis, we know although SAE is simple in principle and easy to operate, it may be inferior to the modeling methods based on PCA. It is also possible that during the training process of SAE, the hyperparameter adjustment is unreasonable and the error becomes larger. These challenges require further verification in the future.

Authors contributions in this article

Zhonghua Chen and **Hongkai Wang** proposed the method. **Zhonghua Chen** conducted the data analysis and wrote the original manuscript. **Hongkai Wang** and **Lauri Kettunen** revised and supervised the manuscript. **Fengyu Cong** supervised the manuscript.



(a)

(b)

FIGURE 23 Modeling error comparison between SAE and PCA-based method

4.4 Article III: Low-dose mouse micro-CT image segmentation based on multi-resolution multi-organ shape prior knowledge model

Zhonghua Chen, Hongkai Wang, Fengyu Cong, and Lauri Kettunen. (2022). Low-dose mouse micro-CT image segmentation based on multi-resolution multi-organ shape prior knowledge model. In 2022 International Conference on Computers, Information Processing and Advanced Education (CIPAE), pp. 349-353. IEEE.

Objective

Multi-organ segmentation is a difficult problem in medical image processing. Automatically segmenting multiple organs in mouse micro-CT images is challenging due to poor image quality. Images often have blurred edges and low contrast between organs, making it hard to develop fully automatic segmentation methods. As a result, manual interaction is often necessary to improve segmentation accuracy. In this study, we propose a new human interactive segmentation method to accurately segment multiple organs in mouse abdominal micro-CT images. Our method uses a deformable model guided by a small amount of human interaction. The goal is to achieve accurate segmentation of multiple organs in these challenging images.

Method

In this study, we used 98 micro-CT images of mice that were collected in Section 2.2 as our experimental subjects. To establish the gold standard, small animal imaging specialists segmented the livers, spleens, left and right kidneys. We then introduced random noise and reduced contrast to create low-quality, low-contrast images. The operators used these low-contrast images to delineate landmarks at the edges of the organs, which were then used to generate the full contour of these organs using VRHBF interpolation. We imported a multi-resolution multi-organ SSM that had been constructed in Article I and used the delineated organ contours to guide the SSM's deformation. By doing so, we were able to limit excessive deformation of the SSM, allowing it to reflect the actual shape of the organs more accurately. The SSM was continually adjusted and deformed under the guidance of the contours until a satisfactory result was obtained. This process was repeated multiple times by adding additional landmarks and generating new contours until the results were achieved.

Results

To test the segmentation, we randomly selected 7 images from a set of 98 low-contrast mouse micro-CT images. We compared the results of the traditional global resolution SSM and this proposed method. The accuracy of the segmentation was evaluated using the Dice value and ASD. In addition, we assessed the method's stability by testing significant differences in segmentation result when different users performed subjective operations.

We began by analyzing the results of the two segmentation methods qualitatively. Figure 24 provides a comparison of the segmented results obtained by the two methods. Each row in the figure represents a randomly selected test image. The gold standard is shown in Figure 24(a). Figure 24(b) and Figure 24(c) respectively depict the segmented results of our proposed method and the traditional global resolution SSM method. Upon comparing the two methods, we observed that the segmented results obtained from our proposed method were superior to those of the traditional global resolution SSM and closer to the gold standard.

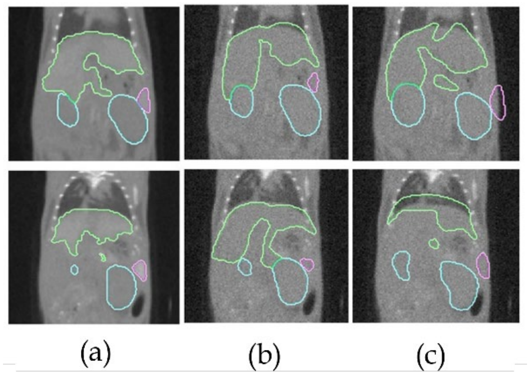


FIGURE 24 Comparison of segmented results between the proposed and traditional global resolution SSM methods

We also performed a quantitative analysis and comparison of the two segmentation methods. We calculated the Dice values and ASD values of the obtained results, as shown in Figure 25. To measure the significance of differences in the segmentation results, we employed the t-test. The figure indicates that our proposed method achieved a higher Dice value and a lower ASD value for each organ. Additionally, we observed significant differences in the segmentation of spleens and kidneys. These results collectively demonstrate the superiority of our proposed method over the traditional global resolution SSM method.

We also performed a quantitative analysis to evaluate the robustness of our method. Figure 26 displays the differences in segmented results obtained by different operators when using our proposed method. The t-test reveals no significant differences in the results, indicating the robustness of our proposed method.

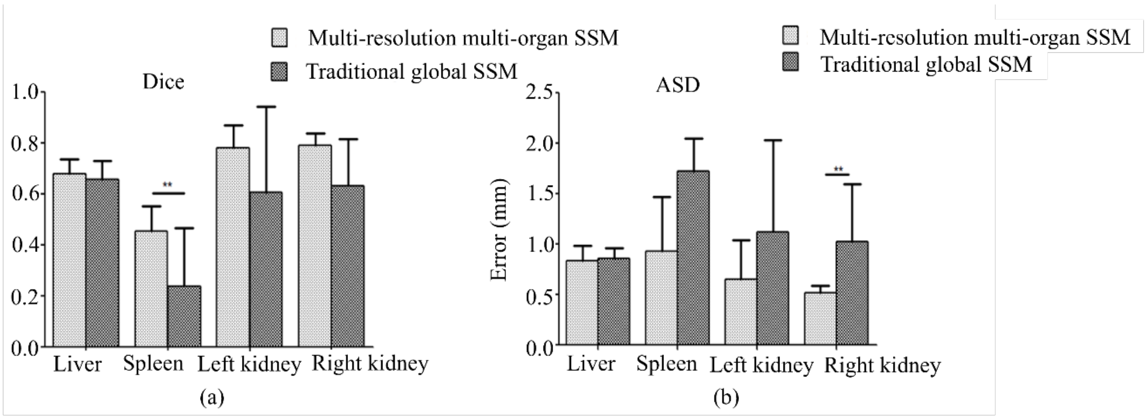


FIGURE 25 Comparison of segmentation accuracy between the proposed and the traditional global resolution SSM methods

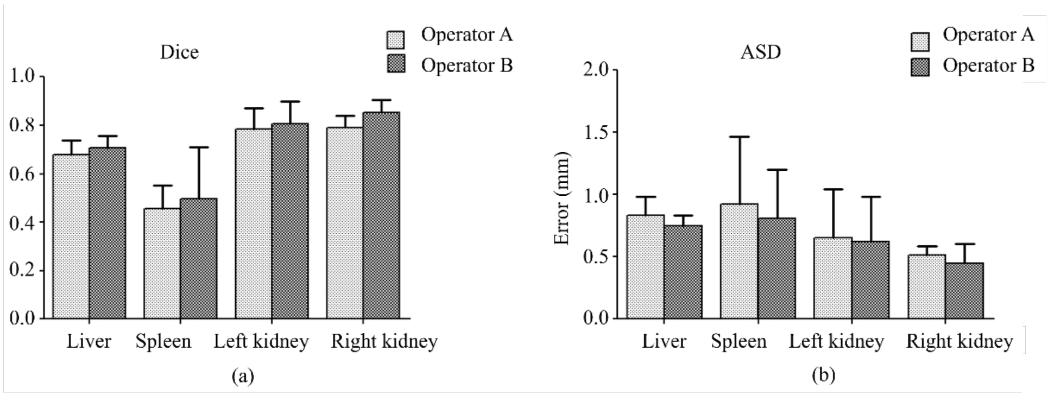


FIGURE 26 Evaluation of the proposed method based on different operators

The proposed method was evaluated on a computer equipped with an Intel Core i7 CPU and 16 GB of memory. User interaction time varied between 30 seconds and 1 minute, while the VHRBF interpolation time ranged between 2 to 3 seconds. The model registration and segmentation time varied from 15 to 25 seconds. The entire image segmentation process was completed within 1 to 1.5 minutes. These results demonstrate that the proposed method meets the requirement for practical interactive segmentation with reasonable waiting time.

Authors contributions in this article

Zhonghua Chen and **Hongkai Wang** proposed the method. **Zhonghua Chen** conducted the data analysis and wrote the original manuscript. **Hongkai Wang** and **Lauri Kettunen** revised and supervised the manuscript. **Fengyu Cong** supervised the manuscript.

4.5 Article IV: AnatomySketch: An extensible open-source software platform for medical image analysis algorithm development

Mingrui Zhuang, Zhonghua Chen, Hongkai Wang, Hong Tang, Jiang He, Bobo Qin, Yuxin Yang, Xiaoxian Jin, Mengzhu Yu, Baitao Jin, Taijing Li, and Lauri Kettunen. (2022). AnatomySketch: An extensible open-source software platform for medical image analysis algorithm development. Journal of Digital Imaging, 35, 1623-1633.

Objective

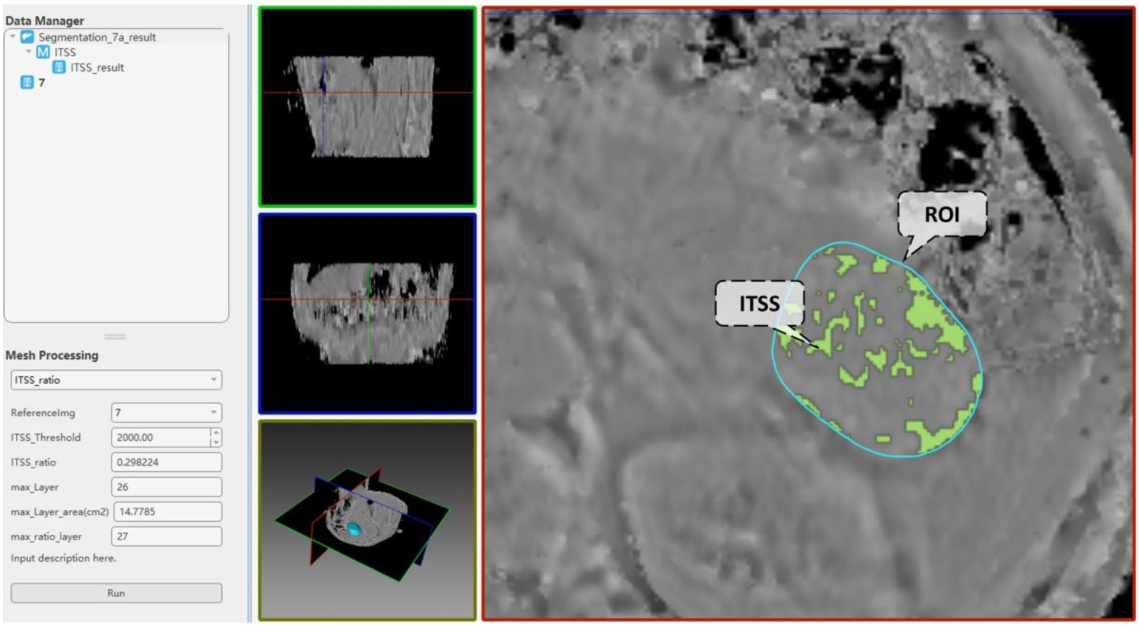
Developing MIA algorithms is complex and challenging. The process involves several steps, such as requirement analysis, algorithm design, coding, testing, data visualization, human-computer interaction, and GUI construction, which require significant resources. Existing development tools do not effectively integrate these steps, particularly for DL algorithms that require a software platform capable of annotating training data and facilitating user interaction. To address this issue, we aim to build an open-source software platform with a friendly GUI and a convenient, extensible plug-in interface that facilitates algorithm integration and development. This platform can integrate multiple steps in the algorithm development process, helping developers improve development efficiency.

Method

The AS software is primarily implemented using C++ and Python. To begin, we have developed a user-friendly GUI for the AS software, which includes a menu bar at the top, a data management list, and a user development module panel on the left-hand side. On the right-hand side, there is a display window that shows three different image planes. AS software is available for both desktop computers and tablets. The software design supports a range of input devices, including mouse, keyboard, stylus, and touch screen. Secondly, in the display window, we have included interactive annotation and proofreading functions for images. Users can use a stylus or mouse to annotate ROIs in the image. These annotations can be used as training data to train network models or saved separately on the computer disk. We have also integrated a flexible plug-in interface in AS, which allows users to integrate their own developed programs into AS. They can use AS to call their own programs to process and display the image. Importantly, AS can help users to annotate training data and train DL models efficiently.

Results

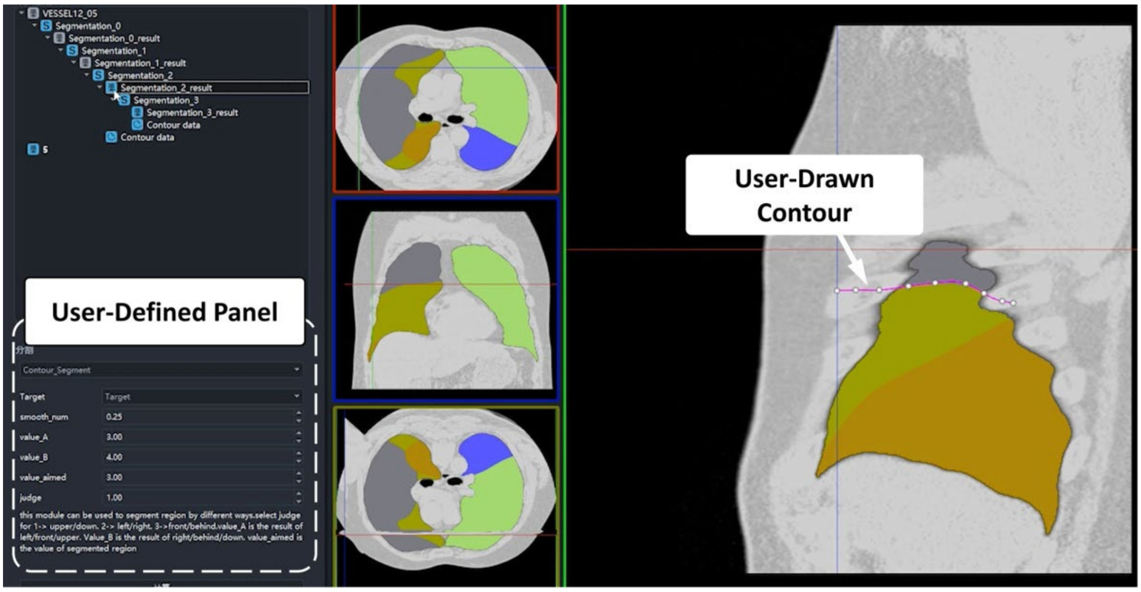
In this study, we have presented four different applications in MIA. These applications include the diagnosis of MRI images, segmentation of lung lobes, segmentation of spinal discs, and the annotation for DL models. For MRI image diagnosis, we developed a plug-in module that analyzes intrahepatic tumors and created a GUI-based software prototype. Physicians used our prototype to delineate ROIs with greater precision in 2D or 3D medical images. The GUI interface of the software prototype is shown in Figure 27. The blue outline represents the ROI boundary, and the green area include the extracted targets.



ITSS_ratio	
ReferenceImg	7
ITSS_Threshold	2000.00
ITSS_ratio	0.298224
max_Layer	26
max_Layer_area(cm2)	14.7785
max_ratio_layer	27

FIGURE 27 An example of a user-defined extension module for target analysis in MR images

To segment lung lobes in CT images, we collaborated with two engineers from a company who developed a lobe annotation plugin. The annotated lung lobes were then used to train the lung segmentation network. Figure 28 displays the GUI interface of the lobe annotation plugin. Each lung lobe image annotation took less than 20 minutes. With the support of a stylus, the engineers were able to annotate 100 lung lobes within three days, significantly reducing the time required compared to other software.



Contour_Segment
Target Target
smooth_num 0.25
value_A 3.00
value_B 4.00
value_aimed 3.00
judge 1.00

this module can be used to segment region by different ways.select judge for 1-> upper/down. 2-> left/right. 3->front/behind.value_A is the result of left/front/upper. Value_B is the result of right/behind/down. value_aimed is the value of segmented region

FIGURE 28 An example of a user-developed plugin module for lung lobe annotation in CT images

To demonstrate the performance of AS for intervertebral disc segmentation, we integrated a user-developed DL model via a plug-in interface. Figure 29 illustrates the results of calling a dense V-Net model (Gibson et al., 2018) in AS to segment the intervertebral disc. In Figure 29(a), without manual annotation, the dense V-Net cannot accurately segment the herniated disc. In Figure 29(b), AS and the dense V-Net were used to manually correct the image, resulting in more accurate segmentation of the herniated part of the intervertebral disc. Although both MITK and 3D Slicer can annotate images, they require a network connection for data processing, which can be inconvenient and raise privacy concerns. Furthermore, existing software tools lack efficient proofreading of the segmentation results of DL models.

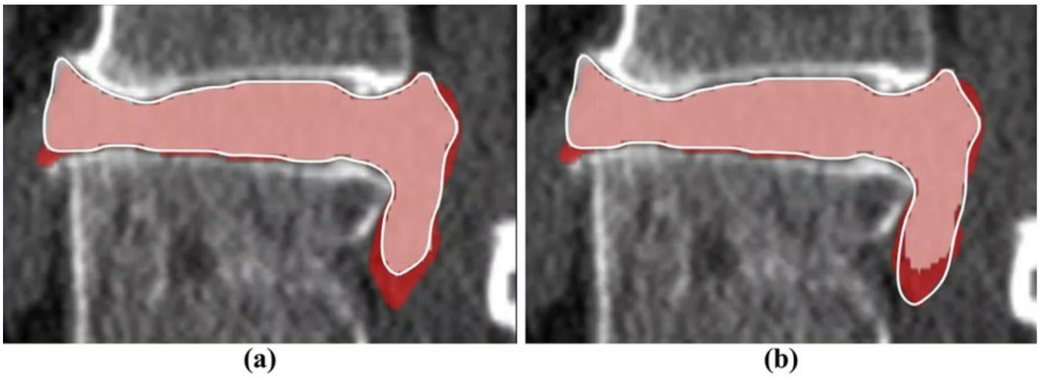


FIGURE 29 An example of manual proofreading. The red areas are ground truths

The most significant core function of AS software is to facilitate an efficient AID process. To demonstrate this, we used a DeepSnake network (Peng et al., 2020) to segment abdominal organs in CT images. Experts initially trained the network by annotating the outlines of organs in a small set of training images. The trained network then segmented more contours for other images, which were subsequently corrected using FFD tools. The rectified images were then added to the network to further optimize it. Figure 30 displays the segmentation results of the initial network and the rectified network. The initial network training data in Figure 30 contained 600 CT slices, while the rectified network contained 1800 CT slices. As shown in Figure 30, correcting the network leads to more accurate results.

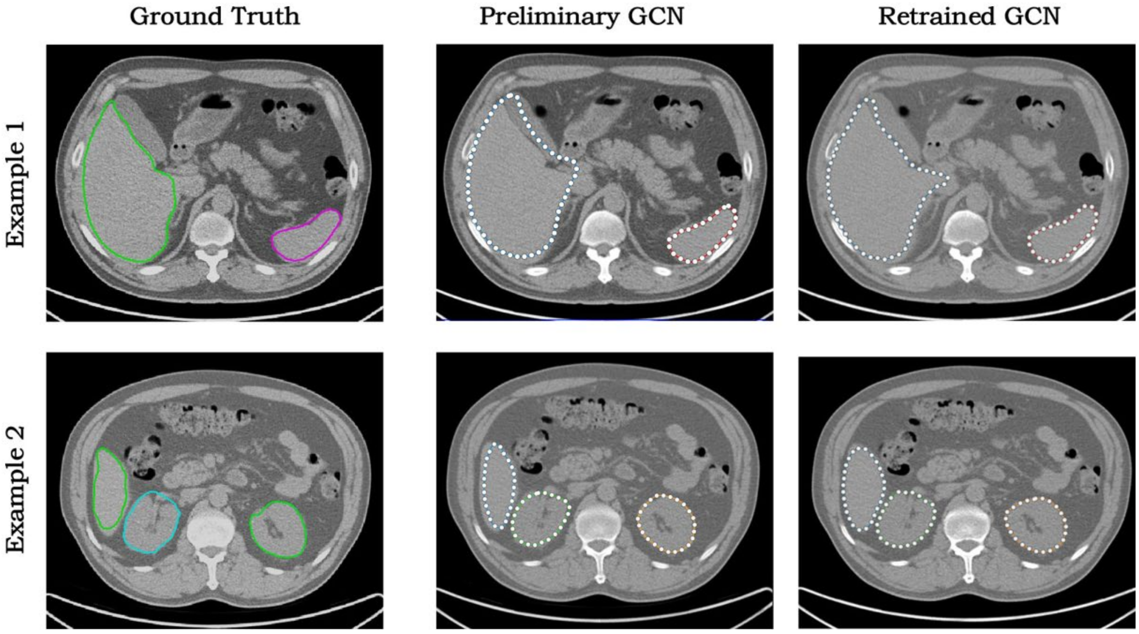


FIGURE 30 AID annotation results of two exemplar CT slices

Authors contributions in this article

Mingrui Zhuang developed software, conducted experiments, tested data, and contributed to manuscript writing. **Zhonghua Chen** developed the function module of statistical shape model segmentation of the software, conducted experiments, tested data, and contributed to manuscript writing. **Hongkai Wang** proposed the method, revised, and supervised the manuscript. **Hong Tang** revised and supervised the manuscript. **Jiang He** processed and tested data. **Bobo Qin** processed and tested data. **Yuxin Yang** processed and tested data. **Xiaoxian Jin** processed and tested data. **Mengzhu Yu** processed and tested data. **Baitao Jin** processed and tested data. **Taijing Li** processed and tested data. **Lauri Kettunen** revised and supervised the manuscript.

5 DISCUSSION

In this chapter, we discuss the whole research including the used dataset, the methods, outcomes, and limitations of each study, as well as potential directions for future research in this thesis.

5.1 Mouse data

Mice share many similarities with humans in terms of genetics (Bryda, 2013), anatomy and organ systems (Rosenthal and Brown, 2007), and disease mechanisms (Tassone et al., 2012). Consequently, mice play a crucial role as valuable models for investigating human biology and disease. For this reason they are commonly used animals in biomedical research. The utility of mice extends to various areas, including the evaluation of drug candidate efficacy, safety, and pharmacokinetics (Festing and Wilkinson, 2007), tumor imaging (Heskamp et al., 2015), cardiovascular research (Wessels and Sedmera, 2003), neuroimaging (Mezzanotte et al., 2017; Xiong et al., 2021), as well as drug development (Sharpless and DePinho, 2006; Weissleder and Ntziachristos, 2003). In vivo mouse imaging holds great significance in facilitating these studies. In medical physics studies, the utilization of micro-CT imaging has become pervasive for visualizing internal organs of small animals. For such reasons, in Articles I and II we used enhanced-contrast mouse micro-CT images to construct different deformable models.

However, in the real world most of the CT images acquired have low contrast. Therefore, in Article III we also employed corresponding low-contrast images for multiple organ segmentation. Practically, the utilization of low-contrast mouse micro-CT images for MIA in research may bring about several benefits and challenges:

- These low-contrast images simulate real-world scenarios where distinguishing subtle boundaries between tissues or structures create a challenge. By working with low-contrast images, the robustness and accuracy of segmentation algorithms in clinical-like scenarios can be evaluated (Clark and Badea, 2021). But the lower contrast levels in the images introduce additional complexity to the segmentation process, requiring more sophisticated algorithms and techniques to achieve accurate results (Schoppe et al., 2020). This is to say, MIA based on low-contrast mouse micro-CT images still remains a big challenge in research.

- The outcomes of segmentation algorithm presented in Article III perform well on low-contrast mouse micro-CT images. This suggests, the algortihms are potential to be applied and tested in more general clinical settings. Algorithms that accurately segment low-contrast structures of mice should perform better when applied to imaging humans, which share the same challenges.
- In addition, mouse micro-CT images serve as a valuable benchmark for algorithm development and validation (Girard et al., 2016). They provide a standard for comparing different segmentation techniques and assessing their performance in challenging scenarios.
- Generation of low-contrast images involves reducing the contrast agent concentration or adjusting imaging parameters. This can help minimize resource usage and potential animal welfare concerns. Researchers can optimize existing data without the need for additional contrast agents, thus promoting a more efficient and ethical use of experimental resources (Kagadis et al., 2010).
- Unfortunately, the decreased contrast may lead to reduced accuracy and increased segmentation errors (Baiker et al., 2010). It is crucial to strike a balance between realistic image simulation and maintaining an acceptable level of image quality for reliable segmentation results. The quality of true low-contrast images is highly dependent on acquisition parameters, such as exposure settings and reconstruction algorithms (Altunbas et al., 2006). These parameters may vary and introduce inconsistencies in image quality.

5.2 Methods

To address research question 1 raised in the introduction section, Article I develops an improved-PCA method for constructing a multi-organ multi-resolution SSM based on enhanced-contrast mouse micro-CT images. The improved-PCA method incorporates improved statistical measures to handle noisy and outlier data, resulting in more reliable shape models. It can capture both global and local shape variations by considering the entire shape distribution as well as local shape deformation modes. We find that this flexibility allows for a more comprehensive representation of shape variations, compared to traditional PCA methods (Cootes et al., 2001; Stegmann and Gomez, 2002).

In addition, we find the improved-PCA method efficiently computes shape models by exploiting the eigenvalue decomposition of the modified covariance matrix. This makes it computationally efficient, enabling the analysis of large datasets by avoiding the HDLSS problem (Aoshima et al., 2018; Hall et al., 2005). Compared to the method for constructing 2D multi-organ multi-resolution SSMs (Wilms et al., 2017), the improved-PCA method can efficiently generate 3D multi-organ multi-resolution SSMs for mouse micro-CT images.

To address research question 2, Article II proposes a SAE neural network to create deformable models with more non-linear deformation modes. We find SAE can capture non-linear deformation modes, allowing for more flexible representation of complex shape variations when compared with PCA methods (Z. Chen et al., 2020; Cootes et al., 2001; Stegmann and Gomez, 2002). The autoencoder architecture of the SAE facilitates unsupervised learning and feature extraction, reducing the need for labeled training data and expert-defined features (P. Vincent et al., 2010). The SAE learns directly from the data, allowing for the discovery of intricate shape patterns and reducing the reliance on predefined statistical assumptions.

To address research question 3, Article III uses human interaction to guide the constructed SSM of Article I to simultaneously segment the livers, spleens, left and right kidneys in mouse micro-CT images. The incorporation of prior knowledge played a crucial role in the study. We find that the segmentation process can be guided towards more accurate and anatomically plausible results as compared to traditional PCA methods (Cootes et al., 2001; Stegmann and Gomez, 2002). CT images often suffer from noise, artifacts, and intensity variations (Davis and Vachhani, 2017; Maier et al., 2018). Incorporating prior knowledge helps in compensating for these issues and enhancing the robustness of the segmentation algorithm. The prior knowledge provides one with contextual information about the expected shape and appearance of structures of CT images. By incorporating the constructed shape model, human interaction can result in more informed decisions during segmentation, leading to improved outcomes (Heimann and Meinzer, 2009).

Multi-organ modeling and segmentation are two of essential analysis in MIA. To address research question 4, Article IV develops the AS software which facilitates the realization of these two approaches. Some existed tools, such as itk (McCormick et al., 2014; Yoo et al., 2002), vtk (Schroeder et al., 2006), MITK (Wolf et al., 2005), have limited extension for user-developed algorithms compared to the developed AS software. Likewise, at present, the 3D Slicer (Pieper et al., 2004) fails to achieve convenient interaction with users and user-friendly programming mechanism. The AS software has a user-friendly GUI interface and a flexible and extensible interface, making it convenient for users to develop algorithms for visual analysis.

5.3 Methods limitations and future directions

However, the methods employed to address the four research questions have some limitations.

For research question 1, the improved-PCA method relies on the assumption that shape variations can be adequately represented by linear combinations of principal components. However, in scenarios where shape variations exhibit significant non-linearities (Z. Chen et al., 2022; Heimann and Meinzer, 2009; Nolte et al., 2016), the linear approximation of PCA may not capture the full complexity of the data. This limitation can result in suboptimal representation of shapes with intricate deformations, such as livers and spleens. In addition, the improved-PCA method heavily relies on the quality and quantity of training data. Insufficient or biased training data can lead to limited model generalization and reduced accuracy in capturing shape variations (Cerrolaza et al., 2019). Moreover, the method assumes that shape variations across different instances are independent and identically distributed, which might not hold true in certain datasets with complex shape relationships.

For research question 2, one significant limitation of the SAE network is the risk of overfitting, especially when the network architecture becomes excessively complex or when the training dataset is limited. In addition, the high capacity of the SAE to learn intricate patterns and non-linear deformations can result in models that are overly specific to the training dataset and fail to generalize well to unseen data (P. Vincent et al., 2010). As compared to improved-PCA method (Z. Chen et al., 2020), this may result in poor construction of deformable models. And they are prone to the chosen dataset, network architecture and hyperparameters. Another limitation of the SAE is the lack of interpretability of their learned representations. While the SAE excels at automatically learning feature representations, the resulting latent space may not have direct and intuitive semantic meanings. This makes it challenging to interpret and extract meaningful insights from the learned deformations (Shen et al., 2017). This limitation can hinder the understanding of the underlying mechanisms of shape variations.

For research question 3, the effectiveness of incorporating prior knowledge, however, relies on the availability and accuracy of the prior information. If the prior knowledge is incomplete, inaccurate, or not representative of the specific imaging conditions or anatomical variations, it can adversely affect the segmentation accuracy. In addition, incorporating prior knowledge often involves an initial registration step to align the prior model with the target image as introduced in Section 2.5. Any errors or misalignments introduced during the registration process can propagate and lead to inaccurate segmentations. Ensuring robust and accurate registration techniques is crucial to mitigate this limitation. The effectiveness of the shape prior knowledge in segmenting organs or tissues relies on the assumption that the target image shares similar characteristics with the training data. If the imaging conditions, anatomical variations, or disease states differ significantly from the training data, the segmentation accuracy may be compromised. Generalizing the method to handle novel cases with different imaging characteristics remains a challenge. While the improved-PCA method can handle shape variations, it may face challenges in accurately segmenting anatomical structures with intricate shapes, irregular boundaries, or overlapping regions. Incorporating prior knowledge alone may not fully address these complexities, and additional techniques may be necessary to improve segmentation accuracy.

For research question 4, the current version of AS is not yet mature, and more functions need to be upgraded to meet user needs. Moreover, the growth of plug-in modules also increases the difficulty of AS community maintenance. The availability of user support, documentation, and an active community of developers and users is crucial for the success and continuous improvement of AS. If the AS platform lacks a strong user community or sufficient resources for support and updates, users may also face challenges in troubleshooting issues or accessing new features and improvements.

To address these limitations, future research can focus on:

- Explore constructing multi-resolution multi-organ models of different organisms using the improved PCA-based method. It is important to evaluate the modeling accuracy of various models to determine the method's universality.
- To generate a wider range of nonlinear deformation modes, we can collect more medical images for training the network. We can also adjust the network structure and parameters to improve its modeling performance.
- Adversarial learning techniques, such as GANs (Yi et al., 2019; Zhou et al., 2023), can be explored to improve the generation of realistic and diverse shape variations. By introducing adversarial training, shape models can better capture the complex and subtle interplay between shape variations and improve the generalization capability.
- We can explore hybrid approaches that combine improved-PCA with other segmentation techniques, such as deep learning or graph-based methods, to better handle complex anatomical structures.
- Another way is to integrate multiple sources of prior knowledge, such as anatomical atlases or probabilistic models, to improve segmentation accuracy.
- AS software is useful for developing MIA algorithms. For future development, expanding the software's functions, upgrading the software, and maintaining the software community are essential considerations.

6 SUMMARY OF THE DISSERTATION

This thesis first describes two different approaches to construct multi-resolution multi-organ models for MIA. The first approach is an improved PCA-based modeling method that aims to build a 3D SSM with multiple levels of resolution. This method also addresses the HDLSS problem that is commonly encountered in PDMs. The results of this approach show that the proposed method has higher modeling accuracy compared to the traditional PCA-based approach.

The second approach described in the thesis is the use of a simple but efficient SAE neural network to construct deformable models. This approach aims to mine more non-linear deformation modes. It demonstrates stronger modeling ability compared to the traditional PCA-based approach in terms of generating deformation modes.

The thesis also presents a method for simultaneously segmenting multiple organs in low-contrast mouse abdominal images using the constructed multi-resolution multi-organ SSM. In this method, the SSM acts as shape prior knowledge in the image segmentation process. The method uses human interaction to guide SSM deformation. It obtains more accurate segmentation results compared to the global resolution SSM approach. The proposed method is also robust and not affected by different user operations.

The last part of this thesis describes the development of AS, a software that accelerates the creation of MIA algorithms. The method for building the multi-resolution multi-organ SSM has been integrated into this software as a semi-automated organ segmentation function module.

In summary, this paper offers valuable contributions to the efficient modeling and segmentation of multiple organs in medical images. Additionally, it provides researchers with a user-friendly, open-source software tool called AS, which can aid in the development of MIA algorithms. The potential benefits of these contributions are promising for future clinical diagnoses.

YHTEENVETO (SUMMARY IN FINNISH)

Tämä väitöskirja esittelee kaksi lähestymistapaa moniresoluutioisten monielinmallien luomiseen lääketieteellisen kuva-analyysin tarpeisiin. Ensimmäinen lähestymistapa on pääkomponenttianalyysista jatkokehitelty mallinnusmenetelmä, jonka tavoite on resoluutioltaan monitasoinen kolmiulotteinen tilastollinen muotomalli. Menetelmä on vastaus pistejakomallissa yleisesti ilmenevään näytteistyksen harvuuteen suurissa dimensioissa. Tulokset osoittavat, että ehdotettu menetelmä mahdollistaa paremman mallinnustarkkuuden perinteisiin pääkomponenttianalyysiin pohjautuviin menetelmiin nähden. Toinen lähestymistapa käyttää yksinkertaista mutta tehokasta pinottua autoenkooderi hermo- eli neuroverkkoa muuttuvien mallien rakentamiseen. Tavoitteena on mahdollistaa laajempia epälineaarisia muodonmuutoksia ja näitä vastaavia moodeja. Tulokset muodonmuutoksen moodeista osoittavat mallinnuskyvyn paranevan pinottuun autoenkooderiin perustuvaan menetelmään verrattuna.

Väitöskirja esittelee myös lähestymistavan useiden elinten samanaikaiseen segmentointiin kontrastiltaan heikkojen hiiren vatsakuvista työssä kehitettyä moniresoluutioisen monielinten tilastollisen muotomallin avulla. Tilastollista muotomallia käytetään erityisesti ennalta tunnetun tiedon hyödyntämiseen kuvien segmentoinnissa käyttäjän vuorovaikutusen tukemana. Menetelmällä saavutetaan tarkempia segmentointeja globaalin resoluution tilastollisen muotomallin verrattuna. Ehdotettu menetelmä on myös robusti eikä se ole altis käyttäjän toimenpiteille. Väitöskirjan viimeinen osa esittelee AnatomySketch ohjelmiston kehitystä, jota käytetään lääketieteellisten kuvantamisen analyysi algoritmien kehittämiseen. Moniresoluutioinen monielinten tilastollinen muotomalli on integroitu ohjelmistoon puoliautomatisoituna segmentoinnin funktiomoduulina.

Väitöskirjan voi yhteenvetona todeta esittelevän uusia tehokkaita menetelmiä monielinmallinnukseen ja -segmentointiin lääketieteellisen kuvantamisen tarpeisiin. Lisäksi työ tarjoaa avoimeen lähdekoodiin perustuvan käyttäjäystävällisen ohjelmistotyökalun kuvantamisen algoritmien kehittämiseen tulevaisuuden kliinisen diagnostiikan tarpeisiin.

REFERENCES

Abramovitch, K., and Rice, D. D. (2014). Basic principles of cone beam computed tomography. *Dental Clinics,* 58(3): 463-484.

Adams, R., and Bischof, L. (1994). Seeded region growing. *IEEE Transactions on Pattern Analysis Machine Intelligence,* 16(6): 641-647.

Adollah, R., Francis, E., Mashor, M., and Harun, N. (2012). Bone marrow image segmentation based on multilevel thresholding. In *2012 International Conference on Biomedical Engineering (ICoBE)*, pages 457-461. IEEE.

Ahmed, A. F. (2017). Efficient Approach for De-Speckling Medical Ultrasound Images Using Improved Adaptive Shock Filter. *Al-Nahrain Journal for Engineering Sciences,* 20(5): 1192-1197.

Alomari, Z., Harput, S., Hyder, S., and Freear, S. (2015). The effect of the transducer parameters on spatial resolution in plane-wave imaging. In *2015 IEEE International Ultrasonics Symposium (IUS)*, pages 1-4. IEEE.

Altunbas, M. C., Shaw, C., Chen, L., Wang, T., and Tu, S. (2006). Effects of scattered radiation and beam quality on low contrast performance in cone beam breast CT. In *Medical Imaging 2006: Physics of Medical Imaging*, pages 1078-1087. SPIE.

Andonie, R. (2019). Hyperparameter optimization in learning systems. *Journal of Membrane Computing,* 1(4): 279-291.

Aoshima, M., Shen, D., Shen, H., Yata, K., Zhou, Y. H., and Marron, J. S. (2018). A survey of high dimension low sample size asymptotics. *Australian New Zealand Journal of Statistics,* 60(1): 4-19.

Audenaert, E. A., Van Houcke, J., Almeida, D. F., Paelinck, L., Peiffer, M., Steenackers, G., et al. (2019). Cascaded statistical shape model based segmentation of the full lower limb in CT. *Computer Methods in Biomechanics Biomedical Engineering, 22*(6): 644-657.

Avants, B. B., Tustison, N., and Song, G. (2009). Advanced normalization tools (ANTS). *Insight Journal, 2*(365): 1-35.

Ayman, A., Funatomi, T., Minoh, M., ELNOMERY, Z., OKADA, T., TOGASHI, K., et al. (2010). New region growing segmentation technique for MR images with weak boundaries (医用画像). *電子情報通信学会技術研究報告. MI, 医用画像,* 110(280): 71-76.

Ayman, A., Mabrouk, E., and Elnomery, Z. (2013). Adaptation of region growing thresholds using memetic programming algorithm. In *2013 IEEE/ACIS 12th International Conference on Computer and Information Science (ICIS)*, pages 29-34. IEEE.

Baer, M., and Kachelrieß, M. (2012). Hybrid scatter correction for CT imaging. *Physics in Medicine Biology,* 57(21): 6849.

Baiker, M., Milles, J., Dijkstra, J., Henning, T. D., Weber, A. W., Que, I., et al. (2010). Atlas-based whole-body segmentation of mice from low-contrast Micro-CT data. *Medical Image Analysis,* 14(6): 723-737.

Baldi, P., and Hornik, K. (1989). Neural networks and principal component analysis: Learning from examples without local minima. *Neural Networks, 2*(1): 53-58.

Bank, D., Koenigstein, N., and Giryes, R. (2020). Autoencoders. *arXiv preprint arXiv:.05991.*

Bertsekas, D. P., and Tsitsiklis, J. N. (2000). Gradient convergence in gradient methods with errors. 10(3): 627-642.

Bi, L., Feng, D., and Kim, J. (2018). Dual-path adversarial learning for fully convolutional network (FCN)-based medical image segmentation. *The Visual Computer,* 34(6): 1043-1052.

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., et al. (2021). Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *Wiley Interdisciplinary Reviews: Data Mining Knowledge Discovery*: e1484.

Boas, F. E., and Fleischmann, D. (2012). CT artifacts: causes and reduction techniques. *Imaging Med,* 4(2): 229-240.

Boone, J. M. (2000). X-ray production, interaction, and detection in diagnostic imaging. *Handbook of Medical Imaging,* 1: 1-78.

Bosch, J. G., Mitchell, S. C., Lelieveldt, B. P., Nijland, F., Kamp, O., Sonka, M., et al. (2002). Automatic segmentation of echocardiographic sequences by active appearance motion models. *IEEE Transactions on Medical Imaging,* 21(11): 1374-1383.

Brazil, E. V., Macedo, I., Sousa, M. C., de Figueiredo, L. H., and Velho, L. (2010). Sketching variational hermite-rbf implicits. In *Proceedings of the Seventh Sketch-Based Interfaces and Modeling Symposium*, pages 1-8.

Brodoefel, H., Burgstahler, C., Tsiflikas, I., Reimann, A., Schroeder, S., Claussen, C. D., et al. (2008). Dual-source CT: effect of heart rate, heart rate variability, and calcification on image quality and diagnostic accuracy. *Radiology,* 247(2): 346-355.

Bryda, E. C. (2013). The Mighty Mouse: the impact of rodents on advances in biomedical research. *Missouri Medicine,* 110(3): 207.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis Machine Intelligence,* (6): 679-698.

Carovac, A., Smajlovic, F., and Junuzovic, D. (2011). Application of ultrasound in medicine. *Acta Informatica Medica,* 19(3): 168.

Caselles, V., Catté, F., Coll, T., and Dibos, F. (1993). A geometric model for active contours in image processing. *Numerische Mathematik,* 66(1): 1-31.

Cates, J., Fletcher, P. T., Styner, M., Hazlett, H. C., and Whitaker, R. (2008). Particle-based shape analysis of multi-object complexes. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 477-485. Springer.

Cerrolaza, J. J., Picazo, M. L., Humbert, L., Sato, Y., Rueckert, D., Ballester, M. Á. G., et al. (2019). Computational anatomy for multi-organ analysis in medical imaging: A review. *Medical Image Analysis,* 56: 44-67.

Cerrolaza, J. J., Villanueva, A., and Cabeza, R. (2011). Multi-shape-hierarchical active shape models. In *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, pages 1. Citeseer.

Chaney, E., Pizer, S., Joshi, S., Broadhurst, R., Fletcher, T., Gash, G., et al. (2004). Automatic male pelvis segmentation from CT images via statistically trained multi-object deformable m-rep models. *International Journal of Radiation Oncology, Biology, Physics,* 60(1): S153-S154.

Chen, C., Dou, Q., Chen, H., Qin, J., and Heng, P. A. (2020). Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. *IEEE Transactions on Medical Imaging,* 39(7): 2494-2505.

Chen, J., and Frey, E. C. (2020). Medical image segmentation via unsupervised convolutional neural network. *arXiv preprint arXiv:.10155*.

Chen, M., Xu, Z., Weinberger, K., and Sha, F. (2012). Marginalized denoising autoencoders for domain adaptation. *arXiv preprint arXiv:1206.4683*.

Chen, S., Bortsova, G., García-Uceda Juárez, A., Van Tulder, G., and De Bruijne, M. (2019). Multi-task attention-based semi-supervised learning for medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22*, pages 457-465. Springer.

Chen, S., Zhong, L., Qiu, C., Zhang, Z., and Zhang, X. (2023). Transformer-based multilevel region and edge aggregation network for magnetic resonance image segmentation. *Computers in Biology and Medicine, 152*: 106427.

Chen, Z., Ristaniemi, T., Cong, F., and Wang, H. (2020). Multi-resolution statistical shape models for multi-organ shape modelling. In *Advances in Neural Networks–ISNN 2020: 17th International Symposium on Neural Networks, ISNN 2020, Cairo, Egypt, December 4–6, 2020, Proceedings 17*, pages 74-84. Springer.

Chen, Z., Wang, H., Cong, F., and Kettunen, L. (2022). Construction of Multi-resolution Multi-organ Shape Model Based on Stacked Autoencoder Neural Network. In *2022 14th International Conference on Advanced Computational Intelligence (ICACI)*, pages 62-67. IEEE.

Cherri, A. K., and Karim, M. A. (1989). Optical symbolic substitution: edge detection using Prewitt, Sobel, and Roberts operators. *Applied Optics,* 28(21): 4644-4648.

Cierniak, R., and Cierniak, R. (2011). Some words about the history of computed tomography. *X-ray Computed Tomography in Biomedical Engineering*, 7-19.

Cirillo, M. D., Abramian, D., and Eklund, A. (2021). Vox2Vox: 3D-GAN for brain tumour segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I 6*, pages 274-284. Springer.

Clark, D., and Badea, C. (2021). Advances in micro-CT imaging of small animals. *Physica Medica,* 88: 175-192.

Coleman, G. B., and Andrews, H. C. (1979). Image segmentation by clustering. *Proceedings of the IEEE,* 67(5): 773-785.

Cootes, T. F., Edwards, G. J., and Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis Machine Intelligence,* 23(6): 681-685.

Cootes, T. F., Hill, A., Taylor, C. J., and Haslam, J. (1994). Use of active shape models for locating structures in medical images. *Image Vision Computing,* 12(6): 355-365.

Cootes, T. F., and Taylor, C. J. (1995). Combining point distribution models with shape models based on finite element analysis. *Image Vision Computing,* 13(5): 403-409.

Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active shape models-their training and application. *Computer Vision Image Understanding,* 61(1): 38-59.

Corot, C., Robert, P., Idée, J.-M., and Port, M. (2006). Recent advances in iron oxide nanocrystal technology for medical imaging. *Advanced Drug Delivery Reviews,* 58(14): 1471-1504.

Cravotto, G., and Cintas, P. (2009). Molecular self-assembly and patterning induced by sound waves. The case of gelation. *Chemical Society Reviews,* 38(9): 2684-2697.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., and Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE Signal Processing Magazine,* 35(1): 53-65.

Davatzikos, C., Tao, X., and Shen, D. (2003). Hierarchical active shape models, using the wavelet transform. *IEEE Transactions on Medical Imaging,* 22(3): 414-423.

Davies, R. H., Twining, C. J., Cootes, T. F., and Taylor, C. J. (2009). Building 3-D statistical shape models by direct optimization. *IEEE Transactions on Medical Imaging,* 29(4): 961-981.

Davis, D. L., and Vachhani, P. (2017). Computed Tomography Artifacts. *Pitfalls in Musculoskeletal Radiology*, 45-59.

Dehariya, V. K., Shrivastava, S. K., and Jain, R. (2010). Clustering of image data set using k-means and fuzzy k-means algorithms. In *2010 International Conference on Computational Intelligence and Communication Networks*, pages 386-391. IEEE.

del Fresno, M., Vénere, M., and Clausse, A. (2009). A combined region growing and deformable model method for extraction of closed surfaces in 3D CT and MRI scans. *Computerized Medical Imaging Graphics,* 33(5): 369-376.

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

Dong, Q., Luo, G., Haynor, D., O'Reilly, M., Linnau, K., Yaniv, Z., et al. (2020). DicomAnnotator: a configurable open-source software program for efficient DICOM image annotation. *Journal of Digital Imaging,* 33: 1514-1526.

Du, G., Cao, X., Liang, J., Chen, X., and Zhan, Y. (2020). Medical image segmentation based on u-net: A review. *Journal of Imaging Science Technology,* 64: 1-12.

Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, 32-57.

Egmentation, S., Pham, D., Xu, C., and Prince, J. (2000). Current Methods in Medical Image Segmentation. *Annual Review of Biomedical Engineering,* 2(1): 315-337.

Elnakib, A., Gimel'farb, G., Suri, J. S., and El-Baz, A. (2011). Medical image segmentation: a brief survey. *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies*: Volume II, 1-39.

Fan, S.-K. S., Lin, Y., and Wu, C.-C. (2008). Image thresholding using a novel estimation method in generalized Gaussian distribution mixture modeling. *Neurocomputing,* 72(1-3): 500-512.

Farag, A. A. (1992). Edge‐based image segmentation. *Remote Sensing Reviews,* 6(1): 95-121.

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., et al. (2012). 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic Resonance Imaging,* 30(9): 1323-1341.

Festing, S., and Wilkinson, R. (2007). The ethics of animal research: talking point on the use of animals in scientific research. *EMBO Reports,* 8(6): 526-530.

Feurer, M., and Hutter, F. (2019). Hyperparameter optimization. *Automated Machine Learning: Methods, Systems, Challenges*, 3-33.

Fischbach, F., Knollmann, F., Griesshaber, V., Freund, T., Akkol, E., and Felix, R. (2003). Detection of pulmonary nodules by multislice computed tomography:

improved detection rate with reduced slice thickness. *European Radiology,* 13: 2378-2383.

Frangi, A. F., Rueckert, D., Schnabel, J. A., and Niessen, W. J. (2002). Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modeling. *IEEE Transactions on Medical Imaging,* 21(9): 1151-1166.

Fu, Y., Lei, Y., Wang, T., Curran, W. J., Liu, T., and Yang, X. (2021). A review of deep learning based methods for medical image multi-organ segmentation. *Physica Medica,* 85: 107-122.

Gerig, G., Styner, M., Jones, D., Weinberger, D., and Lieberman, J. (2001). Shape analysis of brain ventricles using spharm. In *Proceedings IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 2001)*, pages 171-178. IEEE.

Ghojogh, B., Ghodsi, A., Karray, F., and Crowley, M. (2021). Factor analysis, probabilistic principal component analysis, variational inference, and variational autoencoder: Tutorial and survey. *arXiv preprint arXiv:2101.00734*.

Gibson, E., Giganti, F., Hu, Y., Bonmati, E., Bandula, S., Gurusamy, K., et al. (2018). Automatic multi-organ segmentation on abdominal CT with dense V-networks. *IEEE Transactions on Medical Imaging,* 37(8): 1822-1834.

Girard, R., Zeineddine, H. A., Orsbon, C., Tan, H., Moore, T., Hobson, N., et al. (2016). Micro-computed tomography in murine models of cerebral cavernous malformations as a paradigm for brain disease. *Journal of Neuroscience Methods,* 271: 14-24.

Goo, H. W., and Allmendinger, T. (2017). Combined electrocardiography-and respiratory-triggered CT of the lung to reduce respiratory misregistration artifacts between imaging slabs in free-breathing children: initial experience. *Korean Journal of Radiology,* 18(5): 860-866.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2020). Generative adversarial networks. *Communications of the ACM,* 63(11): 139-144.

Grau, V., Mewes, A., Alcaniz, M., Kikinis, R., and Warfield, S. K. (2004). Improved watershed transform for medical image segmentation using prior information. *IEEE Transactions on Medical Imaging,* 23(4): 447-458.

Grosgeorge, D., Petitjean, C., Dacher, J.-N., and Ruan, S. (2013). Graph cut segmentation with a statistical shape model in cardiac MRI. *Computer Vision Image Understanding,* 117(9): 1027-1035.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., et al. (2018). Recent advances in convolutional neural networks. *Pattern Recognition,* 77: 354-377.

Hall, P., Marron, J. S., and Neeman, A. (2005). Geometric representation of high dimension, low sample size data. *Journal of the Royal Statistical Society: Series B,* 67(3): 427-444.

Hamarneh, G., and Li, X. (2009). Watershed segmentation using prior shape and appearance knowledge. *Image Vision Computing,* 27(1-2): 59-68.

Han, X., Hu, Z., Wang, S., and Zhang, Y. (2022). A Survey on Deep Learning in COVID-19 Diagnosis. *Journal of Imaging,* 9(1): 1.

Haralick, R. M., and Shapiro, L. G. (1985). Image segmentation techniques. *Computer vision, Graphics, Image processing,* 29(1): 100-132.

Hasebroock, K. M., and Serkova, N. J. (2009). Toxicity of MRI and CT contrast agents. *Expert Opinion on Drug Metabolism and Toxicology,* 5(4): 403-416.

Heimann, T., and Meinzer, H.-P. (2009). Statistical shape models for 3D medical image segmentation: a review. *Medical Image Analysis,* 13(4): 543-563.

Heinonen, T., Dastidar, P., Eskola, H., Frey, H., Ryymin, P., and Laasonen, E. (1998). Applicability of semi-automatic segmentation for volumetric analysis of brain lesions. *Journal of Medical Engineering Technology,* 22(4): 173-178.

Herman, G. T. (2009). Fundamentals of computerized tomography: image reconstruction from projections. *Springer Science & Business Media*.

Herman, S. (2004). Computed tomography contrast enhancement principles and the use of high-concentration contrast media. *Journal of Computer Assisted Tomography,* 28: S7-S11.

Heskamp, S., Hobo, W., Molkenboer-Kuenen, J. D., Olive, D., Oyen, W. J., Dolstra, H., et al. (2015). Noninvasive Imaging of Tumor PD-L1 Expression Using Radiolabeled Anti–PD-L1 AntibodiesPD-L1 Imaging. *Cancer Research,* 75(14): 2928-2936.

Hill, A., and Taylor, C. J. (1994). Automatic Landmark Generation for Point Distribution Models. In *BMVC*, vol. 2, pages 429-438.

Hill, A., Thornham, A., and Taylor, C. J. (1993). Model-based interpretation of 3d medical images. In *BMVC*, vol. 93, pages 1-10.

Hinton, G. E., and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science,* 313(5786): 504-507.

Holdsworth, D. W., and Thornton, M. M. (2002). Micro-CT in small animal and specimen imaging. *Trends in Biotechnology,* 20(8): S34-S39.

Hoskins, P. R., Martin, K., and Thrush, A. (2019). Diagnostic ultrasound: physics and equipment. CRC Press.

Hsieh, J. (2003). Computed tomography: principles, design, artifacts, and recent advances.

Hsieh, J., Chao, E., Thibault, J., Grekowicz, B., Horst, A., McOlash, S., et al. (2004). A novel reconstruction algorithm to extend the CT scan field ‐ of ‐ view. *Medical Physics,* 31(9): 2385-2391.

Hu, G. (2009). Survey of recent volumetric medical image segmentation techniques. In *Biomedical Engineering*: IntechOpen.

Hu, J., Haworth, S. T., Molthen, R. C., and Dawson, C. A. (2004). Dynamic small animal lung imaging via a postacquisition respiratory gating technique using micro-cone beam computed tomography1. *Academic Radiology,* 11(9): 961-970.

Hu, P., Wu, F., Peng, J., Bao, Y., Chen, F., and Kong, D. (2017). Automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets. *International Journal of Computer Assisted Radiology Surgery,* 12: 399-411.

Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., et al. (2020). Unet 3+: A full-scale connected unet for medical image segmentation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1055-1059. IEEE.

Huang, H., Makedon, F., and McColl, R. (2008). High dimensional statistical shape model for medical image analysis. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1541-1544. IEEE.

Huang, H., Zheng, H., Lin, L., Cai, M., Hu, H., Zhang, Q., et al. (2021). Medical image segmentation with deep atlas prior. *IEEE Transactions on Medical Imaging,* 40(12): 3519-3530.

Huang, Y., Gao, L., Preuhs, A., and Maier, A. (2020). Field of view extension in computed tomography using deep learning prior. In *Bildverarbeitung für die Medizin 2020: Algorithmen–Systeme–Anwendungen. Proceedings des Workshops vom 15. bis 17. März 2020 in Berlin*, pages 186-191. Springer.

Huang, Z.-K., and Chau, K.-W. (2008). A new image thresholding method based on Gaussian mixture model. *Applied Mathematics Computation,* 205(2): 899-907.

Hussain, S., Mubeen, I., Ullah, N., Shah, S. S. U. D., Khan, B. A., Zahoor, M., et al. (2022). Modern Diagnostic Imaging Technique Applications and Risk Factors in the Medical Field: A Review. *BioMed Research International,* 2022.

Jain, R., and Sharma, R. S. (2019). Image segmentation through fuzzy clustering: A survey. In *Harmony Search and Nature Inspired Optimization Algorithms*, pages 497-508. Springer.

Japkowicz, N., Hanson, S. J., and Gluck, M. A. (2000). Nonlinear autoassociation is not equivalent to PCA. *Neural Computation,* 12(3): 531-545.

Kagadis, G. C., Loudos, G., Katsanos, K., Langer, S. G., and Nikiforidis, G. C. (2010). In vivo small animal imaging: current status and future prospects. *Medical Physics,* 37(12): 6421-6442.

Kak, A. C., and Slaney, M. (2001). Principles of computerized tomographic imaging. *Society for Industrial and Applied Mathematics.*

Kalender, W. A., Seissler, W., Klotz, E., and Vock, P. (1990). Spiral volumetric CT with single-breath-hold technique, continuous transport, and continuous scanner rotation. *Radiology,* 176(1): 181-183.

Kang, W.-X., Yang, Q.-Q., and Liang, R.-P. (2009). The comparative research on image segmentation algorithms. In *2009 First International Workshop on Education Technology and Computer Science*, pages 703-707. IEEE.

Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision,* 1(4): 321-331.

Katti, G., Ara, S. A., and Shireen, A. (2011). Magnetic resonance imaging (MRI)–A review. *International Journal of Dental Clinics,* 3(1): 65-70.

Kaur, D., and Kaur, Y. (2014). Various image segmentation techniques: a review. *International Journal of Computer Science Mobile Computing,* 3(5): 809-814.

Kelkar, D., and Gupta, S. (2008). Improved quadtree method for split merge image segmentation. In *2008 First International Conference on Emerging Trends in Engineering and Technology*, pages 44-47. IEEE.

Kim, J., Fisher, J. W., Yezzi, A., Çetin, M., and Willsky, A. S. (2005). A nonparametric statistical method for image segmentation using information theory and curve evolution. *IEEE Transactions on Image Processing,* 14(10): 1486-1502.

Klein, S., Staring, M., Murphy, K., Viergever, M. A., and Pluim, J. P. (2009). Elastix: a toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging,* 29(1): 196-205.

Lacalle, D., Castro-Abril, H. A., Randelovic, T., Domínguez, C., Heras, J., Mata, E., et al. (2021). Spheroidj: An open-source set of tools for spheroid segmentation. *Computer Methods Programs in Biomedicine,* 200: 105837.

Lamecker, H., Lange, T., and Seebass, M. (2004). Segmentation of the liver using a 3D statistical shape model.

Le Bihan, D., and Johansen-Berg, H. (2012). Diffusion MRI at 25: exploring brain tissue structure and function. *Neuroimage,* 61(2): 324-341.

Leal-Taixé, L., Milan, A., Schindler, K., Cremers, D., Reid, I., and Roth, S. (2017). Tracking the trackers: an analysis of the state of the art in multiple object tracking. *arXiv preprint arXiv:1704.02781.*

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature,* 521(7553): 436-444.

Lee, E., Hui, E. S. K., Chan, K. K. L., Tse, K. Y., Kwong, W. K., Chang, T. Y., et al. (2015). Relationship between intravoxel incoherent motion diffusion - weighted MRI and dynamic contrast - enhanced MRI in tissue perfusion of cervical cancers. *Journal of Magnetic Resonance Imaging,* 42(2): 454-459.

Lee, L. K., Liew, S. C., and Thong, W. J. (2015). A review of image segmentation methodologies in medical image. *In Advanced Computer and Communication Engineering Technology: Proceedings of the 1st International Conference on Communication and Computer Engineering,* pages 1069-1080. Springer.

Lee, N., Choi, S. H., and Hyeon, T. (2013). Nano - sized CT contrast agents. *Advanced Materials,* 25(19): 2641-2660.

Lei, T., Sun, R., Du, X., Fu, H., Zhang, C., and Nandi, A. K. (2023). SGU-Net: Shape-Guided Ultralight Network for Abdominal Image Segmentation. *IEEE Journal of Biomedical Health Informatics,* 27(3): 1431-1442.

Lempitsky, V. (2019). Autoencoder. *Computer Vision: A Reference Guide*, 1-6.

Leng, S., Diehn, F., Lane, J., Koeller, K., Witte, R., Carter, R., et al. (2015). Temporal bone CT: improved image quality and potential for decreased radiation dose using an ultra-high-resolution scan mode with an iterative reconstruction algorithm. *American Journal of Neuroradiology,* 36(9): 1599-1603.

Leventon, M. E., Grimson, W. E. L., and Faugeras, O. (2002). Statistical shape influence in geodesic active contours. In *5th IEEE EMBS International Summer School on Biomedical Imaging, 2002*, pages 8. IEEE.

Li, L., Gong, R., and Chen, W. (1997). Gray level image thresholding based on fisher linear projection of two-dimensional histogram. *Pattern Recognition,* 30(5): 743-749.

Litany, O., Bronstein, A., Bronstein, M., and Makadia, A. (2018). Deformable shape completion with graph convolutional autoencoders. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1886-1895. IEEE.

Liu, X., Yang, L., Chen, J., Yu, S., and Li, K. (2022). Region-to-boundary deep learning model with multi-scale feature fusion for medical image segmentation. *Biomedical Signal Processing Control,* 71: 103165.

Lorensen, W. E., and Cline, H. E. (1987). Marching cubes: A high resolution 3D surface construction algorithm. *ACM Siggraph Computer Graphics,* 21(4): 163-169.

Lu, C., Pizer, S. M., Joshi, S., and Jeong, J.-Y. (2007). Statistical multi-object shape models. *International Journal of Computer Vision, 75*: 387-404.

Lu, L., Zhang, X., Zhao, Y., and Jia, Y. (2006). Ear recognition based on statistical shape model. In *First International Conference on Innovative Computing, Information and Control-Volume I (ICICIC'06)*, pages 353-356. IEEE.

Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., and Kim, T.-K. (2021). Multiple object tracking: A literature review. *Artificial Intelligence,* 293: 103448.

Maier, A., Steidl, S., Christlein, V., and Hornegger, J. (2018). Medical imaging systems: An introductory guide.

Maitra, M., and Chatterjee, A. (2008). A novel technique for multilevel optimal magnetic resonance brain image thresholding using bacterial foraging. *Measurement,* 41(10): 1124-1134.

Malik, J., Belongie, S., Leung, T., and Shi, J. (2001). Contour and texture analysis for image segmentation. *International Journal of Computer Vision,* 43(1): 7-27.

Marani, R., Reno, V., Nitti, M., D'Orazio, T., and Stella, E. (2016). A modified iterative closest point algorithm for 3D point cloud registration. *Computer - Aided Civil Infrastructure Engineering,* 31(7): 515-534.

Markl, M., Frydrychowicz, A., Kozerke, S., Hope, M., and Wieben, O. (2012). 4D flow MRI. *Journal of Magnetic Resonance Imaging,* 36(5): 1015-1036.

McCormick, M., Liu, X., Jomier, J., Marion, C., and Ibanez, L. (2014). ITK: enabling reproducible research and open science. *Frontiers in Neuroinformatics,* 8(2014): 13.

McInerney, T., and Terzopoulos, D. (1996). Deformable models in medical image analysis: a survey. *Medical Image Analysis,* 1(2): 91-108.

Mezzanotte, L., van't Root, M., Karatas, H., Goun, E. A., and Löwik, C. W. (2017). In vivo molecular bioluminescence imaging: new tools and applications. *Trends in Biotechnology,* 35(7): 640-652.

Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565-571. IEEE.

Miwa, S., and Otsuka, T. (2017). Practical use of imaging technique for management of bone and soft tissue tumors. *Journal of Orthopaedic Science,* 22(3): 391-400.

Mlynárik, V. (2017). Introduction to nuclear magnetic resonance. *Analytical Biochemistry,* 529 (2017): 4-9.

Moreau, J. T., Saint-Martin, C., Baillet, S., and Dudley, R. W. (2021). MNI SISCOM: an open-source tool for computing subtraction ictal single-photon emission CT coregistered to MRI. *Journal of Digital Imaging,* 34(2021): 357-361.

Natterer, F. (2001). The mathematics of computerized tomography. Society for Industrial and Applied Mathematics.

Natterer, F., and Natterer, F. (1986). Computerized tomography. *The Mathematics of Computerized Tomography*, 1-8.

Nema, S., Dudhane, A., Murala, S., and Naidu, S. (2020). RescueNet: An unpaired GAN for brain tumor segmentation. *Biomedical Signal Processing Control,* 55: 101641.

Nie, D., Gao, Y., Wang, L., and Shen, D. (2018). ASDNet: attention based semi-supervised deep networks for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 370-378. Springer.

Nolte, D., Tsang, C. K., Zhang, K. Y., Ding, Z., Kedgley, A. E., and Bull, A. M. (2016). Non-linear scaling of a musculoskeletal model of the lower limb using statistical shape models. *Journal of Biomechanics,* 49(14): 3576-3581.

Norouzi, A., Rahim, M. S. M., Altameem, A., Saba, T., Rad, A. E., Rehman, A., et al. (2014). Medical image segmentation methods, algorithms, and applications. *IETE Technical Review,* 31(3): 199-213.

Nosrati, M. S., and Hamarneh, G. (2016). Incorporating prior knowledge in medical image segmentation: a survey. *arXiv preprint arXiv:1607.01092*.

O'Shea, K., and Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.

Olabarriaga, S. D., and Smeulders, A. W. (2001). Interaction in the segmentation of medical images: A survey. *Medical Image Analysis,* 5(2): 127-142.

Ortiz, S. H. C., Chiu, T., and Fox, M. D. (2012). Ultrasound image enhancement: A review. *Biomedical Signal Processing Control,* 7(5): 419-428.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, Cybernetics,* 9(1): 62-66.

Panchbhai, A. S. (2015). Wilhelm Conrad Röntgen and the discovery of X-rays: Revisited after centennial. *Journal of Indian Academy of Oral Medicine Radiology,* 27(1): 90-95.

Park, S. I., and Lim, S. J. (2014). Template ‐ Based Reconstruction of Surface Mesh Animation from Point Cloud Animation. *ETRI Journal,* 36(6): 1008-1015.

Patey, S. J., and Corcoran, J. P. (2021). Physics of ultrasound. *Anaesthesia Intensive Care Medicine,* 22(1): 58-63.

Pauwels, E., Van Loo, D., Cornillie, P., Brabant, L., and Van Hoorebeke, L. (2013). An exploratory study of contrast agents for soft tissue visualization by means of high resolution X ‐ ray computed tomography imaging. *Journal of Microscopy,* 250(1): 21-31.

Pelc, N. J. (2014). Recent and future directions in CT imaging. *Annals of Biomedical Engineering,* 42: 260-268.

Peng, S., Jiang, W., Pi, H., Li, X., Bao, H., and Zhou, X. (2020). Deep snake for real-time instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8533-8542. IEEE.

Pham, D. L., Xu, C., and Prince, J. L. (2000). A survey of current methods in medical image segmentation. *Annual Review of Biomedical Engineering,* 2(3): 315-337.

Philbrick, K. A., Weston, A. D., Akkus, Z., Kline, T. L., Korfiatis, P., Sakinis, T., et al. (2019). RIL-contour: a medical imaging dataset annotation tool for and with deep learning. *Journal of Digital Imaging,* 32: 571-581.

Piechowicz, J. (2011). Sound wave diffraction at the edge of a sound barrier. *Acta Physica Polonica A,* 119(6A): 1040-1045.

Pieper, S., Halle, M., and Kikinis, R. (2004). 3D Slicer. In *2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821)*, pages 632-635. IEEE.

Pinter, C., Lasso, A., Wang, A., Sharp, G. C., Alexander, K., Jaffray, D., et al. (2015). Performing radiation therapy research using the open-source SlicerRT toolkit. In *World Congress on Medical Physics and Biomedical Engineering, June 7-12, 2015, Toronto, Canada*, pages 622-625. Springer.

Pitiot, A., Bardinet, E., Thompson, P. M., and Malandain, G. (2006). Piecewise affine registration of biological images for volume reconstruction. *Medical Image Analysis,* 10(3): 465-483.

Pizer, S. M., Fletcher, P. T., Joshi, S., Thall, A., Chen, J. Z., Fridman, Y., et al. (2003). Deformable m-reps for 3D medical image segmentation. *International Journal of Computer Vision,* 55(2): 85-106.

Pizer, S. M., Fritsch, D. S., Yushkevich, P. A., Johnson, V. E., and Chaney, E. L. (1999). Segmentation, registration, and measurement of shape variation via image object shape. *IEEE Transactions on Medical Imaging,* 18(10): 851-865.

Pontana, F., Pagniez, J., Flohr, T., Faivre, J.-B., Duhamel, A., Remy, J., et al. (2011). Chest computed tomography using iterative reconstruction vs filtered back projection (Part 1): evaluation of image noise reduction in 32 patients. *European Radiology,* 21: 627-635.

Poon, K., Hamarneh, G., and Abugharbieh, R. (2007). Segmentation of complex objects with non-spherical topologies from volumetric medical images using 3D livewire. In *Medical Imaging 2007: Image Processing*, pages 1032-1041. SPIE.

Powles, A. E., Martin, D. J., Wells, I. T., and Goodwin, C. R. (2018). Physics of ultrasound. *Anaesthesia Intensive Care Medicine,* 19(4): 202-205.

Prewitt, J. M. (1970). Object enhancement and extraction. *Picture Processing Psychopictorics,* 10(1): 15-19.

Protection, I. C. o. N.-I. R. (2017). ICNIRP statement on diagnostic devices using non-ionizing radiation: existing regulations and potential health risks. *Health Physics,* 112(3): 305-321.

Ramesh, K., Kumar, G. K., Swapna, K., Datta, D., and Rajest, S. S. (2021). A review of medical image segmentation algorithms. *EAI Endorsed Transactions on Pervasive Health Technology,* 7(27): e6-e6.

Ranjan, A., Bolkart, T., Sanyal, S., and Black, M. J. (2018). Generating 3D faces using convolutional mesh autoencoders. In *Proceedings of the European Conference on Computer Vision (ECCV),* pages 704-720.

Ravindraiah, R., and Tejaswini, K. (2013). Survey of image segmentation algorithms based on expectation-maximization. *IOSR Journal of VLSI Signal Processing,* 2(6): 1-7.

Rogosnitzky, M., and Branch, S. (2016). Gadolinium-based contrast agent toxicity: a review of known and proposed mechanisms. *Biometals,* 29(3): 365-376.

Rosenthal, N., and Brown, S. (2007). The mouse ascending: perspectives for human-disease models. *Nature Cell Biology,* 9(9): 993-999.

Roth, H. R., Oda, H., Zhou, X., Shimizu, N., Yang, Y., Hayashi, Y., et al. (2018). An application of cascaded 3D fully convolutional networks for medical image segmentation. *Computerized Medical Imaging Graphics,* 66(2018): 90-99.

Roth, H. R., Shen, C., Oda, H., Sugino, T., Oda, M., Hayashi, Y., et al. (2018). A multi-scale pyramid of 3D fully convolutional networks for abdominal multi-organ segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV 11*, pages 417-425. Springer.

Rudyanto, R. D., Bastarrika, G., de Biurrun, G., Agorreta, J., Montuenga, L. M., Ortiz-de-Solorzano, C., et al. (2013). Individual nodule tracking in micro-CT images of a longitudinal lung cancer mouse model. *Medical Image Analysis,* 17(8): 1095-1105.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature,* 323(6088): 533-536.

Sahiner, B., Pezeshk, A., Hadjiiski, L. M., Wang, X., Drukker, K., Cha, K. H., et al. (2019). Deep learning in medical imaging and radiation therapy. *Medical Physics,* 46(1): e1-e36.

Salvolini, L., Secchi, E. B., Costarelli, L., and De Nicola, M. (2000). Clinical applications of 2D and 3D CT imaging of the airways—a review. *European Journal of Radiology,* 34(1): 9-25.

Sarvazyan, A. P., Urban, M. W., and Greenleaf, J. F. (2013). Acoustic waves in medical imaging and diagnostics. *Ultrasound in Medicine Biology,* 39(7): 1133-1146.

Scarfe, W. C., and Farman, A. G. (2008). What is cone-beam CT and how does it work? *Dental Clinics of North America,* 52(4): 707-730.

Schoppe, O., Pan, C., Coronel, J., Mai, H., Rong, Z., Todorov, M. I., et al. (2020). Deep learning-enabled multi-organ segmentation in whole-body mouse scans. *Nature Communications,* 11(1): 5626.

Schroeder, W., Martin, K., and Lorensen, B. (2006). The visualization toolkit, 4th edn. Kitware. New York.

Scott, I. M., Cootes, T. F., and Taylor, C. J. (2003). Improving appearance model matching using local image structure. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 258-269. Springer.

Sederberg, T. W., and Parry, S. R. (1986). Free-form deformation of solid geometric models. In *Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques*, pages 151-160.

Seeram, E. (2018). Computed tomography: a technical review. *Radiologic Technology,* 89(3): 279CT-302CT.

Seibert, J. A. (2004). X-ray imaging physics for nuclear medicine technologists. Part 1: Basic principles of x-ray production. *Journal of Nuclear Medicine Technology,* 32(3): 139-147.

Serai, S. D. (2022). Basics of magnetic resonance imaging and quantitative parameters T1, T2, T2*, T1rho and diffusion-weighted imaging. *Pediatric Radiology,* 52(2): 217-227.

Shah, N., Bansal, N., and Logani, A. (2014). Recent advances in imaging technologies in dentistry. *World Journal of Radiology,* 6(10): 794.

Shamonin, D. P., Bron, E. E., Lelieveldt, B. P., Smits, M., Klein, S., Staring, M., et al. (2014). Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. *Frontiers in Neuroinformatics,* 7: 50.

Sharpless, N. E., and DePinho, R. A. (2006). The mighty mouse: genetically engineered mouse models in cancer drug development. *Nature Reviews Drug Discovery,* 5(9): 741-754.

Shen, D., Herskovits, E. H., and Davatzikos, C. (2001). An adaptive-focus statistical shape model for segmentation and shape modeling of 3-D brain structures. *IEEE Transactions on Medical Imaging,* 20(4): 257-270.

Shen, D., Wu, G., and Suk, H.-I. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering,* 19: 221-248.

Shlens, J. (2014). A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100.*

Shriki, J. (2014). Ultrasound physics. *Critical Care Clinics,* 30(1): 1-24.

Shung, K. K. (2011). Diagnostic ultrasound: Past, present, and future. *Journal of Medical and Biological Engineering,* 31(6): 371-374.

Siddique, N., Paheding, S., Elkin, C. P., and Devabhaktuni, V. (2021). U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access, 9*: 82031-82057.

Stegmann, M. B., and Gomez, D. D. (2002). A brief introduction to statistical shape analysis. *Informatics Mathematical Modelling, Technical University of Denmark, DTU,* 15(11).

Stout, D. B., Chatziioannou, A. F., Lawson, T. P., Silverman, R. W., Gambhir, S. S., and Phelps, M. E. (2005). Small animal imaging center design: the facility at the UCLA Crump Institute for Molecular Imaging. *Molecular Imaging Biology,* 7: 393-402.

Suckow, C. E., and Stout, D. B. (2008). MicroCT liver contrast agent enhancement over time, dose, and mouse strain. *Molecular Imaging Biology,* 10: 114-120.

Sun, Y., Yuan, P., and Sun, Y. (2020). MM-GAN: 3D MRI data augmentation for medical image segmentation via generative adversarial networks. In *2020 IEEE International Conference on Knowledge Graph (ICKG)*, pages 227-234. IEEE.

Szabo, T. L. (2004). Diagnostic ultrasound imaging: inside out. Academic press.

Taha, A. A., and Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging,* 15(1): 1-28.

Takane, Y., Young, F. W., and De Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika,* 42: 7-67.

Tassone, P., Neri, P., Burger, R., T Di Martino, M., Leone, E., Amodio, N., et al. (2012). Mouse models as a translational platform for the development of new therapeutic agents in multiple myeloma. *Current Cancer Drug Targets,* 12(7): 814-822.

Tayebi, S., Momani, S., and Arqub, O. A. (2022). The cubic B-spline interpolation method for numerical point solutions of conformable boundary value problems. *Alexandria Engineering Journal,* 61(2): 1519-1528.

Terzopoulos, D., and Szeliski, R. (1992). Tracking with Kalman snakes. *Active Vision, 20*: 3-20.

Tsai, A., Yezzi, A., Wells, W., Tempany, C., Tucker, D., Fan, A., et al. (2003). A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Transactions on Medical Imaging,* 22(2): 137-154.

Van Griethuysen, J. J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., et al. (2017). Computational radiomics system to decode the radiographic phenotype. *Cancer Research,* 77(21): e104-e107.

Vardhanabhuti, V., Loader, R., and Roobottom, C. A. (2013). Assessment of image quality on effects of varying tube voltage and automatic tube current modulation with hybrid and pure iterative reconstruction techniques in abdominal/pelvic CT: a phantom study. *Investigative Radiology,* 48(3): 167-174.

Vera, S., Gil, D., Borras, A., Sánchez, X., Pérez, F., Linguraru, M. G., et al. (2011). Computation and Evaluation of Medial Surfaces for Shape Representation of Abdominal Organs. In *Abdominal Imaging*, pages 223-230. Springer.

Vijay, M., Subha, S., and Karthik, K. (2012). Adaptive spatial and wavelet multiscale products thresholding method for medical image denoising. In *2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET)*, pages 1109-1114. IEEE.

Vincent, O. R., and Folorunso, O. (2009). A descriptive algorithm for sobel image edge detection. In *Proceedings of Informing Science & IT Education Conference (InSITE)*, vol. 40, pages 97-107.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.-A., and Bottou, L. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research,* 11(12).

Vuong, Q. L., Gillis, P., Roch, A., and Gossuin, Y. (2017). Magnetic resonance relaxation induced by superparamagnetic particles used as contrast agents in magnetic resonance imaging: a theoretical review. *Wiley Interdisciplinary Reviews: Nanomedicine Nanobiotechnology,* 9(6): e1468.

Wang, H., Han, Y., Chen, Z., Hu, R., Chatziioannou, A. F., Zhang, B., et al. (2019). Prediction of major torso organs in low-contrast micro-CT images of mice using a two-stage deeply supervised fully convolutional network. *Physics in Medicine,* 64(24): 245014.

Wang, H., Stout, D. B., and Chatziioannou, A. F. (2011). Estimation of mouse organ locations through registration of a statistical mouse atlas with micro-CT images. *IEEE Transactions on Medical Imaging,* 31(1): 88-102.

Wang, H., Stout, D. B., Taschereau, R., Gu, Z., Vu, N. T., Prout, D. L., et al. (2012). MARS: a mouse atlas registration system based on a planar x-ray projector and an optical camera. *Physics in Medicine Biology,* 57(19): 6063.

Wang, J., and Fleischmann, D. J. R. (2018). Improving spatial resolution at CT: development, benefits, and pitfalls. *Radiology,* 289(1): 261-262.

Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H., and Nandi, A. K. (2022). Medical image segmentation using deep learning: A survey. *IET Image Processing,* 16(5): 1243-1267.

Wang, W., Huang, Y., Wang, Y., and Wang, L. (2014). Generalized autoencoder: A neural network framework for dimensionality reduction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 490-497.

Wang, Y., Yao, H., and Zhao, S. (2016). Auto-encoder based dimensionality reduction. *Neurocomputing,* 184: 232-242.

Wang, Y., Zhou, Y., Shen, W., Park, S., Fishman, E. K., and Yuille, A. L. (2019). Abdominal multi-organ segmentation with organ-attention networks and statistical fusion. *Medical Image Analysis, 55*: 88-102.

Warfield, S. K., Zou, K. H., and Wells, W. M. (2002). Validation of image segmentation and expert quality with an expectation-maximization algorithm. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 298-306. Springer.

Webb, A. (2022). Introduction to biomedical imaging. John Wiley & Sons.

Weissleder, R., and Ntziachristos, V. (2003). Shedding light onto live molecular targets. *Nature Medicine,* 9(1): 123-128.

Wessels, A., and Sedmera, D. (2003). Developmental anatomy of the heart: a tale of mice and man. *Physiological Genomics,* 15(3): 165-176.

Willekens, I., Lahoutte, T., Buls, N., Vanhove, C., Deklerck, R., Bossuyt, A., et al. (2009). Time-course of contrast enhancement in spleen and liver with Exia 160, Fenestra LC, and VC. *Molecular Imaging and Biology,* 11: 128-135.

Wilms, M., Ehrhardt, J., and Handels, H. (2012). A 4D statistical shape model for automated segmentation of lungs with large tumors. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012: 15th International Conference, Nice, France, October 1-5, 2012, Proceedings, Part II 15*, pages 347-354. Springer.

Wilms, M., Handels, H., and Ehrhardt, J. (2017). Multi-resolution multi-object statistical shape models based on the locality assumption. *Medical Image Analysis, 38*: 17-29.

Wilson, K., Bengtson, M., and Schaub, H. (2020). X ‐ ray Spectroscopic Determination of Electrostatic Potential and Material Composition for Spacecraft: Experimental Results. *Space Weather,* 18(4): e2019SW002342.

Wilson, K., and Schaub, H. (2019). X-ray spectroscopy for electrostatic potential and material determination of space objects. *IEEE Transactions on Plasma Science,* 47(8): 3858-3866.

Wolf, I., Vetter, M., Wegner, I., Böttger, T., Nolden, M., Schöbinger, M., et al. (2005). The medical imaging interaction toolkit. *Medical Image Analysis,* 9(6): 594-604.

Wormanns, D., Kohl, G., Klotz, E., Marheine, A., Beyer, F., Heindel, W., et al. (2004). Volumetric measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility. *European Radiology,* 14: 86-92.

Wu, J., Poehlman, S., Noseworthy, M. D., and Kamath, M. V. (2008). Texture feature based automated seeded region growing in abdominal MRI segmentation. In *2008 International Conference on BioMedical Engineering and Informatics*, pages 263-267. IEEE.

Xi, X., Shi, H., Han, L., Wang, T., Ding, H. Y., Zhang, G., et al. (2017). Breast tumor segmentation with prior knowledge learning. *Neurocomputing,* 237: 145-157.

Xiong, M., Roshanbin, S., Rokka, J., Schlein, E., Ingelsson, M., Sehlin, D., et al. (2021). In vivo imaging of synaptic density with [11C] UCB-J PET in two mouse models of neurodegenerative disease. *Neuroimage,* 239: 118302.

Xun, S., Li, D., Zhu, H., Chen, M., Wang, J., Li, J., et al. (2022). Generative adversarial networks in medical image segmentation: A review. *Computers in Biology,* 140: 105063.

Yan, D., Zhang, Z., Luo, Q., and Yang, X. (2017). A novel mouse segmentation method based on dynamic contrast enhanced micro-CT images. *PLoS One,* 12(1): e0169424.

Yang, L., and Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing,* 415: 295-316.

Yang, Z., Xu, B., Luo, W., and Chen, F. (2022). Autoencoder-based representation learning and its application in intelligent fault diagnosis: A review. *Measurement,* 189: 110460.

Yi, X., Walia, E., and Babyn, P. (2019). Generative adversarial network in medical imaging: A review. *Medical Image Analysis,* 58: 101552.

Yoo, T. S., Ackerman, M. J., Lorensen, W. E., Schroeder, W., Chalana, V., Aylward, S., et al. (2002). Engineering and algorithm design for an image processing API: a technical report on ITK-the insight toolkit. In *Medicine Meets Virtual Reality 02/10*, pages 586-592. IOS press.

Yoshizawa, S., Takagi, R., and Umemura, S.-i. (2017). Enhancement of high-intensity focused ultrasound heating by short-pulse generated cavitation. *Applied Sciences,* 7(3): 288.

Yu-Qian, Z., Wei-Hua, G., Zhen-Cheng, C., Jing-Tian, T., and Ling-Yun, L. (2006). Medical images edge detection based on mathematical morphology. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 6492-6495. IEEE.

Yushkevich, P. A., Gao, Y., and Gerig, G. (2016). ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3342-3345. IEEE.

Zaidi, H., Montandon, M.-L., and Alavi, A. (2008). The clinical role of fusion imaging using PET, CT, and MR imaging. *PET Clinics,* 3(3): 275-291.

Zanaty, E., and Asaad, A. (2013). Probabilistic region growing method for improving magnetic resonance image segmentation. *Connection Science,* 25(4): 179-196.

Zanaty, E., and Ghoniemy, S. (2016). Medical image segmentation techniques: an overview. *International Journal of Informatics Medical Data Processing,* 1(1): 16-37.

Zarb, F., Rainford, L., and McEntee, M. F. (2010). Image quality assessment tools for optimization of CT images. *Radiography,* 16(2): 147-153.

Zhang, G., Liu, Y., and Jin, X. (2020). A survey of autoencoder-based recommender systems. *Frontiers of Computer Science,* 14: 430-450.

Zhang, J., Yan, C.-H., Chui, C.-K., and Ong, S.-H. (2010). Fast segmentation of bone in CT images using 3D adaptive thresholding. *Computers in Biology Medicine,* 40(2): 231-236.

Zhang, L., Xiang, D., Jin, C., Shi, F., Yu, K., and Chen, X. (2019). OIPAV: an integrated software system for ophthalmic image processing, analysis, and visualization. *Journal of Digital Imaging,* 32: 183-197.

Zhang, S., Yao, L., and Xu, X. (2017). Autosvd++ an efficient hybrid collaborative filtering model via contractive auto-encoders. In *Proceedings of the 40th*

International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 957-960.

Zhang, Y.-D., Dong, Z., Wang, S.-H., Yu, X., Yao, X., Zhou, Q., et al. (2020). Advances in multimodal data fusion in neuroimaging: overview, challenges, and novel orientation. *Information Fusion,* 64: 149-187.

Zhao, F., and Xie, X. (2013). An overview of interactive medical image segmentation. *Annals of the BMVA,* 2013(7): 1-22.

Zhou, T., Li, Q., Lu, H., Cheng, Q., and Zhang, X. (2023). GAN review: Models and medical image fusion applications. *Information Fusion,* 91: 134-148.

Zhou, X., Takayama, R., Wang, S., Hara, T., and Fujita, H. (2017). Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method. *Medical Physics,* 44(10): 5221-5233.

Zhou, Y., Lower, E. E., Li, H., Farhey, Y., and Baughman, R. P. (2017). Clinical characteristics of patients with bone sarcoidosis. In *Seminars in Arthritis and Rheumatism*, pages 143-148. Elsevier.

Zucker, S. W. (1976). Region growing: Childhood and adolescence. *Computer Graphics Image Processing,* 5(3): 382-399.

ORIGINAL PAPERS

I

MULTI-RESOLUTION STATISTICAL SHAPE MODELS FOR MULTI-ORGAN SHAPE MODELLING

by

Zhonghua Chen, Tapani Ristaniemi, Fengyu Cong, and Hongkai Wang 2020

Advances in Neural Networks – ISNN 2020: 17th International Symposium on
Neural Networks, pp. 74-84

https://doi.org/10.1007/978-3-030-64221-1_7

Multi-Resolution Statistical Shape Model for Multi-Organ Shape Modelling

Zhonghua Chen[1, 2], Tapani Ristaniemi[2], Fengyu Cong[1, 2], and Hongkai Wang[1 *]

[1]School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China
`chenzh4693@foxmail.com, cong@dlut.edu.cn, wang.hongkai@dlut.edu.cn`
[2]Faculty of Information Technology, University of Jyväskylä, Jyväskylä 40100, Finland
`tapani.e.ristaniemi@jyu.fi`

Abstract. Statistical shape models (SSMs) are widely used in medical image segmentation. However, traditional SSM methods suffer from the High-Dimension-Low-Sample-Size (HDLSS) problem in modelling. In this work, we extend the state-of-the-art multi-resultion SSM approach from two dimension (2D) to three dimension (3D) and from single organ to multiple organs. Then we proposed a multi-resolution multi-organ 3D SSM method that uses a downsampling-and-interpolation strategy to overcome HDLSS problem. We also use an inter-surface-point distance thresholding scheme to achieve multi-resolution modelling effect. Our method is tested on the modelling of multiple mouse abdominal organs from mouse micro-CT images in three different resolution levels, including multi-organ level, single organ level and local structure level. The minimum specificity error and generalization error of this method are less than 0.3 mm, which are close to the pixel resolution of mouse CT images (0.2 mm) and better than traditional principal component analysis (PCA) method.

Keywords: Multi-resolution multi-organ SSM, PCA, HDLSS, Mouse micro-CT image, Liver, Spleen, Left kidney, Right kidney, Geodesic distance, Euclidean distance.

1 Introduction

In the last three decades, SSM approaches have been used as one of the most important methods to segment and register organs for medical image analysis [4]. The applications of SSMs includes but not limit to the following fields: 1) Medical image segmentation and registration [2, 3, 7]. 2) Clinical diagnosis and treatment [1, 11]. 3) Analysis of organ contraction [10].

Due to the complexity of medical images, SSMs of 3D organs are playing an increasingly important role in medical image segmentation. To represent 3D organ shapes, landmarks are sampled from the organ surface. However, a challenging problem of 3D SSM construction is that the number of training samples is small while the number of landmarks is large [12]. To fully capture the great variability of 3D shape, the traditional PCA modelling method needs to provide a large number of representative training samples to achieve a good modelling effect, which usually requires a lot of labor and is even impossible to complete. This problem is later called High-Dimension-Low-Sample-Size (HDLSS) problem, which leads to insufficient and inaccurate expression of the model.

In order to solve the HDLSS problem, Wilms et al. [16] proposed a multi-resolution statistical shape model with the traditional PCA method based on local distance constraints in 2017, and they used this method to construct a 2D multi-resolution SSM of human hand shape and cardiopulmonary shape. The method is an important extension of the traditional SSM method, which can be used to obtain statistical deformation models of objects at different resolution levels. In addition, it makes the resulting models achieve better generalization and specificity based on fewer training samples. However, one limitation of this method is that it takes up a lot of memory and is not suitable for cases where there are many sampling points. Therefore, Wilms et al. only modelled simple 2D shapes in their study. Unfortunately, the 3D shape vectors of multiple organs usually contain thousands or even tens of thousands of sampling points. Moreover, since multiple organ shape modelling usually requires more sampling points than single organ, this method is not applicable to multi-organ modelling as well. These drawbacks limit the application of this method to multi-organ 3D shape modelling.

In this article, we propose a solution to extend the multi-resolution SSM approach to 3D shape modelling of multiple organs with large number of surface points. Our methods combines 3D object surface downsampling with Laplace diffusion equation to construct multi-resolution multi-organ 3D SSM. We obtain deformation components in three resolution levels, which are the "multi-organ level", "single organ level" and "local structure level". The models obtained from the above three resolution levels are compared quantitatively with the traditional PCA modelling methods in terms of model generalization and specific performance, and we obtain better modelling performance than the traditional methods.

2 Materials and Methods

2.1 Description of Mouse Micro-CT Data

The multi-organ shape training samples of mouse micro-CT images are taken from the Molecular Imaging Centre of the University of California, Los Angeles [14, 15]. During the imaging process, mice are injected with liver contrast agent Fenestra LC (ART, Montreal, QC, Canada) for clear imaging of abdominal organs. The weights of the tested mice range from 15 to 30 grams, and the data are selected according to the following principles for modelling: (1) Boundaries of the abdominal organs of the mouse: livers, spleens, and kidneys are clear. (2) There are no motion artefacts in the CT images of mice. (3) There are no cases where the livers, spleens and kidneys of the mice deviate from the normal shape. Mice are imaged in a multi-mode indoor prone position that provides anesthesia and heating [13]. Although the imaging room limits the possible postures of the mice, these postures are not strictly normalized. The random body bending postures in the left, right, and backward directions are included in the data set. The imaging system is MicroCAT II Small Animal CT (Siemens Preclinical Solutions, Knoxville, TN). Equipment acquisition parameters for imaging: exposure setting 70 kVp, 500 mAs, 500 ms and 360 step rotation, 2.0 mm aluminum filter. In the image acquisition process, an improved Feld Kamp process is used to reconstruct the image so that the isotropic voxel size is 0.2 mm, the image matrix size is 256 × 256 × 496, and the pixel resolution is 0.2 mm.

In this study, 98 mouse micro-CT images are collected as training samples for model construction. Small animal imaging experts are invited to segment the 3D regions of livers, spleens, left kidneys, and right kidneys from the images, and then use the moving cube algorithm [6] to convert the segmented label maps to mesh surfaces. On this basis, one of the 98 sample surfaces is selected as the template surface, and the point cloud registration algorithm [8, 9] is used to register the template to all other training samples, so that different samples have the same number of mesh vertices, and each vertex corresponds to the same anatomical position in different samples, thus completing the preparation of all training data. Fig.1 illustrates the entire training data preparation process.

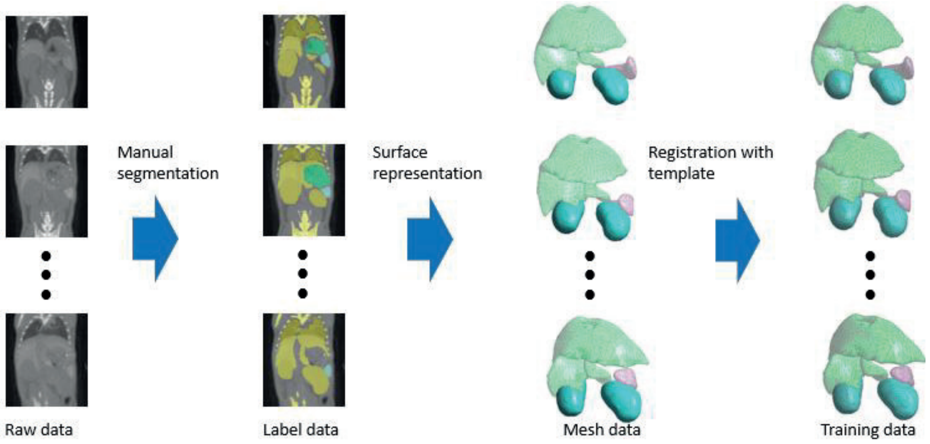


Fig. 1. The construction process for preparing the training data of the mouse multi-organ shape model

2.2 Description of Algorithms

The Construction of Multi-Resolution SSM.

1. Given N d (d = 3) dimensional mouse abdominal multi-organ training models $\{\mathbf{S_i}\}_1^N$, where $\mathbf{S_i} = \{\mathbf{X_{1,i}}, \dots, \mathbf{X_{j,i}}, \dots, \mathbf{X_{M,i}}\}$ contains M points, each point $\mathbf{X}_{j,i}=(x_{j,i}, y_{j,i}, z_{j,i})^T$ is distributed on the surface of the training model. Then calcu-

late the average model of the training models, the calculation formula is shown in formula (1):

$$\vec{\boldsymbol{\mu}} = \frac{\mathbf{1}}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathbf{S}_{\mathrm{i}} \tag{1}$$

2. After normalization, the covariance matrix of the coordinates of different dimensional points is calculated, and the calculation formula is shown in formula (2):

$$\mathrm{C} = \frac{1}{\mathrm{N}-1}\sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathbf{S}_{\mathrm{i}} - \vec{\boldsymbol{\mu}})(\mathbf{S}_{\mathrm{i}} - \vec{\boldsymbol{\mu}})^{\mathrm{T}} \tag{2}$$

3. Calculate the point-to-point geodesic distance $\mathrm{d_{geo}}(\mathbf{X_i}, \mathbf{X_j})$ on each model surface, and use rule (3) to set the values of the two sides of the covariance matrix C symmetrical to 0:

$$\boldsymbol{\rho}_{\mathrm{i,j}} = \begin{cases} \frac{\mathrm{cov}(\mathrm{X_i}, \mathrm{X_j})}{\sigma_{\mathrm{i}}\sigma_{\mathrm{j}}} & \text{if } \mathrm{d_{geo}}(\mathbf{X_i}, \mathbf{X_j}) \leq \boldsymbol{\tau} \\ 0 & \text{else} \end{cases} \tag{3}$$

where σ_{i} and σ_{j} are the standard deviations of the i-th and j-th dimensions, respectively, and $\boldsymbol{\tau}$ is the threshold given in the experiment according to the relevant rules, so that a simplified symmetric matrix $\mathrm{R_1} = \begin{pmatrix} \boldsymbol{\rho}_{1,1} & \cdots & \boldsymbol{\rho}_{1,\mathrm{M}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\rho}_{\mathrm{M},1} & \cdots & \boldsymbol{\rho}_{\mathrm{M,M}} \end{pmatrix}$.

4. Since $\mathrm{R_1}$ is not positive semi-definite and cannot be implemented eigenvalue decomposition, it is necessary to use the approximation method [5] to find an approximate positive semi-definite matrix $\mathrm{R_2}$ replacing $\mathrm{R_1}$ with formula (4):

$$\begin{gathered} \mathrm{R_2} = \min_{\mathrm{A}} \|\mathrm{A} - \mathrm{R_1}\|_{\mathrm{F}} \\ \det(\mathrm{A}) \geq 0 \\ \mathrm{diag}(\mathrm{A}) = 1 \end{gathered} \tag{4}$$

5. Calculate the eigenvector matrix U_τ of $\mathrm{R_2}$ and the corresponding eigenvalue matrix Λ_τ, as shown in formula (5):

$$\begin{pmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{\mathrm{dM}} \end{pmatrix} \mathrm{R_2} \begin{pmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{\mathrm{dM}} \end{pmatrix} = \mathrm{U}_\tau \Lambda_\tau \mathrm{U}_\tau^{\mathrm{T}} \tag{5}$$

where the eigenvector set included in U_τ is represented as $\mathbf{P}$, and the eigenvalue vector on the diagonal of Λ_τ is represented as $\vec{\lambda}$.

6. When different values of distance threshold $\boldsymbol{\tau}$ are selected in equation (3), the model will show different deformation capabilities locally; when$\boldsymbol{\tau} \geq \max_{\mathrm{i,j}} \mathrm{d_{geo}}(\vec{\mathrm{x}}_{\mathrm{i}}, \vec{\mathrm{y}}_{\mathrm{i}})$, the constructed model is the traditional SSM; when $\boldsymbol{\tau} = 0$, the physical coordinate points on all training samples lose their relevance, and the constructed model cannot be deformed, which has no practical significance. By defining a series of thresholds $\boldsymbol{\tau_1} > \boldsymbol{\tau_2} > \cdots > \boldsymbol{\tau_L}$, a multi-resolution scheme is defined to obtain a set of shape models $\{\vec{\boldsymbol{\mu}}, \mathbf{P}_1, \ldots, \mathbf{P}_{\mathrm{L}}, \vec{\lambda}_1, \ldots, \vec{\lambda}_{\mathrm{L}}\}$ that vary from global to local. However, these models are highly dependent and redundant, and do not constitute a single shape space. Therefore, it is necessary to retain global information to combine

them into a subspace, so that the feature vectors provided by the local SSM can optimally represent more local information. Based on step 1 to 5, the algorithm for constructing a multi-resolution shape model is derived as follows:

Suppose that there are N training model data matrices $\mathbf{X} = (\overrightarrow{\mathbf{s_1}} \,|\, \overrightarrow{\mathbf{s_2}} \cdots \,|\, \overrightarrow{\mathbf{s_N}}) \boldsymbol{\epsilon} \mathbf{R}^{m\times N}$, the thresholds of geodesic distance on each model surface are $\boldsymbol{\tau_1} > \boldsymbol{\tau_2} > \cdots > \boldsymbol{\tau_L}$. Calculate the average model $\vec{\boldsymbol{\mu}} = \frac{\mathbf{1}}{N}\sum_{\mathbf{i=1}}^{N}\overrightarrow{\mathbf{s_i}}$ of the training models. And define the distance matrix d_{geo} on the average model. Assuming that the iteration index r ranges from 1 to L in the calculation process, where r represents the number of models, the deformation coefficient of the local SSM is defined as $\overrightarrow{\lambda_{\tau_r}}$, and the deformation component corresponding to the coefficient is defined as $\mathbf{P}_{\tau_r}$.

When $r = 1$, it means that there is only one shape model space, and the multi-organ statistical shape model can be obtained by directly using the traditional PCA method; when $r > 1$, it means that there are multiple shape model spaces, and these model spaces need to be combined for singular value decomposition. The decomposition process is as follows:

$$\mathbf{U}(\cos\theta)\mathbf{V}^T \leftarrow \mathrm{svd}(\mathbf{P}_{MR}^T\mathbf{P}_{\tau_r}) \tag{6}$$

$$\mathbf{S} = \begin{pmatrix} \mathbf{cos\,\theta_{k\times k}} & \cdots & 0 \\ \vdots & . & \vdots \\ 0 & \cdots & \mathbf{I}_{(\mathbf{l-k})\times(\mathbf{l-k})} \end{pmatrix} \tag{7}$$

where svd() represents singular value decomposition and calculates the transform base $\widehat{\mathbf{B}}$, $\widehat{\mathbf{B}} = \mathbf{P}_{\tau_r}\mathbf{V}\mathbf{S}^T$. Then, calculate the covariance matrix after spatial transformation:

$$\widehat{\boldsymbol{\Sigma}}_{\tau_{MR}} = \left(\sigma_{i,j}^{\tau_{MR}}\right) = \mathbf{U}^T \begin{pmatrix} \lambda_{\tau_{MR},1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_{\tau_{MR},k} \end{pmatrix} \mathbf{U} \tag{8}$$

$$\widehat{\boldsymbol{\Sigma}}_{\tau_r} = \left(\sigma_{i,j}^{\tau_r}\right) = \mathbf{V}^T \begin{pmatrix} \lambda_{\tau_r,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_{\tau_r,l} \end{pmatrix} \mathbf{V} \tag{9}$$

$\widehat{\boldsymbol{\Sigma}}_{MR,r} = (\sigma_{i,j}^{\tau_{MR},\tau_r})$, where $\sigma_{i,j}^{\tau_{MR},\tau_r} = \begin{cases} \sigma_{i,j}^{\tau_{MR}} & \text{if } \ i,j \in [1,k] \\ \sigma_{i,j}^{\tau_r} & \text{if } \ i,j \in [k+1,l] \\ 0 & \text{else} \end{cases}$, finally calculate the uncorrelated basis vectors and the corresponding feature values:

$$[P_{MR}, \hat{\lambda}_{MR}] \leftarrow \mathrm{eig}(\widehat{\mathbf{B}}\widehat{\boldsymbol{\Sigma}}_{MR,r}\widehat{\mathbf{B}}^T) \tag{10}$$

where eig() represents eigenvalue decomposition. And the multi-resolution multi-organ SSM $\boldsymbol{\mu}$ is represented as follows:

$$\boldsymbol{\mu} = \vec{\boldsymbol{\mu}} + \hat{\lambda}_{MR}P_{MR} \tag{11}$$

Multi-Resolution Multi-Organ SSM.
In order to extend this method to the construction of 3D multi-organ models with a large number of sample points, the idea adopted in this study is to first downsample the vertices on the model surface. Then we extend the method of Wilms et al. to train the down-sampled 3D point sets to obtain a multi-resolution model shape. Finally, we interpolate the deformation vectors of the down-sampled vertices to generate deformation vectors of all vertices on the entire surface.

Fig.2 shows the idea of this improved method, where Fig.2 (a) shows the down-sampled vertices (marked in red) on the surface of the organ. There are a total of 3759 vertices in four kinds of organs (livers, spleens, left kidneys, right kidneys) in this study, and 375 vertices are obtained after 10 times down sampling, which can be used to construct a model in a computer with 16G memory. Fig.2 (b) shows the deformation vectors (represented by black arrows) on the down-sampled vertices of this model, and Fig.2 (c) shows the deformation vectors on all vertices by interpolating the deformation vectors of down-sampled vertices over the entire model surface, The interpolation method used is the Laplace iteration diffusion algorithm of the surface:

$$\vec{X}_{i_{(n+1)}} = \vec{X}_{i_n} + \frac{\lambda}{M}\sum_{j=0}^{M}(\vec{X}_{j_n} - \vec{X}_{i_n}) \qquad (12)$$

where n represents the number of iterations, i is the vertex index, $\vec{x}_{i_n}$ represents the deformation vector of the i-th vertex coordinate at the n-th iteration, $j = 0,1,\dots,M$ represents $M+1$ indexes of neighbor vertices around vertex i, λ represents the smooth intensity coefficient. At the beginning of the iteration ($n = 0$), first set the deformation vector $\vec{x}_{i_0}$ of the down-sampled vertices to the modeled feature vector P_{MR}, and set the deformation vector of the other vertices to 0; during the iteration process, keep the deformation vectors of down-sampled vertices always to be P_{MR}, and the deformation vectors of other vertices are calculated by formula (12). In order to obtain the desirable interpolation effect, through repeated testing, the maximum number of iterations is set to 1000, and the value of λ is set to 0.8. It can be seen from Fig.2 (c) that after interpolation, the deformation vectors on a small number of down-sampled points smoothly spread to the entire model surface. According to this method, the whole deformation components of the shape model are obtained by interpolating all vertices on the model surface, and the overall deformation of the model is further realized according to the interpolation results. It should be noted that although the deformation vectors are obtained by interpolation instead of by training all vertices, in the case of limited memory, this method can obtain reasonable deformation vectors for a large number of vertices of multiple organs. After observation and quantitative measurement (see this in Results section), the modelling results are better than the traditional global shape model.

In addition to the interpolation algorithm described above, the selection of the resolution in Wilms et al. method is also improved in this study to make it more applicable to multiple organs. Because the original algorithm of Wilms et al. did not specifically consider the problem of modelling multiple organs, but only imposed geodesic distance constraints on the range of the local deformation to generate multi-resolution models under different distance constraints. However, in the case of multi-organ modelling, a simple geodesic distance cannot properly describe the distance relationship between two points belonging to different organs. For example, if the Euclidean distance between a point at the bottom of a lung and another point at the top of the liver is very close, these two points should have strong correlation in terms of common deformation because of the bottom of the lung and the top of the liver always coincide with each other. But according to the principle of the geodesic distance constraint in the Wilms et al. Algorithm, these two points belong to different organs, and the geodesic distance will be farther, so that the correlation between them in a model becomes smaller and does not meet the deformation regulation of adjacent organs. Based on the above considerations, this study uses Euclidean distance instead of geodesic distance as a constraint. The approach is as follows:

Based on the local deformed multi-organ SSM constructed in steps 1 to 6, the distances of surface vertices are calculated by combining the modified Euclidean distance between different organs with the geodesic distance expressed in equation (3). This article specifies that the geodesic distance of points on different organ models is infinite, and the Euclidean distance of points on different organ models can be calculated. Given a model vector $\vec{s}$ containing O target calibration points, and $\vec{s} = \left(\vec{x}_{1,1}^{T}, \dots, \vec{x}_{1,M_1}^{T}, \vec{x}_{2,1}^{T}, \dots, \vec{x}_{2,M_2}^{T}, \dots, \vec{x}_{O,1}^{T}, \dots, , \vec{x}_{O,M_O}^{T}\right)^{T}$, where M_i represents the number of landmarks of the i-th model, $i \in \{1, \dots, O\}$. Define the undirected graph$G_g(V, E_g)$, $V = \left\{\vec{x}_{i,j} \,\middle|\, i \in \{1, \dots, O\}, j \in \{1, \dots, M_i\}\right\}$ represents the vertices of the undirected graph,$E_g = \left\{\left(\vec{x}_{i,j}, \vec{x}_{i,k}\right) \middle| i \in \{1, \dots, O\}, j \in \{1, \dots, M_i\}, k \in N\left(\vec{x}_{i,j}\right)\right\}$, $N\left(\vec{x}_{i,j}\right)$ is the direct neighbourhood of point $\vec{x}_{i,j}$ on target i. The weight $w_{i,j,k}^{g}$ of edge $\left(\vec{x}_{i,j}, \vec{x}_{i,k}\right)$ is represented by the Euclidean distance between two points:

$$w_{i,j,k}^{g} = \left\|\vec{x}_{i,j} - \vec{x}_{i,k}\right\| \tag{13}$$

The geodesic distance $d_{geo}(\vec{x}, \vec{y})$ between two points on the target surface can be estimated by the shortest path in G_g. There is no connection relationship between different targets in G_g, so the distance between points on different targets is infinite.

Define the second fully connected undirected graph $G_e(V, E_e)$, and the edge weights represent the Euclidean distances of the scaled translation:

$$w_{i,j,l,k}^{e} = \eta\left\|\vec{x}_{i,j} - \vec{x}_{i,k}\right\| + \delta, \quad \eta, \delta \epsilon R \tag{14}$$

Use $d_e(\vec{x}, \vec{y})$, $\vec{x}, \vec{y} \in V$ represents the shortest distance in G_e. Combine equation (13) with (14), It can be seen from the two equations that on the same target, when $\eta \leq 1, \delta = 0$, $d_e(\vec{x}, \vec{y}) \leq d_{geo}(\vec{x}, \vec{y})$; when $\eta \geq 1, \delta \geq \max[d_{geo}\left(\vec{x}_{i,j}, \vec{x}_{i,k}\right)]$, $d_e(\vec{x}, \vec{y}) > d_{geo}(\vec{x}, \vec{y})$. On the same target surface, assumed that the energy required from point $\vec{x}$ to point $\vec{y}$ is equal to $d_{geo}(\vec{x}, \vec{y})$, and the coefficient η represents the energy ratio of moving the same distance in the embedding space R^d , which also means the relative viscosity of the space, δ represents the energy required to overcome the adhesion force to leave the target surface. No matter moving on the target surface or moving in the embedding space, the merged distance $d(\vec{x}, \vec{y})$ of two points $\vec{x}, \vec{y}$ in V is a path with minimum energy. Therefore, the shortest path $d(\vec{x}, \vec{y})$ of the combined fully connected graph $G(V, E)$ can be obtained with edge weights equation (15):

$$
w_{i,j,l,k} = \begin{cases} \min\left(w^{g}_{i,j,k}, w^{e}_{i,j,l,k}\right) & \text{if } \left(\vec{x}_{i,j}, \vec{x}_{l,k}\right) \in E_e \\ w^{e}_{i,j,l,k} & \text{else} \end{cases} \tag{15}
$$

We set the ratio σ of the distance threshold **τ** to 0.99 and 0.5 respectively, and we can get two shape models with different resolutions. Since the geodesic distance between organs is defined as infinity, σ = 0.99 retains the deformation of a single organ very well, and each deformation component in the obtained model corresponds to a certain deformation mode of a single organ. On the other hand, based on the definition of Euclidean distance, when σ = 0.5, the common deformation between vertices is limited to local areas of the organ, and the deformation components of the obtained model correspond to deformation modes of the local area of the organ.

In summary, we combine the traditional global model ($\sigma = 1$, the traditional PCA modeling method), the single organ level model ($\sigma = 0.99$) and the local organ level model ($\sigma = 0.5$) with formulas (6)-(11). A 3D SSM based on prior knowledge is obtained which is suitable for modelling multiple organs, and the deformation components contained in this model divided into three resolution levels, that is, global level(multi-organ level), single organ level and local structure level. This modelling method well reflects the different levels of deformation in a multi-organ combination system and describes the deformation of multiple organs better than traditional global models.

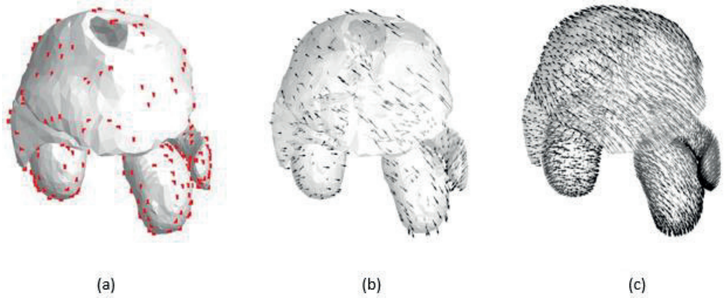


Fig. 2. Schematic diagram of improved method based on downsampling training and deformation vectors interpolation. (a) Down-sampled vertices; (b) Deformation of the down-sampled vertices; (c) The interpolated deformation vectors of all vertices on the surface.

3 Results

Fig.3 shows the modelling effect of the deformation components of the multi-resolution multi-organ shape model constructed by the method in this article. Due to the limited space of this article, the model of each resolution level only shows the results of the first three deformation components on the average shape model. Deformation components are denoted by PC_1, PC_2 and PC_3, respectively. $\lambda_1, \lambda_2, \lambda_3$ are the corresponding eigenvalues, and α_1, α_2, α_3, the shape coefficients of the multi-resolution multi-organ SSM, are set as the weights of model deformation. For each resolution level, the first row show the average models (the average models of the three resolution levels are the same), and the second to fourth rows show the deformation results of the average model with the first three deformation components. Fig. 3 (a), 3 (b) and 3 (c) show the different shapes when the shape coefficient of a component takes different values, and the parts with obvious deformation are circled in the right column. From Fig.3 (a), we can see that the organ deformations reflected by different components all occur together among multiple organs. For example, PC_1 reflects the change in the distance between the left lower lobe of the liver and the spleen, which is most likely caused by the size change of the stomach between them. PC_2 reflects the closeness between the anterior half of the spleen and the left kidney, and PC_3 reflects the change in the distance between the liver and the two kidneys. Fig.3 (b) reflects the deformation of a single organ level, in which PC_1, PC_2 and PC_3 correspond to the deformation of the livers, left kidneys and spleens, respectively. Fig.3 (c) Reflects the local deformation of each organ, such as PC_1 reflects the deformation of the left lower lobe of the liver, PC_2 reflects the deformation of the anterior half of the spleen, PC_3 reflects changes in the anterior curvature of the right kidney. When local deformation is performed, other parts of the same organ keep unchanged. These results show that the method in this study can effectively model the deformation of organs at different resolution levels.

In addition to the above qualitative observation of model deformation modes, two quantitative indicators of generalization and specificity [4] are also used to evaluate the accuracy of model construction in the study.

The generalization of the model is used to measure the model's ability to represent new shapes (that is, shapes not included in the training samples). Generalization can be measured by using Leave-One-Out (LOO) method: assuming there are N training samples, one sample is left as the test sample $\mathrm{S_j}$, and the other $\mathrm{N}-1$ samples $\{\mathrm{S_i} | \mathrm{i} = 1,2,\dots,\mathrm{N}, \mathrm{i} \neq \mathrm{j}\}$ are used to train the model $\mathbf{M}^*$, and then fit $\mathrm{S_j}$ through the deformation of $\mathbf{M}^*$, and calculate the average distance between the fitting result and $\mathrm{S_j}$ as the fitting error $\mathrm{e_j}$. This process is repeated N times (ie j=1,2, …, N), and then set the average error $\mathrm{e_g} = \frac{\sum_{\mathrm{j}=1}^{\mathrm{N}} \mathrm{e_j}}{\mathrm{N}}$ of N times as a measure of model generalization ability, the smaller the value of $\mathrm{e_g}$ is, the better the model generalization ability is.

The specificity of the model is used to measure the model's ability to represent its own training samples. The specificity of the SSM can be tested by randomly generating shape samples: when we get the model based on N training samples, the shape coefficient vectors $\{\vec{\alpha}_\mathrm{j} | \mathrm{j} = 1,2,\dots,\mathrm{K}\}$ of K group models is randomly generated based on the normal distribution, where the mean value of the normal distribution is 0, and the standard deviation is the standard deviation obtained by eigenvalue decomposition of PCA method. Based on each randomly generated coefficient $\vec{\alpha}_j$, generate its corresponding 3D shape, and find a sample whose surface distance is closest to this shape in the training sample set, and set this surface distance as the error $\mathrm{e_j}$ of the jth random sample. Then calculate the average error $\mathrm{e_s} = \frac{\sum_{\mathrm{j}=1}^{\mathrm{K}} \mathrm{e_j}}{\mathrm{K}}$ of K random samples as a measure of model specificity, the smaller the value of $\mathrm{e_s}$ is, the better the model specificity ability is.

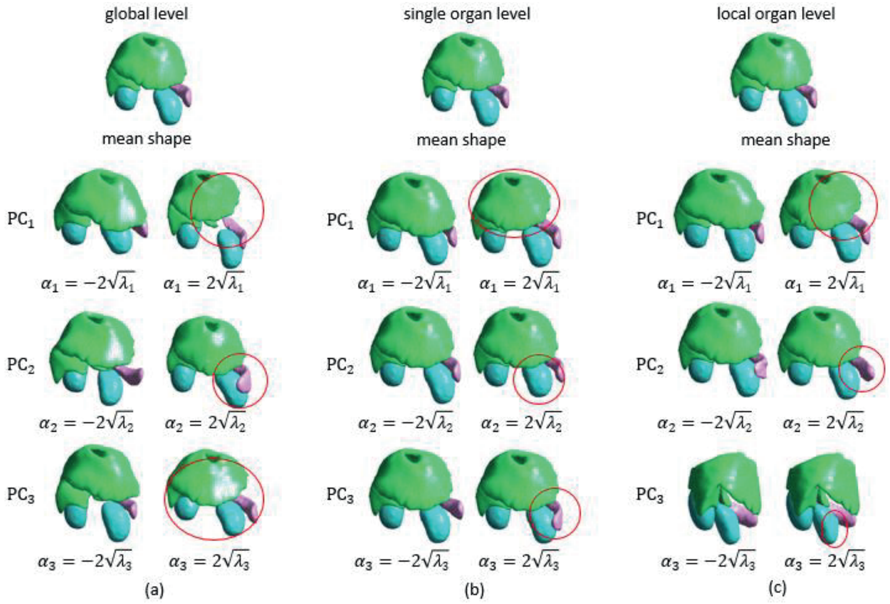


Fig. 3. Deformation components in multi-resolution multi-statistical shape model. (a) Deformation components at the global resolution level; (b) Deformation components at the single-organ resolution level; (c) Deformation components at the local structure level of the organ.

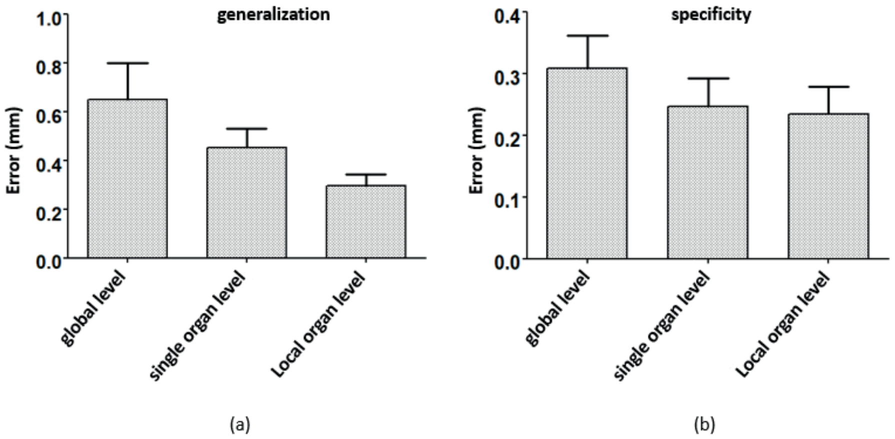


Fig. 4. Quantitative performance evaluation of multi-resolution multi-organ model (a) Generalization error； (b) Specificity error

In order to reflect the improvement effect of the multi-resolution model on the generalization error e_g and the specificity error e_s, this experiment calculates the results of e_g and e_s at different resolution levels, as shown in Fig.4. Both for generalization error and specificity error, the mean value and standard deviation of the three model errors from global resolution level to local structure resolution level are within 1.0 mm. When the model changes from global level to local structure level, the mean value and variance of the errors are gradually decreasing, which indicates that the multi-organ model with local structure level is more accurate for the boundary registration. This means that the multi-resolution multi-organ model constructed in this paper has better generalization and specificity than the global model constructed by traditional PCA method. Encouragingly, even for generalization errors, the minimum mean value of the multi-resolution multi-organ model has reached about 0.3 mm, which is close to the minimum mean value of the specificity error of 0.25 mm. More importantly, it is also close to the pixel resolution of mouse CT images of 0.2 mm, and is lower than the average specificity error of the traditional global model of 0.31 mm.

4 Conclusion

This article proposes a multi-resolution multi-organ shape prior knowledge model construction method and uses it to model multiple abdominal organs of mouse micro-CT images. Compared to the recently proposed state-of-the-art 2D multi-resolution SSM method by Wilms et al., our method solves the shortcomings of memory occupation and thus extend the method to 3D space. On the other hand, this work extend the method to multi-organ modelling and can be used for modelling the inter-subject shape changes of multi-organ, single organ and local structure levels. This method surpasses the traditional PCA modelling method in terms of both generalization and specificity. It should be pointed out that although this work builds a model based on the abdominal organs of mouse, the method in this study is also applicable to the multi-organ modelling of human or other animal bodies. The model constructed in this work lays the foundation of shape prior knowledge for further multi-organ image segmentation.

Acknowledgements

This study was funded by the general program of the National Natural Science Fund of China (No. 81971693, 81401475), the Science and Technology Innovation Fund of Dalian City (2018J12GX042) and the Fundamental Research Funds for the Central Universities (DUT19JC01). We thank the Molecular Imaging Centre of the University of California, Los Angeles for providing 98 mouse CT images, and the scholarships from China scholarship Council (No. 201806060163).

References

1. BONDIAU, P.Y. and MALANDAIN, G., 1997. Eye reconstruction and CT-retinography fusion for proton treatment planning of ocular diseases. In CVRMed-MRCAS'97, J.

TROCCAZ, E. GRIMSON and R. MÖSGES Eds. Springer Berlin Heidelberg, Berlin, Heidelberg, 705-714.

2. COOTES, T.F., EDWARDS, G.J., and TAYLOR, C.J., 2001. Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence 23, 6, 681-685.
3. COOTES, T.F., TAYLOR, C.J., COOPER, D.H., and GRAHAM, J., 1995. Active Shape Models-Their Training and Application. Computer Vision and Image Understanding 61, 1 (1995/01/01/), 38-59.
4. HEIMANN, T. and MEINZER, H.-P., 2009. Statistical shape models for 3D medical image segmentation: A review. Medical Image Analysis 13, 4 (2009/08/01/), 543-563.
5. HIGHAM, N.J., 2002. Computing the nearest correlation matrix—a problem from finance. IMA Journal of Numerical Analysis 22, 3, 329-343.
6. LORENSEN, W. and CLINE, H., 1987. Marching Cubes: A High Resolution 3D Surface Construction Algorithm. ACM SIGGRAPH Computer Graphics 21(08/01), 163.
7. LORENZ, C. and KRAHNSTOVER, N., 1999. 3D statistical shape models for medical image segmentation. In Second International Conference on 3-D Digital Imaging and Modeling (Cat. No.PR00062), 414-423.
8. MARANI, R., RENÒ, V., NITTI, M., D'ORAZIO, T., and STELLA, E., 2016. A Modified Iterative Closest Point Algorithm for 3D Point Cloud Registration. Computer-Aided Civil and Infrastructure Engineering 31, 7 (2016/07/01), 515-534.
9. PARK, S. and LIM, S.-J., 2014. Template-Based Reconstruction of Surface Mesh Animation from Point Cloud Animation. ETRI Journal 36(12/01), 1008-1015.
10. RUECKERT, D. and BURGER, P., 1997. Shape-based segmentation and tracking in 4D cardiac MR images. In CVRMed-MRCAS'97, J. TROCCAZ, E. GRIMSON and R. MÖSGES Eds. Springer Berlin Heidelberg, Berlin, Heidelberg, 43-52.
11. SANDOR, S. and LEAHY, R., 1997. Surface-based labeling of cortical anatomy using a deformable atlas. IEEE Transactions on Medical Imaging 16, 1, 41-54.
12. STEGMANN, M. and GOMEZ, D., 2002. A Brief Introduction to Statistical Shape Analysis. Informatics and Mathematical Modelling, Technical University of Denmark, DTU 15(01/01).
13. SUCKOW, C., KUNTNER, C., CHOW, P., SILVERMAN, R., CHATZIIOANNOU, A., and STOUT, D., 2009. Multimodality Rodent Imaging Chambers for Use Under Barrier Conditions with Gas Anesthesia. Molecular Imaging and Biology 11, 2 (2009/03/01), 100-106.
14. SUCKOW, C.E. and STOUT, D.B., 2008. MicroCT Liver Contrast Agent Enhancement Over Time, Dose, and Mouse Strain. Molecular Imaging and Biology 10, 2 (2008/03/01), 114-120.
15. WANG, H., STOUT, D.B., and CHATZIIOANNOU, A.F., 2012. Estimation of Mouse Organ Locations Through Registration of a Statistical Mouse Atlas With Micro-CT Images. IEEE Transactions on Medical Imaging 31, 1, 88-102.
16. WILMS, M., HANDELS, H., and EHRHARDT, J., 2017. Multi-resolution multi-object statistical shape models based on the locality assumption. Medical Image Analysis 38 (02/01).

II

CONSTRUCTION OF MULTI-RESOLUTION MULTI-ORGAN SHAPE MODEL BASED ON STACKED AUTOENCODER NEURAL NETWORK

by

Zhonghua Chen, Hongkai Wang, Fengyu Cong, and Lauri Kettunen 2022

2022 14th International Conference on Advanced Computational Intelligence (ICACI), pp. 62-67

https://doi.org/10.1109/ICACI55529.2022.9837706

Construction of Multi-resolution Multi-organ Shape Model Based on Stacked Autoencoder Neural Network

1st Zhonghua Chen
School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering
Faculty of Information Technology
Dalian University of Technology
University of Jyväskylä
Dalian, China, 116024
Jyväskylä, Finland, 40014
zhonghua.x.chen@student.jyu.fi

2nd Hongkai Wang
School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering
Dalian University of Technology
Dalian, China, 116024
wang.hongkai@dlut.edu.cn

3rd Fengyu Cong
School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering
Faculty of Information Technology
School of Artificial Intelligence, Faculty of Electronic Information and Electrical Engineering
Key Laboratory of Integrated Circuit and Biomedical Electronic System
Dalian University of Technology
University of Jyväskylä
Dalian, China, 116024
Jyväskylä, Finland, 40014
cong@dlut.edu.cn

4th Lauri Kettunen
Faculty of Information Technology
University of Jyväskylä
Jyväskylä, Finland, 40014
lauri.y.o.kettunen@jyu.fi

***Abstract*— The construction of statistical shape models (SSMs) is an important method in the field of medical image segmentation. Most SSMs are constructed by using traditional modeling methods based on principal component analysis (PCA), which cannot fully present the true deformation ability of models. To solve the insufficient deformation ability of SSMs, we propose a stacked autoencoder (SAE) neural network to construct a multi-resolution multi-organ shape model based on mouse micro-CT images, which can express more linear and non-linear deformations than SSMs based on PCA. The main advantage of this method is that the SAE neural network is simple and flexible and it can learn more deformation modes from training data. We have quantitatively compared the modeling performance of this method with the constructed SSMs based on PCA in terms of model generalization and specificity.**

***Index Terms*— statistical shape model, principal component analysis, mouse micro-CT images, stacked autoencoder neural network, multi-resolution multi-organ shape model**

I. INTRODUCTION

The construction of statistical shape models (SSMs) is an important segmentation method in the field of medical image segmentation. The applications of medical image segmentation methods focus on the segmentation of biological tissues and organs. Some applications consist of liver segmentation for volume measurement [1], breast tumor segmentation for diagnosis [2], bone localization and segmentation [3], and automatic heart segmentation for treatment [4], etc. In the past few decades, SSM methods have been employed to construct segmentations for single specific organs, especially for research purposes. Only a few other modeling methods focused on the segmentation of multiple organs. The most important approach in constructing SSMs is the principal component analysis (PCA) based on eigenvalue decomposition [5]. Given a set of medical images, such as computed tomography (CT) images, PCA approaches are used to calculate the deformation components and corresponding deformation coefficients. Thereafter the deformation components are added to the mean shape of these medical images to obtain a traditional standard SSM. Based on this rationale, various techniques have been proposed to improve the calculation of deformation components. Many other features of images are added to the PCA process to construct more representative models for images: they include different variants based on active shape models [6] and active appearance models [7] proposed by Cootes et al. Later in 2009 Tresadern et al. [8] combined an MRF-based local shape model with a PCA-based global shape model for modeling and locating deformable objects. Wilms et al. [9] expanded this work to construct a two-dimensional (2D) multi-resolution multi-organ model for the segmentation of 2D CT images.

Although PCA-based shape modeling methods have been used widely in medical image segmentation and many

---

National Natural Science Fund of China (No. 81971693, 81401475),
Science and Technology Innovation Fund of Dalian City (2018J12GX042),
Fundamental Research Funds for the Central Universities (DUT19JC01).
China Scholarship Council (No. 201806060163)

different improved variants have been proposed in the last three decades, each modification can only be used in a specific application. Their performance may degrade dramatically when there are outliers in the data of the training models. Due to the limited deformation ability of linear SSMs, some non-linear PCA methods have been proposed to represent variations of bending and rotation in a more natural fashion [10, 11, 12]. Furthermore, when there are large amounts of three-dimensional (3D) training models, the computer's memory requirements increase exponentially, which leads to the failure of model construction [9, 13].

On the other hand, in recent years autoencoder-based shape modeling approaches have been proposed. They can learn latent non-linear deformations of shape models and represent more refined expressions in deformation models. In 2018, Litany et al. used a variational autoencoder to learn the latent space of objects with non-rigid deformations to reconstruct missing parts of objects [14]. At the same time, Ranjan et al. proposed to use convolutional mesh autoencoders learning non-linear variations for the construction of 3D faces [15].

A trained autoencoder network is simple for model construction and it only needs to save a few trained network parameters compared with PCA-based methods. In this paper, we have stacked autoencoder networks to build an SAE neural network that is used to construct a multi-resolution multi-organ shape model based on abdominal mouse micro-CT images. We have compared the models obtained from the SAE neural network with traditional global SSMs and local SSMs in terms of model generalization and specificity performance.

II. MATERIALS AND METHODS

A. Mouse Micro-CT Data

In this study, we have collected micro-CT images of 98 mice from the Molecular Imaging Centre of the University of California, Los Angeles. For neural networks, sufficient training data should be provided to train the network. Practically, there are certain challenges in the collection of medical images, and 98 mouse images can be regarded as large sample data in the domain of medical image analysis. The detailed operating rules of mouse micro-CT imaging and the parameter settings of imaging equipment are described in our previous study [16].

After we have collected 98 mouse micro-CT images, we need to preprocess these images to obtain the final training data. The specific process is as follows: (1) We invited small animal imaging experts to manually segment 98 mouse micro-CT images to obtain labels of the liver, spleen, left kidney, and right kidney. (2) Next, these labels have been converted into multi-organ surface meshes with the moving cube algorithm [17]. (3) After this, we chose a multi-organ surface mesh from the meshes as a reference template and used the point cloud registration method [18, 19] to register it to the other 97 meshes. Finally, each mesh has the same number of points and each point on the mesh corresponds to the same anatomical position. (4) We performed the same downsampling on all points on every mesh to obtain point cloud data that can be trained by the network.

B. Methods

1) Description of autoencoder network

An autoencoder network is an unsupervised neural network that learns input data and reconstructs the input data to the greatest extent. Normally, the simplest autoencoder has three layers: an input layer, a hidden layer, and an output layer. The number of nodes in the output layer is the same as the input layer. Since the input data is lossy and relevant during compression and decompression, the autoencoder network can learn compression and decompression functions that contain specific features of the data and can be used for data dimensionality reduction and feature extraction. In practical applications, an autoencoder network must try to obtain the most important features that can represent the input data.

An autoencoder consists of an encoder ϕ, and a decoder Ψ, which can be defined as follows:

$$\begin{gathered} \phi: X \rightarrow F \\ \Psi: F \rightarrow X \\ \phi, \Psi = \operatorname{argmin}_{\phi,\Psi} \|X - (\Psi \circ \phi)X\|^2 \end{gathered} \qquad (1)$$

where *argmin* represents obtaining minimum values of ϕ, Ψ simultaneously.

In the simplest case, given one hidden layer, the encoder stage of an autoencoder takes the input $x \epsilon R^d = X$ and maps it to $h \in R^p = F$:

$$h = \sigma(Wx + b), \qquad (2)$$

where h is referred to as code, latent variables, or latent representation, σ is an element-wise activation function such as a sigmoid function or a rectified linear unit, W is a weight matrix, and b is a bias vector.

The decoder stage of the autoencoder maps h to the reconstruction x' of the same shape x:

$$x' = \sigma'(W'h + b'), \qquad (3)$$

where σ' is also an element-wise activation function, W' is a weight matrix and b' is a bias vector.

Autoencoders are trained to minimize reconstruction errors which are often referred to as the "loss":

$$\mathrm{L}(x,x') = \|x - x'\|^2 = \left\|x - \sigma'\left(W'\left(\sigma(Wx + b)\right) + b'\right)\right\|^2 \qquad (4)$$

The basic structure of an autoencoder network is shown in Fig. 1. The first column is the input layer, the middle column is a hidden layer and the last column is the output layer, +1 indicates an offset constraint condition added to the corresponding layer.

2) Multi-resolution shape modeling based on SAE neural network

A stacked autoencoder (SAE) neural network is composed of multiple networks connected in series. In this paper, the purpose of using the SAE neural network to train input mesh data is to extract high-dimensional features of the data layer

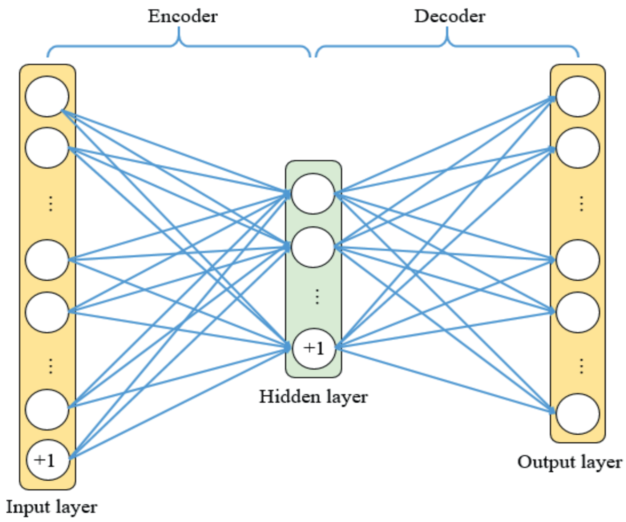


Fig. 1. Basic structure of an autoencoder network.

by layer and to reduce the dimension of the input data layer by layer. Therefore, the SAE neural network transforms high-dimensional mesh data into a series of low-dimensional feature vectors. Finally, these low-dimensional feature vectors are decompressed to reconstruct mesh data approximately equal to the input mesh data. In the process of decompressing features and restoring data, if we regularly modify high-dimensional feature parameters, we can see that the reconstructed data change regularly, which is also reflected in the deformation of 3D models.

Fig. 2 shows the multi-resolution shape modeling process based on SAE neural network. E represents a compression network (i.e., an encoder) and D represents a decompression network (i.e., a decoder). Each combination of E and D is an autoencoder network introduced in section 1).

This study uses an SAE neural network with three autoencoder networks. The training data for the first level E_1 and D_1 is the shape vector of meshes described in section A. After training, E_1 compresses the shape vector set into a shorter feature vector set $\overrightarrow{F_1} = \{\vec{f}_{1,1}, \vec{f}_{1,2}, \dots, \vec{f}_{1,M}\}$, where M is the number of training meshes, and the subscript i, j of $\vec{f}_{i,j}$ represents the *j-th* sample of the *i-th* level. The $\vec{F}_1$ obtained from the first level network is used as the training sample of the second level network to obtain the training results of E_2 and D_2. The second level network generates a further compressed feature vector set $\vec{F}_2 = \{\vec{f}_{2,1}, \vec{f}_{2,2}, \dots, \vec{f}_{2,M}\}$. Then $\vec{F}_2$ is used to train the third level network (i.e., E_3 and D_3). As a result, the vector lengths of $\vec{F}_1$, $\vec{F}_2$, and $\vec{F}_3$ decrease gradually, which means the compression ratio gradually increases. From the perspective of shape characteristics, the resolution of the shape deformation should gradually transition from a global model to a local part of the model, thus forming a multi-resolution training method.

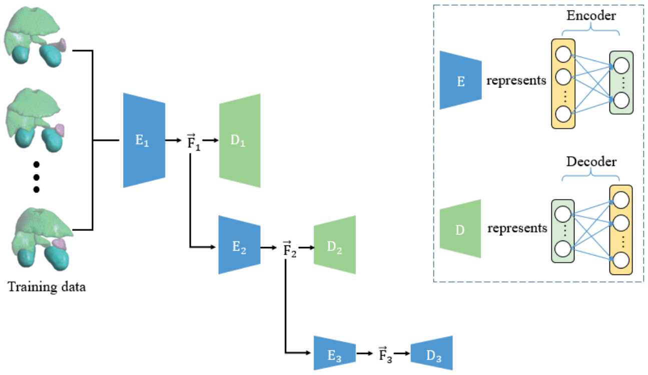


Fig. 2. The overall structure of a three-level SAE network with 3 autoencoder networks.

The SAE neural network used in this research encounters the problem of too large model data. Each 3D training mesh contains 3759 surface sampling points and the shape vector of a mesh contains 11277 dimensions, which takes up a lot of memory and causes a computational burden for SAE neural network training. Due to the limitation of the memory capacity of the computer used in this study, the points of each training mesh are downsampled to 752 as described in section A. Then the down-sampled model vertices are trained by the SAE neural network, and the generated deformation vectors are interpolated through the Laplace iteration diffusion algorithm (5) to obtain deformation vectors of 3759 vertices of a mesh.

$$\vec{X}_{i_{(n+1)}} = \vec{X}_{i_n} + \frac{\lambda}{M}\sum_{j=0}^{m}(\vec{X}_{j_n} - \vec{X}_{i_n}) \qquad (5)$$

where n represents the number of iterations, i is the vertex index, $\vec{x}_{i_n}$ represents the deformation vector of the *i-th* vertex at the *n-th* iteration, $j = 0,1,\dots,m$ represents *m+1* indexes of neighbor vertices around vertex i, λ represents the smooth intensity coefficient.

In terms of the SAE neural network parameters set, we try to select the appropriate number of nodes in the hidden layer, and output layer as well as the number of levels in the SAE neural network by changing different parameter values in this study. After that, we determine to use a three-level SAE neural network in this study. The number of input nodes of the first level encoder is 752 × 3 = 2256 (i.e., the number of points of down-sampled mesh multiplied by the spatial dimension), and the corresponding number of output nodes is 1000. The number of input and output nodes of the second level encoder is 1000 and 200, respectively. The number of input and output nodes of the third level encoder is 200 and 30, respectively. Finally, the shape model including multiple organs is represented by the 30-dimensional feature vector $\vec{f}_3$, which is the output of the third level encoder.

3) *Shape fitting and generation with SAE neural network*

Based on the trained three-level SAE neural network, we can do multi-organ shape fitting and generation for new mouse micro-CT images. As shown in Fig. 3 (a), the shape vectors of multiple organs are input to a three-level SAE neural network. Through the process of three compressions and three decompressions, a new shape is reconstructed as the fitting result of the input shape. According to the principle of the SAE neural network, the output shape should be as similar as possible to the input shape, and the error between them indicates the fitting error. The smaller the fitting error is, the better the training performance of the SAE neural network is. In this paper, we calculate the average surface distance [20]

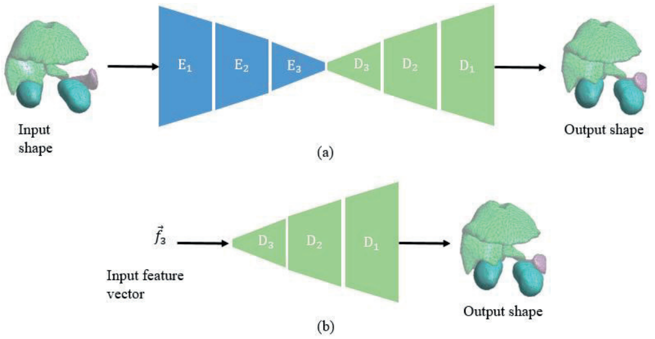


Fig. 3. Shape fitting and shape generation using the SAE neural network. (a) Shape fitting; (b) Shape generation.

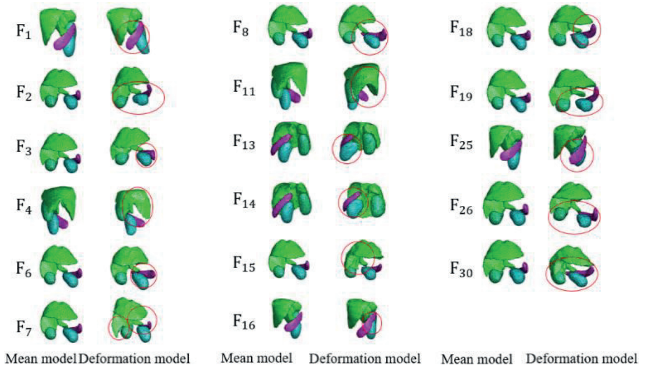


Fig. 4. Multi-organ model reconstructed by SAE neural network.

between the constructed model and a training mesh to evaluate the modeling performance [21] of our method. If an input shape is one of the training samples, the fitting error reflects the specificity of the SAE neural network; if an input shape is not included in the training sample set, the fitting error reflects the generalization of the SAE neural network.

Fig. 3 (b) shows the shape generation process based on the three-level decoder. If a randomly selected feature vector $\vec{f}_3$ is input to the network, the corresponding output shape can be obtained; if the training performance of the network is effective, the output shape should correspond to the true anatomical shape. Further, if the mean shape of the training meshes is input to the three-level SAE neural network, the feature vector $\overline{\vec{f}_3}$ corresponding to the mean shape can be obtained. And if any value in any dimension of $\overline{\vec{f}_3}$ is modified, the deformation effect of the mean shape can be constructed, thus realizing the construction of the multi-resolution shape model for multiple organs.

III. RESULTS

A. Multi-resolution Shape Modeling Results Based on SAE Neural Network

According to the shape generation method in the previous section, we take the 30-dimensional feature vector $\overline{\vec{f}_3}$ of the mean shape as the basis and make the unit size, which deviates from the average value, to be 0.1 by changing the value of each dimension in the study. At last, deformation results of the mean shape corresponding to 30 dimensions are obtained. we have found that in the process of controlling the deformation of the 30-dimensional components, some of the deformation laws of the model show a certain similarity. Therefore, in Fig. 4, 17 representative deformation components are selected, where $F_i(i = 1,2,3,4,6,7,8,11,13,14,15,16,18,19,25,26,30)$ represents the *i-th* deformation component of the third level in the SAE neural network. The first, third, and fifth columns are the mean models, and the second, fourth, and sixth columns are the deformation models reconstructed by adjusting these deformation components in this network. Some areas of the reconstruction model that have obvious deformation are circled. As Fig. 4 shows, in the reconstruction of the multi-organ shape model, different deformation components in the third level of the SAE neural network control the deformation of different areas of the model. Such as components F_4, F_7, F_{11}, F_{15} can control liver deformation, components F_2, F_3, F_6, F_{13} can control spleen deformation, components F_7, F_{15}, F_{30} can control right kidney deformation, components F_{13}, F_{14}, F_{25}, F_{30} can control left kidney deformation, and the components F_8, F_{26}, etc. can control the relative position changes of certain organs. Each deformation component of constructing a multi-organ model not only controls the deformation of a certain organ, but also affects the changes in the shape and position of multiple organs. In the constructed model, there is not only a single organ with obvious local deformations, but also multiple organs with obvious global deformations, which shows a multi-resolution deformation effect.

B. Comparison of SAE Neural Network and SSM Modeling

In the previous study, we used a modified traditional PCA method to construct traditional global SSM and multi-resolution SSM for the mouse micro-CT images and obtained global and local deformation modes [16]. On this basis, we further compare the model constructed by the SAE network in this study with the multi-resolution model constructed in the previous study. We quantitatively analyze the generalization and specificity of the models constructed by these two different approaches, based on the average surface distance errors of the constructed model and each training data.

For comparison, Fig. 5 shows the traditional PCA method to construct a global SSM for realizing the global deformation effect. The first and fourth columns are the mean models, which are represented by $\boldsymbol{\mu}$. The deformation components are represented by PC_1, PC_2 and PC_3. $\lambda_i(i = 1,2,\dots,10)$ is the eigenvalue obtained by implementing eigenvalue decomposition on the covariance matrix of training sample points, which corresponds to the deformation component. $\alpha_i(i = 1,2,\dots,10)$ is the model coefficient, which is set as the weight of the deformation model. Comparing Fig. 4 and Fig. 5, we can find that almost all the deformation models constructed by PCA methods can also be constructed by using SAE neural network. However, the multi-organ deformation model reconstructed by SAE neural network has more deformation modes than that constructed by PCA methods.

In this study, the SAE neural network is mainly used to reconstruct the mouse abdomen multi-organ model. As a comparison, the traditional global SSM and multi-resolution SSM are introduced to further analyze the accuracy of these

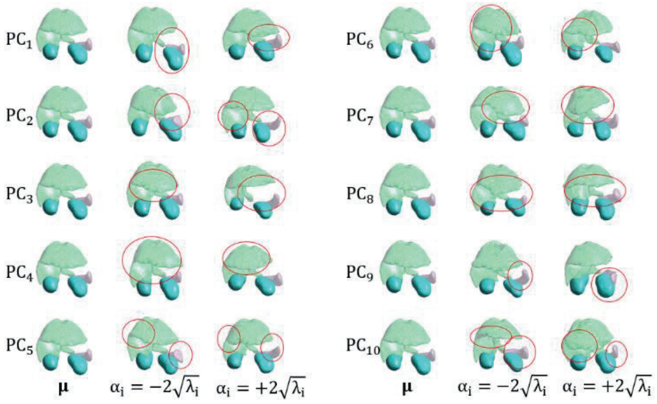


Fig. 5. Multi-organ model constructed by PCA method.

three modeling methods. It can be seen from Fig. 6 (a) that in terms of generalization, the modeling error of the traditional global multi-organ SSM and multi-resolution model are smaller than that of the SAE neural network. Especially for the multi-resolution model, the error is around 0.3 mm, while the modeling error of the SAE neural network is about 1.5 mm. In Fig. 6 (b), in terms of specificity, the modeling errors of the traditional global multi-organ SSM and the multi-resolution model have reached the level of 0.3 mm and 0.2 mm, respectively, but the modeling error of the SAE neural network is still around 1.5 mm. According to the analysis, although the principle of the SAE neural network is simple and easy to implement, it is still inferior to the linear SSM construction method based on the PCA method. We also find that the number of hidden layers, the number of nodes, and the sparseness of the SAE neural network may cause the training error to become larger, which needs to be verified in future work.

IV. CONCLUSION

The method proposed in this paper is more concise and feasible, and the model constructed by it can express much richer deformation modes. The SAE neural network is simpler than PCA-based methods in terms of algorithm complexity. But the deformation successfully reflects multi-resolution changes in shapes. Moreover, SAE neural network is a new type of multi-resolution modeling method based on shape prior knowledge and is easy to implement. SAE neural network has potential application for multi-resolution organ segmentation as the multi-resolution features are important for the shape modeling of multiple organs. The advantages of the proposed method are that the constructed diverse deformation components consist of nonlinear and linear modeling characteristics. For example, the deformation components obtained by SAE neural network are much more than that of PCA-based methods. Likely, the model built by SAE neural network consists of much more local deformation components which can change the local shape of an organ in detail. Although the error rates of generalization and specificity of the SAE neural network are still a little bit higher, the excellent nonlinear multi-resolution modeling characteristics deserve further improvement and optimization. On the other hand, SAE neural network is a nonlinear modeling method, which is more suitable for simulating nonlinear shape space with a manifold distribution. In theory, it can obtain a more accurate modeling effect than traditional linear modeling methods. Furthermore, SAE neural network modeling method has multi-resolution characteristics not shared by traditional methods, which means SAE neural network modeling method performs better theoretically.

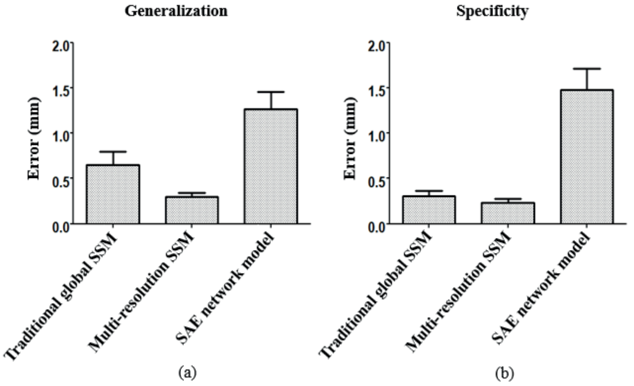


Fig. 6. Shape reconstruction accuracy comparison among the SAE neural network with the traditional and multi-resolution models. (a) Generalization errors; (b) Specificity errors.

ACKNOWLEDGMENT

We thank the Molecular Imaging Center of the University of California, Los Angeles for providing 98 mouse micro-CT images to support our work.

REFERENCES

[1] S. J. Lim, Y. Y. Jeong, and Y. S. Ho, “Automatic liver segmentation for volume measurement in CT Images,” *Journal of Visual Communication and Image Representation*. vol. 17, pp. 860–875, 2006.

[2] N. Karunanayake, P. Aimmanee, W. Lohitvisate, and S.S. Makhanov, “Particle method for segmentation of breast tumors in ultrasound images,” *Mathematics and Computers in Simulation*. vol. 170, pp. 257–284, 2020.

[3] A. Klein, J. Warszawski, J. Hillengaß, and K. H. Maier-Hein, “Automatic bone segmentation in whole-body CT images,” *International Journal of Computer Assisted Radiology and Surgery*. vol. 14, pp. 21–29, 2019.

[4] X. Zhuang, K. S. Rhode, R. S. Razavi, D. J. Hawkes, and S. Ourselin, "A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI," *IEEE Transactions on Medical Imaging*. vol. 29, pp. 1612–1625, September 2010.

[5] M. B. Stegmann and D. D. Gomez, “A brief introduction to statistical shape analysis,” *Informatics and Mathematical Modelling, Technical University of Denmark*, Denmark, pp. 1–15, March, 2002.

[6] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, “Active shape models – their training and application,” *Computer Vision and Image Understanding*. vol. 61, pp. 38–59, 1995.

[7] T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active appearance models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*. vol. 23, pp. 681–685, June 2001.

[8] P. Tresadern, H. Bhaskar, S. A. Adeshina, and C. J. Taylor, “Combining local and global shape models for deformable object matching,” *British Machine Vision Conference*. London: BMVA Press, 2009, pp. 1–12.

[9] M. Wilms, H. handels, and J. Ehrhardt, “Multi-resolution multi-object statistical shape models based on the locality assumption,” *Medical Image Analysis*. vol. 38, pp. 17–29, 2017.

[10] P. D. Sozou, T. F. Cootes, C. J. Taylor, and E. C. Di-Mauro, “A non-linear generalisation of PDMs using polynomial regression,” *British Machine Vision Conference*. York : BMVA Press, 1994, pp. 397–406.

[11] P. D. Sozou, T. F. Cootes, C. J. Taylor, E. C. Di Mauro, and A. Lanitis, “Non-linear point distribution modelling using a multi-layer perceptron,” *Image and Vision Computing*. vol. 15, pp. 457–463, 1997.

[12] C. J. Twining and C. J. Taylor, “Kernel principal component analysis and the construction of non-linear active shape models,” *British Machine Vision Conference*. Manchester: BMVA Press, 2001, pp. 1–10.

[13] H. Huang, F. Makedon, and R. McColl, “High dimensional statistical shape model for medical image analysis,” *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. pp. 1541–1544, 2008.

[14] O. Litany, A. Bronstein, M. Bronstein, and A. Makadia, “Deformable shape completion with graph convolutional autoencoders,” *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 1886–1895, 2018.

[15] A. Ranjan, T. Bolkart, S. Sanyal, and M. J. Black, “Generating 3D faces using convolutional mesh autoencoders,” *European Conference on Computer Vision*. Cham: Springer, 2018, pp. 725–741.

[16] Z. Chen, T. Ristaniemi, F. Cong, and H. Wang, “Multi-resolution statistical shape models for multi-organ shape modelling,” *International Symposium on Neural Networks*. Cham: Springer, 2020, pp. 74–84.

[17] W. E. Lorensen and H. E. Cline, “Marching Cubes: A high resolution 3D surface construction algorithm,” *Computer Graphics*. vol. 21, pp. 163–169, July 1987.

[18] R. Marani, V. Renò, M. Nitti, T. D'Orazio, and E. Stella, “A modified iterative closest point algorithm for 3D point cloud registration,” *Computer-Aided Civil and Infrastructure Engineering*. vol. 31, pp. 515–534, July 2016.

[19] S. I. Park and S.-J. Lim, “Template-based reconstruction of surface mesh animation from point cloud animation,” *ETRI Journal*. vol. 36, pp. 1008–1015, December 2014.

[20] H. Wang, D. B. Stout, R. Taschereau, Z. Gu, N. T. Vu, and D. L. Prout et al, “MARS: a mouse atlas registration system based on a planar x-ray projector and an optical camera,” *Physics in Medicine and Biology*. vol. 57, pp. 6063–6077, October 2012.

[21] T. Heimann and H. P. Meinzer, “Statistical shape models for 3D medical image segmentation: a review,” *Medical Image Analysis*. vol. 13, pp. 543–563, August 2009.

III

LOW-DOSE MOUSE MICRO-CT IMAGE SEGMENTATION BASED ON MULTI-RESOLUTION MULTI-ORGAN SHAPE PRIOR KNOWLEDGE MODEL

by

Zhonghua Chen, Hongkai Wang, Fengyu Cong, and Lauri Kettunen 2022

2022 International Conference on Computers, Information Processing and
Advanced Education (CIPAE), pp. 349-353

https://doi.org/10.1109/CIPAE55637.2022.00079

Low-dose mouse micro-CT image segmentation based on multi-resolution multi-organ shape prior knowledge model

Zhonghua Chen
School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering
Faculty of Information Technology
Dalian University of Technology
University of Jyväskylä
Dalian, China
Jyväskylä, Finland
zhonghua.x.chen@student.jyu.fi

Hongkai Wang
School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering
Dalian University of Technology
Dalian, China
wang.hongkai@dlut.edu.cn

Fengyu Cong
School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering
Faculty of Information Technology
School of Artificial Intelligence, Faculty of Electronic Information and Electrical Engineering
Key Laboratory of Integrated Circuit and Biomedical Electronic System
Dalian University of Technology
University of Jyväskylä
Dalian, China
Jyväskylä, Finland
cong@dlut.edu.cn

Lauri Kettunen
Faculty of Information Technology
University of Jyväskylä
Jyväskylä, Finland
lauri.y.o.kettunen@jyu.fi

Abstract—**Automatic segmentation of computed tomography (CT) images of mice is a step toward computer-assisted preclinical image analysis. Due to the low image quality of micro-CT images, fully-automatic methods may not achieve robust segmentation. For this reason, human interventions are needed to achieve higher segmentation accuracy. In this paper, we propose a human interactive segmentation method incorporating anatomical prior knowledge for multiple abdominal organs in mouse micro-CT images. The method automatically fits a multi-organ shape model to the user-sketched partial boundary contours. Segmentation accuracy is validated by comparing the proposed method against existing shape models. The robustness of our proposed method was evaluated with different users. Finally, the results suggest the proposed method generates accurate segmentation with good robustness.**

Keywords—mouse micro-CT images, statistical shape model (SSM), human interaction, Variational Hermite Radial Basis Function (VHRBF), multi-resolution multi-organ shape model

I. INTRODUCTION

The imaging of small animals plays an important role in preclinical drug development and cancer research. Among various small animal imaging, computed tomography (CT) is the most widely used anatomical imaging approach. Many functional imaging techniques, such as positron emission tomography (PET), and single-photon emission computed tomography (SPECT), use CT images as anatomy references. Mice are one of the most commonly used small animals in preclinical research. Internal organ segmentation in mouse micro-CT images is important for lesion localization, organ morphometric measurements, anatomical structure modeling and simulation, and the research of organ-level pharmacokinetics.

However, in mouse micro-CT images the contrast between soft tissues is poor. This is a limitation. In the process of imaging, to minimize the X-ray dose, shorter exposure time or limited projection angles are usually employed. But in mouse micro-CT images these result in blurred edges of soft tissues or organs. Although a variety of soft-tissue contrast agents are available, the use of contrast agents can increase the complexity and cost of experiments. Sometimes, they even affect the experimental results. Therefore, contrast agents are not used in imaging a large number of mice. This implies challenges to the segmentation of soft tissues and organs.

Due to the low soft-tissue contrast in mouse micro-CT images, human intervention and organ shape prior knowledge are needed to achieve robust and accurate segmentation. As a growing subfield of image analysis, the study of medical image segmentation based on statistical shape models (SSMs) has become an important tool in medical image analysis. In medical applications, SSMs delineate regions of interest (ROI) accurately. Different shape models, be they active shape models (ASMs) [1], or active appearance models (AAMs) [2] lay the foundations for more feasible segmentation. As later variant shape models are constructed with more shape information and improved algorithms. A systematic review of variants of ASMs and AAMs for three-dimensional (3D) medical image segmentation concluded that more detailed shape information contributed to better shape segmentation [3].

The most detailed SSM models are designed to segment specific organs, tissues, or anatomical structures [4-6]. There exists a classic and well-known method for building SSMs [7] with a point distribution model (PDM) representing a shape appearance [8]. In a subsequent study, it was found that a priori knowledge of the shape of bones and the angles between joints can facilitate the segmentation of single regions [9]. Furthermore, a priori knowledge of shape, especially on the weights assigned to points in the PDM is one of the most critical factors in achieving a correspondence between points and target edges in CT images [10]. Currently, more generic SSMs have been developed to segment anatomical structures with severe shape abnormality or missing, such as finding effects between the reference model and specific heart individual through remodeling transformation [11], segmenting limbs in CT images by updating and refining the model with newly segmented training samples [12].

Another challenge is the so-called High-Dimension-Low-Sample-Size (HDLSS) problem in model construction [13]. As a traditional method, principal component analysis (PCA), especially the variants based on it, has become popular for the segmentation, registration, or reconstruction of different anatomical structures. While PCA-based methods have been applied to more complex anatomies, such as liver, lung, bone, brain structure, or even the whole human skeleton, a large number of training samples is needed and should be pre-processed [14]. Fortunately, to some extent, the local distance constraints for points in SSMs can solve the HDLSS problem [15]. Furthermore, it is not easy to collect and provide more than hundreds of medical images for training an SSM. Manual segmentation of hundreds of medical images is a very time-consuming task. There is either not enough prior knowledge about the training shape to build an SSM based on traditional PCA. Likely, the local segmentation of an organ is inaccurate, if there is not enough information on the shape available.

In this study, we combine our existing multi-resolution multi-organ shape model [16] with Variational Hermite Radial Basis Function (VHRBF) [17] for multiple organ segmentation in low-contrast mouse micro-CT images. The original mouse micro-CT images are obtained from mice injected with contrast agents. The training data was generated by adding some random noise and by changing manually the contrast of the original images. The use of VHRBF followed by the SSM deformation results in an effective segmentation process. We evaluate the segmentation accuracies of several organs, both qualitatively and quantitatively, in terms of the multi-resolution multi-organ shape model and traditional shape model. The robustness of the proposed method was also evaluated by several users.

II. METHOD

A. *Experimental data*

Fig. 1 is an example of the abdominal organs in a typical low-dose mouse micro-CT image. As can be observed, the boundaries between the abdominal organs, the liver, spleen, left kidney, and right kidney, are difficult to distinguish. Only the upper boundaries of the liver (i.e., the connection with the bottom of the lungs) and the lower boundaries of the kidneys (i.e., the connection with the fat in the abdomen) are roughly observable by the eye. The low-quality image with grayscale information does not yield enough information to complete the segmentation of multiple organs. However, a priori knowledge of the shape of organs makes it possible to complete the incomplete local boundaries in the image to come up with a better estimate of the desired segmentation. The multi-resolution multi-organ shape model constructed in [16] is created for this purpose.

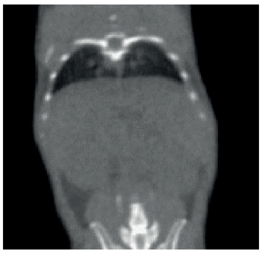
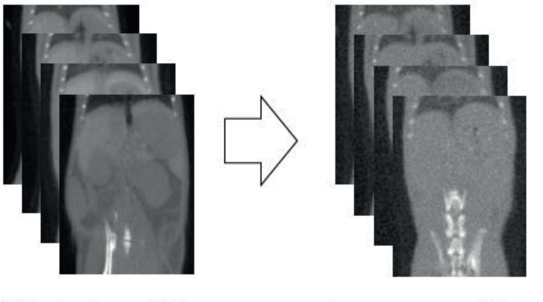


Fig. 1. A typical low-contrast micro-CT image of mouse abdomen.

The low-contrast mouse micro-CT images used in this study contain abdominal organs that are difficult even for human experts to accurately delineate. However, to evaluate the accuracy of the segmentation, it is necessary to know the actual regions of the organs in the image. To tackle this problem, we first exploit the high-contrast images generated in our previous study. (They are micro-CT images obtained from mice injected with contrast agents.) In these images the boundaries of abdominal organs are distinguishable, and experts segmented and labeled the organs from high-contrast images. Next, we decreased the contrast of the labeled organs and added some noise to come up with low-contrast images suitable for the algorithm testing (as shown in Fig. 2). Finally, we compared the segmentation results with the golden standards of experts to evaluate the accuracy of our methods. The source, amount, and acquisition parameters of the images used in this study are the same as in the references [16, 18, 19].

B. *Algorithm*

Since the boundaries in the low-contrast images are blurred, we introduce user interaction to control the shape model [16] to segment desired targets. In this study, a specialized user interface was developed to allow users to draw discrete edges with a computer mouse (Fig. 3 (b)). Exploiting the edges inserted by the user, the VHRBF interpolation algorithm [17] generates the curves bounding organs, see Fig. 3 (c) and Fig. 3 (d). The segmentation speed is greatly improved when users first manually segment the organs. An unavoidable problem is that the VHRBF algorithm shows good segmentation performance for objects with smooth surfaces, while it is not ideal for objects with non-smooth surfaces.



Hight-contrast mouse CT images

Low-contrast mouse CT images

Fig. 2. Generation of low-contrast mouse micro-CT images

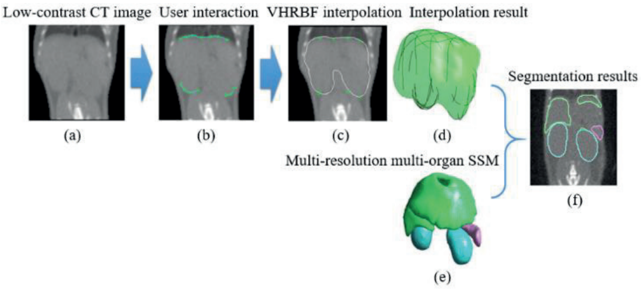


Fig. 3. The workflow of low-contrast mouse micro-CT image segmentation based on multi-resolution multi-organ shape model. (a) The input CT image; (b) Partial edge contours drawn via minor amount of user interaction; (c) The VHRBF interpolation result based on partial user contours in the coronal slice view; (d) Surface rendering of the VHRBF interpolation method based on partial user contours; (e) The multi-resolution multi-organ shape priori model; (f) The segmentation result.

Step 1: In the process of human interaction, the control points of boundaries are collected from the region of interest (ROI) in low-contrast images. The normal vectors of all control points are calculated at the same time, as well as the interpolation coefficients $\alpha_1, \dots, \alpha_j, \dots, \alpha_n, b \epsilon R$, $\boldsymbol{\beta}^1, \dots, \boldsymbol{\beta}^j, \dots, \boldsymbol{\beta}^N, \boldsymbol{a} \epsilon \boldsymbol{R}^3$. The interpolation function, $f(\mathbf{X}) = \sum_{j=1}^{N} \left\{ \alpha_j \left\| \mathbf{X}\text{-}\mathbf{X}^j \right\|^3 \text{-} 3\langle \boldsymbol{\beta}^j, \mathbf{X}\text{-}\mathbf{X}^j \rangle \left\| \mathbf{X}\text{-}\mathbf{X}^j \right\| \right\} + \langle \mathbf{a}, \mathbf{X} \rangle + b$, can be obtained as follows:

$$\sum_{j=1}^{N} \left\{ \alpha_j \left\| \boldsymbol{X}\text{-}\boldsymbol{X}^j \right\|^3 \text{-} 3\langle \boldsymbol{\beta}^j, \boldsymbol{X}\text{-}\boldsymbol{X}^j \rangle \left\| \boldsymbol{X}\text{-}\boldsymbol{X}^j \right\| \right\} + \langle \boldsymbol{a}, \boldsymbol{X} \rangle + b = 0 \tag{1}$$

$$\nabla f\left(\boldsymbol{X}^j\right) = \boldsymbol{n}^j \tag{2}$$

$$\sum_{j=1}^{N} \alpha_j = 0 \tag{3}$$

$$\sum_{j=1}^{N} \left\{ \alpha_j \boldsymbol{X}^j + \boldsymbol{\beta}^j \right\} = 0 \tag{4}$$

where $\boldsymbol{X}$ represents a 3D coordinate, $\boldsymbol{X}^j$ is a 3D coordinate of the control point j of ROI, and $\boldsymbol{n}^j$ is the corresponding normal vector.

Step 2: Perform coordinate transformation and center shift on the deformation model of ROI, so that the center of the deformation model coincides with the center of ROI. The deformation model includes the displacement direction as well as the amplitude of each point. The center point of the deformation model is first shifted to coincide with the center of the corresponding ROI. Then the following transformation is made.

$$x_c = \frac{x_a - x_{a_0}}{l_{x_a}} l_{x_b} + x_{b_0} \tag{5}$$

$$y_c = \frac{y_a - y_{a_0}}{l_{y_a}} l_{y_b} + y_{b_0} \tag{6}$$

$$z_c = \frac{z_a - z_{a_0}}{l_{z_a}} l_{z_b} + z_{b_0} \tag{7}$$

where $\boldsymbol{X_a} = \left(x_a, y_a, z_a\right)$ and $\boldsymbol{X_c} = \left(x_c, y_c, z_c\right)$ are the coordinates before and after transformation, respectively, $\boldsymbol{X_{a_0}} = \left(x_{a_0}, y_{a_0}, z_{a_0}\right)$ is the center of the deformation model before the transformation, $\boldsymbol{X_{b_0}} = \left(x_{b_0}, y_{b_0}, z_{b_0}\right)$ is the center of ROI, $l_{x_a}, l_{y_a}, l_{z_a}$ are the lengths of the deformation model in x, y, and z directions before the coordinate transformation, and $l_{x_b}, l_{y_b}, l_{z_b}$ are the lengths of ROI in the x, y, and z directions, respectively.

Step 3: Substitute the coordinate $\boldsymbol{X_c} = \left(x_c, y_c, z_c\right)$ obtained in step 2 into the interpolation function $f(\boldsymbol{X})$ to obtain the corresponding function values. We use a multithreading mechanism to perform parallel acceleration processing for solving the interpolation function value.

Step 4: Insert $\boldsymbol{X_c} = \left(x_c, y_c, z_c\right)$ into (2) and solve it to obtain the gradient of function f at the point. The process of solving the gradient can be accelerated with a multithreading mechanism and parallel processing.

Step 5: As shown by (1) in step 1, the value $f\left(\boldsymbol{X}^j\right) = 0$ of the interpolation function corresponds to the control point $\boldsymbol{X}^j$ on the edge of ROI. Consequently, the positive and negative values of the interpolation function in step 3 specify whether the deformation model surface approaches the control points outside or inside the ROI. In step 4, the gradients of the interpolation points correspond with the speed and orientation of each move of the deformation model.

Step 6: The deformation equation is as follows:

$$\boldsymbol{X} = \varphi(\boldsymbol{X_1} + \boldsymbol{G}) + (1 - \varphi)\boldsymbol{X_2} \tag{8}$$

here, $\boldsymbol{X_1}$ is the coordinate of the interpolation point on the deformation model before the next iteration. $\boldsymbol{G}$ is the gradient vector field of interpolation points on the deformation model. $\boldsymbol{X_2}$ is the coordinate of the point obtained by fitting the deformation model to $\boldsymbol{X_1} + \boldsymbol{G}$. $\boldsymbol{X}$ is the coordinate of the point on the model after deformation. $\boldsymbol{\varphi}$ is the weight factor, which is 0.8.

Controlled by steps 3, 4, and 5, we use (8) to continuously move and deform the deformation model. The size of the deformation model is adjusted according to the delineated area of ROI during the first deformation. Steps 3, 4, and 5 are used to iteratively calculate the function values and the gradients of interpolation points. Thereafter, the model is modified with these changes bringing the boundary of the model close to that of the ROI. On the boundary points, function f gets a value of zero. Once the iteration stops, a 3D image is generated from the ROI.

Step 7: If the coordinates $\boldsymbol{X}$ obtained depart significantly from the boundary points of the target organs, an additional manual interactive correction will take place. The end-user may add more boundary control points, and then re-run steps 1 to 6 until the users are satisfied with the accuracy of the edge match.

III. RESULTS ANALYSIS

In this study, we chose the high-contrast mouse micro-CT images segmented by experts as the golden standard. To simulate a true segmentation process of multiple organs, we first convert the high-contrast images to grayscale ones. The grayscale values of these images containing contrast agents are reduced to match with images without agents. As a result, we got low-contrast mouse micro-CT images (as shown in Fig. 2). From the low-contrast images, we randomly selected 7 images for segmentation tests.

To compare the difference in segmentation accuracy between the multi-resolution model and the traditional global resolution model, we invoke the VHRBF [17] interpolation method to guide these two models to segment the livers, spleens, and kidneys, respectively. We compare and evaluate the segmentation accuracy of these models in terms of Dice value and average surface distance (ASD) [20] of the segmented results. The method used in this study involves user interaction. In order to evaluate the influence of subjective operations from different users, we invite two operators (A and B) to conduct interactive segmentation tests on the above 7 images. We also use the Dice value and ASD value to evaluate the segmented results of A and B.

Fig. 4 shows the segmentation effect of the multi-resolution multi-organ model and the traditional global resolution model. Fig. 4 exemplifies two representative test images. Each row corresponds to a test image. Fig. 4 (a) is manually labeled results from a high-contrast CT image (i.e., from the golden standard), Fig. 4 (b) is the segmentation results of the multi-resolution multi-organ model, and Fig. 4 (c) is the segmentation results of the traditional global resolution model. We find that the segmentation results of the multi-resolution multi-organ model are closer to the golden standard than the traditional global resolution model, thanks to the better generalization of the multi-resolution model in shape modeling.

In Fig. 5, we compare the segmentation accuracy of the multi-resolution multi-organ model and the traditional global resolution model and use the *t*-test to measure the difference. Asterisks above the horizontal lines of the histograms indicate significant differences between the two data, and no asterisks indicate no significant differences. One asterisk indicates the significance level at which the P-value of the test function is between 0.05 and 0.1. Two asterisks indicate the significance level at which the P-value of the test function is between 0.01 and 0.05, and three asterisks indicate the significance level at which the P-value of the test function is less than 0.01.

When using the multi-resolution multi-organ model and the traditional global resolution model to segment the liver, spleen, and the kidney, the Dice value and ASD value of each organ show that the segmentation accuracy of the multi-resolution multi-organ model outperforms the traditional global resolution model, see Fig. 5 (a) and (b). Fig. 5 (a) shows that there is no significant difference in the Dice value of livers, left kidneys, and right kidneys for both models, but there is a difference in the spleen segmentation. In contrast, Fig. 5 (b) shows that between the models there is no significant difference in the ASD value of the liver, spleen, and left kidney. However, there is a difference in the segmentation of the right kidney.

Fig. 6 shows the differences in segmentation performed by different operators using the multi-resolution multi-organ model. As Fig. 6 (a) shows, there is no significant difference in the Dice value of each segmented organ from different operators. Fig. 6 (b) also shows that the difference in the segmentation results of different operators in terms of ASD value is small. Correspondingly, the evaluation of segmentation results suggests that the methods proposed in this study are robust.

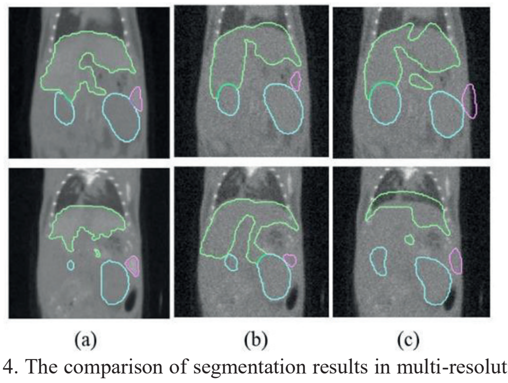

(a) (b) (c)

Fig. 4. The comparison of segmentation results in multi-resolution model and the traditional global resolution model. Each row is a representative test image. (a) Golden standard; (b) Segmentation results of the multi-resolution model; (c) Segmentation results of the traditional global resolution model.

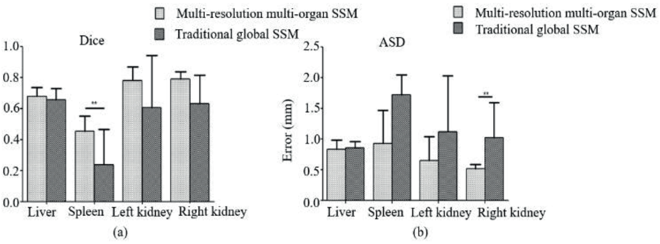


Fig. 5. Comparison of segmentation accuracy of multi-resolution multi-organ model and traditional global resolution model. (a) Dice value comparison; (b) ASD value comparison.

From Fig. 5 and Fig. 6 we conclude that the multi-resolution multi-organ model used in this study yields an advantage compared to the traditional global resolution model in terms of segmentation accuracy. The result can be explained by the better shape-fitting ability of the multi-resolution multi-organ model. We also evaluated this deformation ability in our previous study [16]. Although the algorithm proposed in this study requires a bit of manual interaction, there are no significant differences between the users in the two-subject test. The main reason for this is the limited amount of user interaction needed. In addition, each operator only inserts a small number of edges. The edges and their positions are clear, implying that the sketches drawn by users are close to each other. In this study, the user interaction only plays a guiding role for the model registration, while the actual segmentation is completed by models automatically, thus minimizing the uncertainty caused by different operators.

Although the method in this study performs well in terms of user stability, we still find it has a limited segmentation accuracy for mouse micro-CT images. The Dice values for the liver, left kidney and right kidney are above 0.6, while the Dice value for the spleen is below 0.5. The main factor that affects the segmentation accuracy is the image quality. Compared with the clinical diagnostic CT of humans, the soft tissue contrast of mouse micro-CT images is not on the same level. This results in poorer segmentation accuracy. Even so, the multi-resolution multi-organ segmentation algorithm proposed in this study still achieves higher accuracy than the traditional global resolution model used for human CT image segmentation, confirming the effectiveness of our methods.

In this study, the method proposed was tested on a computer with an Intel Core i7 CPU with 16GB memory. The user interaction process took from 30 seconds to 1 minute, the VHRBF interpolation process required 2.6 seconds, and the shape model registration and segmentation took 15-25

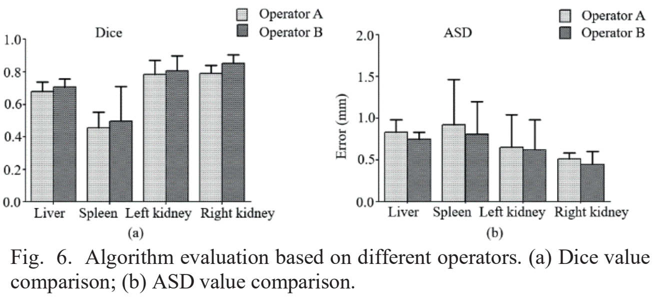


Fig. 6. Algorithm evaluation based on different operators. (a) Dice value comparison; (b) ASD value comparison.

seconds. The whole image segmentation process ran in a total of 1 to 1.5 minutes. This meets the requirements of an interactive segmentation method with a reasonable short waiting time.

IV. CONCLUSION

This study combines VHRBF and the multi-resolution multi-organ model in an innovative way to segment multiple organs simultaneously. We deform the multi-resolution multi-organ model to segment targets in images with a limited number of user interactions. Accordingly, we obtain multiple organs simultaneously in low-contrast mouse micro-CT images. In addition, we get better segmentation accuracy than the traditional global resolution model. On the one hand, combining user interaction with a priori knowledge of the multi-resolution multi-organ model enables successful segmentation for organs with severely missing boundaries. On the other hand, the user tests suggest the stability of the proposed method is very good, indicating that the algorithm is robust. In the next stage, we will consider introducing the segmentation of organs with clear boundaries such as lungs and bones in mouse micro-CT images. In addition, we will try to locate low-contrast organs collaboratively in the abdomen with high accuracy.

ACKNOWLEDGMENT

We thank the Molecular Imaging Center of the University of California, Los Angeles for providing 98 mouse micro-CT images to support our work. We also thank the general program of the National Natural Science Fund of China (No. 81971693, 81401475), the Science and Technology Innovation Fund of Dalian City (2018J12GX042), the Fundamental Research Funds for the Central Universities (DUT19JC01), and the scholarships from China Scholarship Council (No. 201806060163).

[1] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, “Active shape models – their training and application,” Comput. Vis. Image. Und. vol. 61, pp. 38-59, 1995.

[2] T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active appearance models,” IEEE Trans. Pattern. Anal. Mach. Intell. vol. 23, no. 6, pp. 681-685, June 2001.

[3] T. Heimann, H. P. Meinzer, “Statistical shape models for 3D medical image segmentation: a review,” Med. Image. Anal. vol. 13, pp. 543-563, 2009.

[4] H. Lamecker, T. Lange, and M. Seebass, “Segmentation of the liver using a 3D statistical shape model,” 2004.

[5] R. Huang et al., “A locally constrained statistical shape model for robust nasal cavity segmentation in computed tomography,” 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), 2016, pp. 1334-1337.

[6] D. Shen, E. H. Herskovits, and C. Davatzikos, “An adaptive-focus statistical shape model for segmentation and shape modeling of 3-D brain structures,” IEEE Trans. Med. Imaging. vol. 20, no. 4, pp. 257-270, April 2001.

[7] M. B. Stegmann and D. D. Gomez, “A brief introduction to statistical shape analysis,” 2002.

[8] L. Lu, X. Zhang, Y. Zhao, and Y. Jia, “Ear recognition based on statistical shape model,” First International Conference on Innovative Computing, Information and Control - Volume I (ICICIC'06), 2006, pp. 353-356.

[9] D. Kainmueller, H. Lamecker, S. Zachow, and H. C. Hege, “An articulated statistical shape model for accurate hip joint segmentation,” Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. pp. 6345-6351, 2009.

[10] J. H. Noble, R. H. Gifford, R. F. Labadie, and B. M. Dawant, “Statistical shape model segmentation and frequency mapping of cochlear implant stimulation targets in CT,” Med. Image. Comput. Comput. Assist. Interv. vol. 15, pp. 421-428, 2012.

[11] X. Albà et al., “An algorithm for the segmentation of highly abnormal hearts using a generic statistical shape model,” IEEE Trans. Med. Imaging. vol. 35, pp. 845-859, March 2016.

[12] E. A. Audenaert et al., “Cascaded statistical shape model based segmentation of the full lower limb in CT,” Comput. Methods. Biomech. Biomed. Engin. vol. 22, pp. 644-657, May 2019.

[13] M. Aoshima et al., “A survey of high dimension low sample size asymptotics,” Aust. N. Z. J. Stat. vol. 60, pp. 4-19, March 2018.

[14] C. R. Henak, A. E. Anderson, J. A. Weiss, “Subject-specific analysis of joint contact mechanics: application to the study of osteoarthritis and surgical planning,” J. Biomech. Eng. vol. 135, pp. 1-26, February 2013.

[15] M. Wilms, H. Handels, J. Ehrhardt, “Multi-resolution multi-object statistical shape models based on the locality assumption,” Med. Image. Anal. vol. 38, pp. 17-29, May 2017.

[16] Z. Chen, T. Ristaniemi, F. Cong, and H. Wang, “Multi-resolution statistical shape models for multi-organ shape modelling,” Lect. Notes. Comput. Sci. vol. 12557, pp. 74–84, November 2020.

[17] E. Brazil, I. Macedo, M. C. Sousa, L. D. figueiredo, L. Velho, “Sketching variational hermite-rbf implicits,” SBIM. Eurographics Association, pp. 1-8, 2010.

[18] C. E. Suckow, D. B. Stout, “MicroCT liver contrast agent enhancement over time, dose, and mouse strain,” Mol. Imaging. Biol. vol. 10, pp. 114-120, 2008.

[19] H. Wang, D. B. Stout, A. F. Chatziioannou, “Estimation of mouse organ locations through registration of a statistical mouse atlas with micro-CT images,” IEEE Trans. Med. Imaging. vol. 31, pp. 88-102, January 2012.

[20] H. Wang et al., “MARS: a mouse atlas registration system based on a planar x-ray projector and an optical camera,” Phys. Med. Biol. vol. 57, pp. 6063-6077, October 2012.

IV

ANATOMYSKETCH: AN EXTENSIBLE OPEN-SOURCE SOFTWARE PLATFORM FOR MEDICAL IMAGE ANALYSIS ALGORITHM DEVELOPMENT

by

Mingrui Zhuang, **Zhonghua Chen**, Hongkai Wang, Hong Tang, Jiang He, Bobo Qin,
Yuxin Yang, Xiaoxian Jin, Mengzhu Yu, Baitao Jin, Taijing Li, and Lauri Kettunen
2022

Journal of Digital Imaging, 35, 1623–1633

https://doi.org/10.1007/s10278-022-00660-5

Journal of Digital Imaging
https://doi.org/10.1007/s10278-022-00660-5

ORIGINAL PAPER

AnatomySketch: An Extensible Open-Source Software Platform for Medical Image Analysis Algorithm Development

Mingrui Zhuang[1] · Zhonghua Chen[1,2] · Hongkai Wang[1,3] · Hong Tang[1] · Jiang He[1] · Bobo Qin[1] · Yuxin Yang[1] · Xiaoxian Jin[1] · Mengzhu Yu[1] · Baitao Jin[1] · Taijing Li[1] · Lauri Kettunen[2]

Received: 10 October 2021 / Revised: 7 May 2022 / Accepted: 18 May 2022

Abstract

The development of medical image analysis algorithm is a complex process including the multiple sub-steps of model training, data visualization, human–computer interaction and graphical user interface (GUI) construction. To accelerate the development process, algorithm developers need a software tool to assist with all the sub-steps so that they can focus on the core function implementation. Especially, for the development of deep learning (DL) algorithms, a software tool supporting training data annotation and GUI construction is highly desired. In this work, we constructed AnatomySketch, an extensible open-source software platform with a friendly GUI and a flexible plugin interface for integrating user-developed algorithm modules. Through the plugin interface, algorithm developers can quickly create a GUI-based software prototype for clinical validation. AnatomySketch supports image annotation using the stylus and multi-touch screen. It also provides efficient tools to facilitate the collaboration between human experts and artificial intelligent (AI) algorithms. We demonstrate four exemplar applications including customized MRI image diagnosis, interactive lung lobe segmentation, human-AI collaborated spine disc segmentation and Annotation-by-iterative-Deep-Learning (AID) for DL model training. Using AnatomySketch, the gap between laboratory prototyping and clinical testing is bridged and the development of MIA algorithms is accelerated. The software is opened at https://github.com/DlutMedimgGroup/AnatomySketch-Software.

Keywords Medical image analysis · Image annotation · User interaction · Algorithm development · Deep learning

Background

Nowadays, computer-assisted medical image analysis (MIA) algorithms are increasingly used in disease diagnosis and treatment. The development of MIA algorithms is a complex process involving algorithm design, model training, software implementation and performance testing. To speed up the development, researchers need a convenient software platform to assist with different sub-step of the process. Ideally, the platform should include a graphical user interface (GUI) for user interaction and data visualization, as well as a plugin interface for user-developed algorithm integration. Many software tools have been established to meet these needs. Several code libraries were developed to help with algorithm implementation, such as the classical itk [1, 2], vtk [3], elastix [4, 5], ANTS [6] and the recently published libraries for radiomics (e.g. pyradiaomics [7]) and deep learning (e.g. monai[1]). GUI-based software tools were also developed for image segmentation (e.g. ITK-SNAP [8], MITK [9], TurgleSeg [10], Seg3D), data annotation (DicomAnnotator [11] and Pair[2]) and the analysis of specific imaging modalities (e.g. SpheroidJ [12], MNI SISCOM [13] and OIPAV [14]). Few of these tools are extensible for user-developed algorithms including deep neural networks. The 3D Slicer [15] software has powerful extension capabilities and a rich

Mingrui Zhuang and Zhonghua Chen contributed equally to this paper.

✉ Hongkai Wang
wang.hongkai@dlut.edu.cn

1 School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China

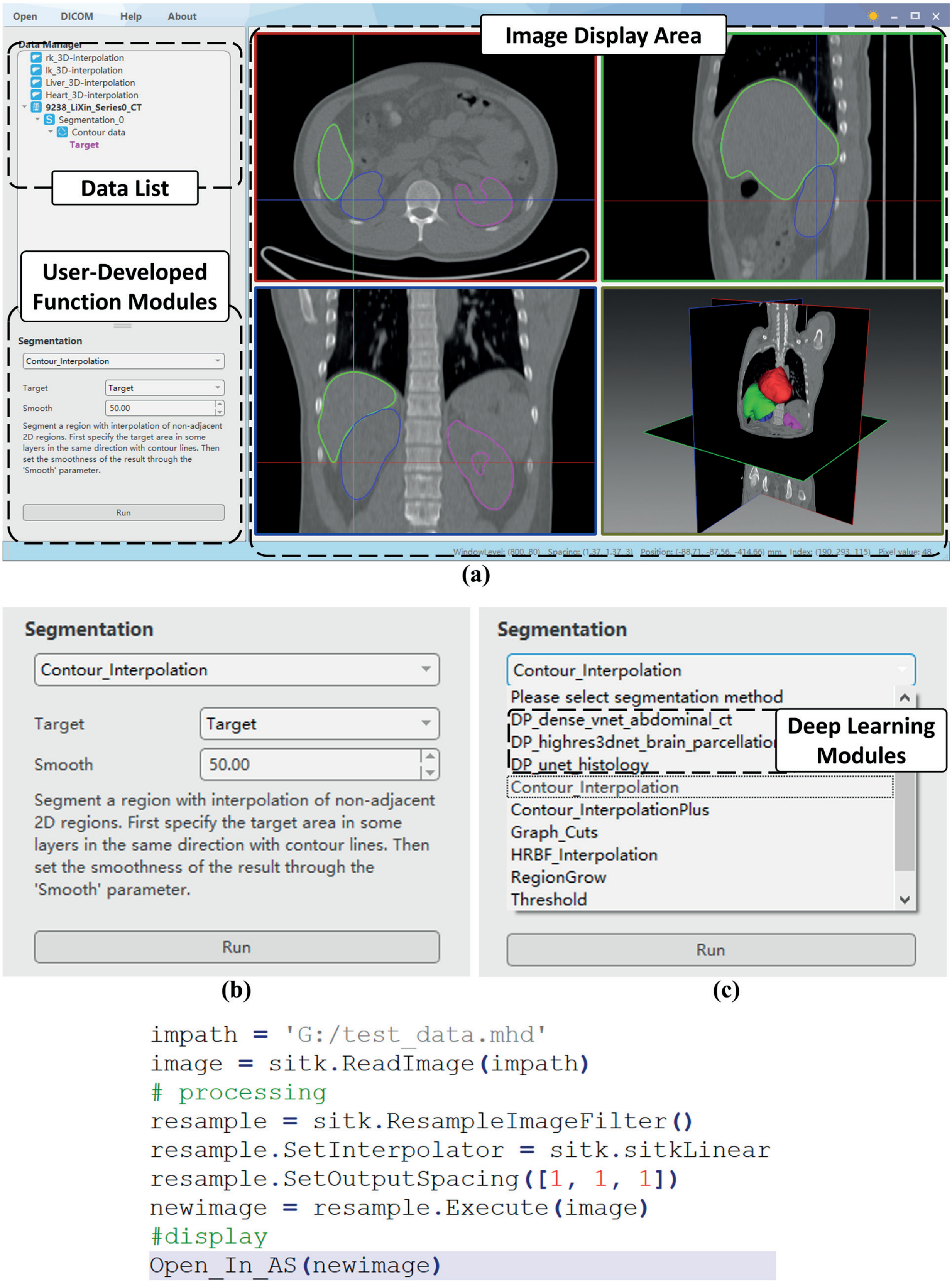
2 Faculty of Information Technology, University of Jyväskylä, 40100 Jyväskylä, Finland

3 Liaoning Key Laboratory of Integrated Circuit and Biomedical Electronic System, Dalian 116024, China

---

[1] Monai: https://monai.io/

[2] Pair: http://www.aipair.com.cn/



Contour_Interpolation

Target: Target

Smooth: 50.00

Segment a region with interpolation of non-adjacent 2D regions. First specify the target area in some layers in the same direction with contour lines. Then set the smoothness of the result through the 'Smooth' parameter.

WindowLevel: (800, 80) Spacing: (1.37, 1.37, 3) Position: (-88.71, -87.56, -414.66) mm Index: (190, 293, 115) Pixel value: 48

(a)

(b)

(c)

```
impath = 'G:/test_data.mhd'
image = sitk.ReadImage(impath)
# processing
resample = sitk.ResampleImageFilter()
resample.SetInterpolator = sitk.sitkLinear
resample.SetOutputSpacing([1, 1, 1])
newimage = resample.Execute(image)
#display
Open_In_AS(newimage)
```

(d)

◂**Fig. 1** **a** AnatomySketch interface. **b** The function module panel. **c** An example of the drop list of user-defined function modules, mostly segmentation methods in this case, including deep network models. **d** An example of calling the software GUI (the highlighted line of code) to visualize intermediate variables.

library of extension modules [16, 17], but its programming mechanism and interaction workflow are relatively complex for junior programmers.

In recent years, deep learning (DL) algorithms are increasingly used in clinical applications. To alleviate the heavy burden of data annotation for DL model training, the procedure of Annotation-by-iterative-Deep-Learning (AID) becomes popular. In the AID workflow, the human experts first annotate a small set of training data which are used for training a preliminary annotation network. The preliminary network is then used to automatically annotate more training data with imperfect accuracy. Next, the human expert proofread the network-annotated data to ensure the annotation accuracy, and the proofread data are supplemented to the training set to retrain a more accurate annotation network. As this process is repeated, the network becomes more and more accurate; thus, fewer and fewer human effort is needed for data annotation. To support the AID workflow, a software platform with convenient annotation/proofreading tools and a plugin interface for the annotation network model is necessary. Philbrick et al. developed the RIL-Contour [18] software with AID function which supports multiuser collaboration and network model version management. However, this software only focuses on data annotation; it is not a general assistance tool for the entire algorithm development workflow.

Besides the needs of a software assisting the classical MIA algorithm development, the recently trend of DL algorithm research also requires a software platform to integrate neural network models and the AID workflow. Due to the lack of such a platform, researchers tediously switch between different assistance tools, making the development process complicated and slow. Moreover, without a GUI-based platform, many algorithms are published as command-line tools or even source codes that are unfriendly to clinical users.

In response to the existing needs, we developed a software platform named AnatomySketch for fast MIA algorithm integration and GUI-based software prototyping. AnatomySketch (AS) offers convenient tools for data annotation, image visualization and algorithm integration, so that the algorithm developers can focus on core function development and rapidly produce a software prototype for algorithm demonstration and testing. Compared to the existing software tools, AnatomySketch has an easier interface for DL model integration and more convenient supports for multi-touch and stylus-based image annotation. The software is developed with the following key features:

(i) *A convenient GUI for data visualization and human–computer interaction.* AnatomySketch has a concise interface for multi-modality and multi-dimensional image visualization. It also incorporates a library of basic processing tools for medical images and graphical models to save the time of basic processing function implementation. To support the development of semi-automated algorithms and the annotation of training data, AnatomySketch provides simple workflow for data annotation mouse, shortcut keys, stylus and multi-touch screen. It also facilitates a simple correction of the automatic segmentation with an inbuilt contour editing method.

(ii) *Flexible plugin interface for user-developed algorithms.* With a flexible plugin interface, the software allows the users to integrate their algorithms (including DL models) as extension function modules. This feature facilitates rapid prototyping of GUI-based software for specific clinical applications, making the evaluation and demonstration of novel algorithms easier and faster. By combining the DL model plugins with the annotation tools, the AID workflow can be realized to speed up DL model training.

The following sections will introduce the detailed software feature and demonstrate exemplar applications of fast software prototyping and AID workflow realization.

Method

Software Design and Architecture

AnatomySketch is designed with a concise GUI consisting of a menu bar (on the top), a data list for data property management (on the top-left), a customizable widget panel for user-developed function modules (on the bottom-left) and a display area (on the right) with three orthogonal section windows and one three-dimensional (3D) view window. The design philosophy of the GUI is being simple, intuitive and familiar to the MIA researchers. The GUI has two operation modes, namely the desktop mode for mouse and keyboard interaction and the tablet mode for stylus and touch screen interaction.

The desktop mode has a classical layout (Fig. 1a) similar to the well-known MITK, ITK-SNAP and 3D Slicer. Distinctively, the layout has a customizable widget panel for user-developed algorithms on the bottom-left (Fig. 1b). A drop list (Fig. 1c) on top of the panel allows the selection of algorithm modules, and the panel layout changes according to the selected algorithm. The algorithm-specific panel layout is defined via a configuration file specifying the positions and appearances of the control widgets (including drop

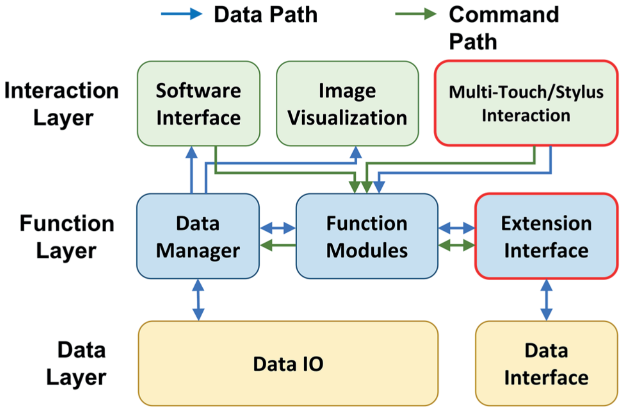


Fig. 2 The architecture diagram of the software platform

menu, text box, press button and explanatory text). Details of the layout definition will be introduced in the "Extension Module" section.

To assist with data visualization during algorithm debugging, we also allow the desktop mode GUI to be called as an inline function of the user programs (e.g. Python, MATLAB or C + +code). This feature is especially useful for accepting user interaction during the algorithm workflow or for inspecting the intermediate variables of the data flow. The intermediate variables can be images, graphical shape models and user-annotations. It is possible to overlay multiple variables in the display window to check the accuracy of image segmentation and/or registration. Figure 1d shows a line of Python code (namely the "Open_In_AS" function) calling the GUI for intermediate image inspection. This Python code is available on the website of AnatomySketch.

The tablet mode is developed to take advantage of the multi-touch screen and the stylus (if available) for efficient annotation of region boundaries. This mode can be activated by clicking expansion button on the top-right of the display windows. The clicked window is enlarged as a palette for stylus sketching. As shown in Fig. 3a, the operator can use the stylus with one hand to drawn contours and scribbles, and meanwhile use the other hand to zoom, pan or rotate the image via the multi-touch screen.

The architecture of the AS software is shown in Fig. 2. The software architecture is composed of three layers including the interaction layer, the function layer and the data layer. The blue and green arrows denote the data and command paths between the modules, respectively. As the core of the software, the function modules coordinate all the other modules to handle data processing and user interactions.

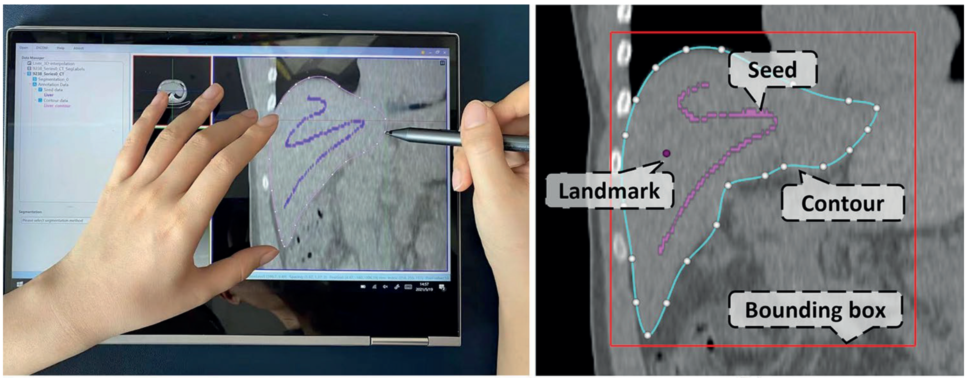
Interactive Annotation and Proofreading

As shown in Fig. 3b, AnatomySketch supports the annotation of anatomical landmarks, bounding boxes, edge contours, curves, scribbles and object regions using the mouse or the stylus. All the annotations can be accessed by the user-developed plugin modules as inputs. The annotations can also be exported into computer discs as separate files for offline algorithm training.

AnatomySketch also provides a convenient boundary correction tool for proofreading the segmentation results of automatic algorithms. This tool is implemented based on the free-form deformation (FFD) method [19]. It allows the user to drag the inaccurate boundary towards the correct position (as shown in Fig. 4). We applied both 2D and 3D versions of the FFD method for adjusting 2D contours and 3D surfaces, respectively.

FFD is a point-controlled contour deformation method. After the user dragging operation, a 6×6 grid is constructed around the starting point of the dragging. The mouse/stylus motion vector (yellow arrow in Fig. 4) is first extrapolated to the 6×6 grid vertices by solving an inverse interpolation function of the cubic B-spline. Then, the deformation vectors of contour points are interpolated by solving the cubic B-sample interpolation function,

Fig. 3 Interactive data annotation. **a** The tablet mode layout supporting stylus sketching and multi-touch gestures. **b** Multiple annotation tools are provided by the software



(a) (b)

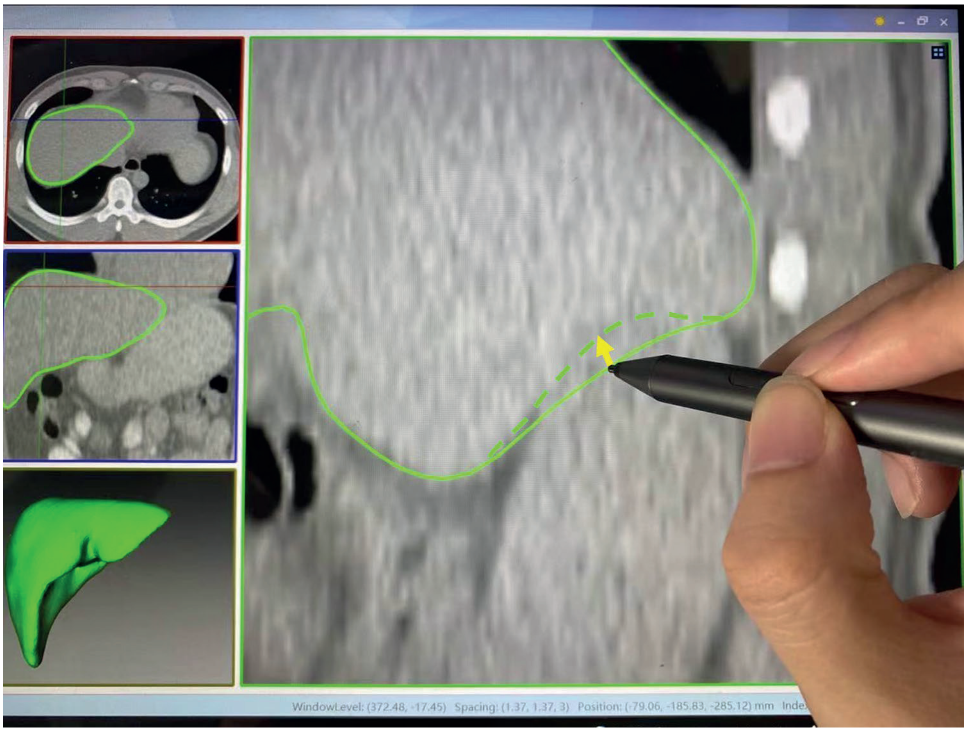


Fig. 4 Proofreading of the automatic segmentation result. The inaccurate boundary can be dragged towards the correct position (the dashed curve) using the stylus or the mouse

$$\mathrm{p} = \sum_{i=0}^{3} \binom{3}{i} (1-m)^{3-i} m^{i} \left(\sum_{j=0}^{3} \binom{3}{j} (1-n)^{3-j} n^{j} \mathrm{p}_{i,j} \right) \tag{1}$$

where p is the interpolated deformation vector of a contour vertex, (m, n) is the normalized local coordinate of the dragging start point, and $\mathrm{p}_{i,j}$ is the deformation vector of grid node (i, j). The readers are referred to [19] for more details of the method.

Extension Module

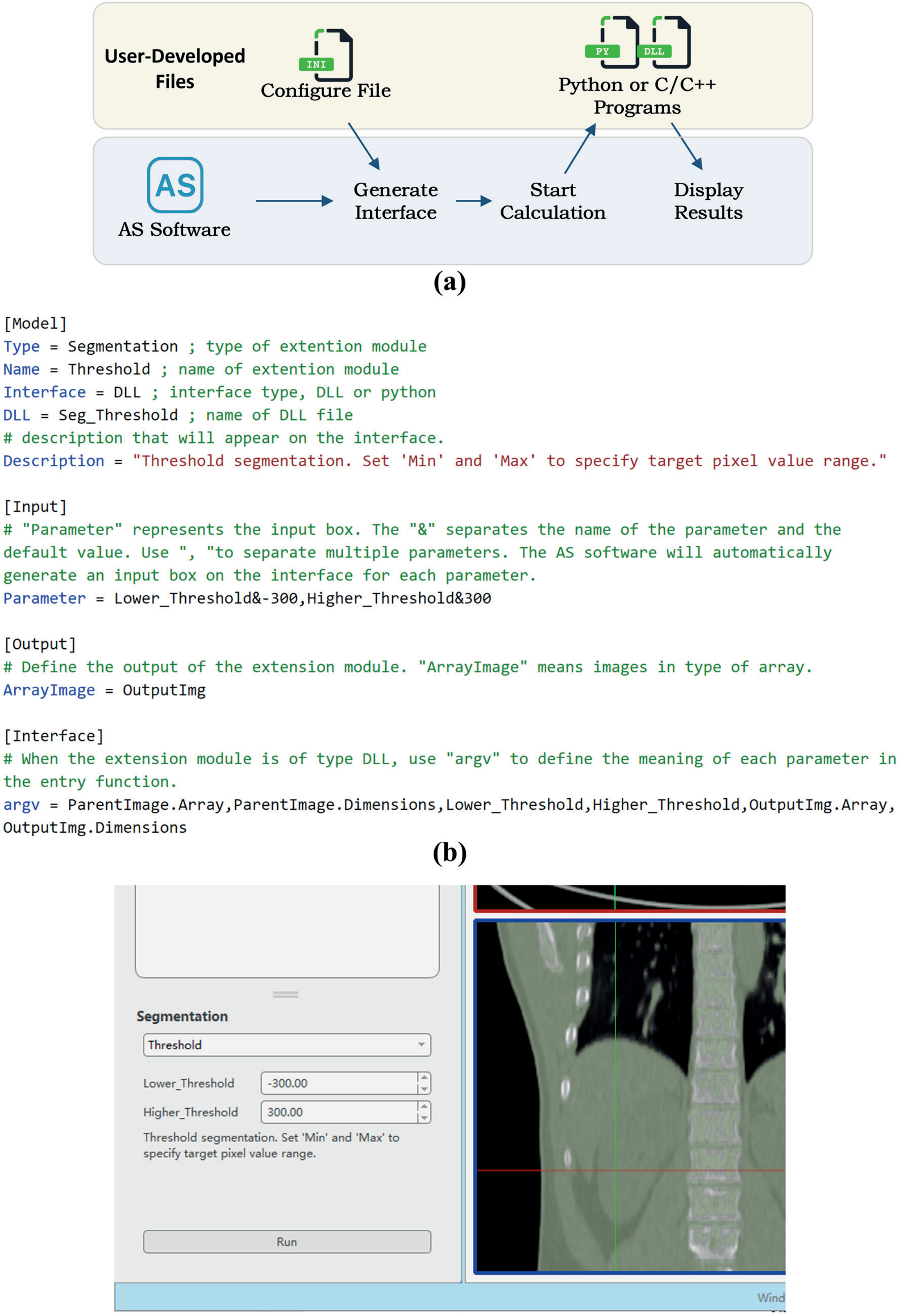
AnatomySketch provides a flexible plugin interface for the integration of user-developed algorithms. Figure 5a shows the workflow of the extension modules. A configuration file and a program file simply form the plugin interface of the software. The configuration file is a text-format definition file specifying the extension module type, input and output parameters and the GUI design. The software will read this definition file and realize the extension interface accordingly. The program file is the user-programmed executable (exe) file or a dynamic link library (dll) of the core algorithm function. Figure 5b shows an example of the configuration file for thresholding segmentation. The configuration file is imported into the software to create the customized widget panel (Fig. 5c). The program file is invoked by clicking the “Calculate” button on the widget panel. To transfer input and output data between the software and the program file, a loose coupling mechanism is adopted. The software first writes the input data (i.e. image arrays, polygonal meshes or annotations) into the computer disc, then the user program imports them for computation and writes the outputs to the computer disc. Finally, the software gets the output results from the disc and updates the GUI display.

On the AnatomySketch website, several quick-start templates of the configuration files and program files are provided. The developers can publish their algorithm modules on the website to promote the usage of their works.

Deep Learning Support

Thanks to the flexible plugin interface, deep neural networks can be integrated into the software as plugin modules. The developer needs to create a Python-based program file to get input data from the software and calls the network model to process the data. The network can be developed using any DL platform (e.g. PyTorch, TensorFlow or Keras) and compiled as an executable file to be called by the user-created Python program. The outputs of the network are written into the hard drive and then imported into the software via the plugin interface.

In AnatomySketch, the AID workflow is realized by combining the annotation tools, the proofreading tools and the plugin modules. Rich annotation tools of AnatomySketch are used for labelling the primary training data, then a user-defined Python program is called to train the network and use the trained model to label more images. During the proofreading, the 2D and 3D FFD tools are used to correct



(a)

```
[Model]
Type = Segmentation ; type of extention module
Name = Threshold ; name of extention module
Interface = DLL ; interface type, DLL or python
DLL = Seg_Threshold ; name of DLL file
# description that will appear on the interface.
Description = "Threshold segmentation. Set 'Min' and 'Max' to specify target pixel value range."

[Input]
# "Parameter" represents the input box. The "&" separates the name of the parameter and the
default value. Use ", "to separate multiple parameters. The AS software will automatically
generate an input box on the interface for each parameter.
Parameter = Lower_Threshold&-300,Higher_Threshold&300

[Output]
# Define the output of the extension module. "ArrayImage" means images in type of array.
ArrayImage = OutputImg

[Interface]
# When the extension module is of type DLL, use "argv" to define the meaning of each parameter in
the entry function.
argv = ParentImage.Array,ParentImage.Dimensions,Lower_Threshold,Higher_Threshold,OutputImg.Array,
OutputImg.Dimensions
```

(b)

Segmentation

Threshold

Lower_Threshold -300.00

Higher_Threshold 300.00

Threshold segmentation. Set 'Min' and 'Max' to specify target pixel value range.

(c)

◂**Fig. 5** The extension module. **a** The workflow of the extension module. **b** An example of the configuration file. **c** The widget panel generated by AnatomySketch according to the configuration file of (**b**)

the segmentation errors. The entire AID workflow is customized as a plugin module with a widget panel supporting iterative training and proofreading. Video 1 attached to this paper demonstrates the AID support features of AnatomySketch.

Results

In this section, we will demonstrate the examples of using AnatomySketch for fast plugin module development and software prototyping. We will also show examples of human-AI collaborated image segmentation and AID-based data annotation. These examples demonstrate the convenience of integrating DL models or interactive algorithms into our software, while such integration can be time consuming or even infeasible for the existing tools.

Software Prototyping

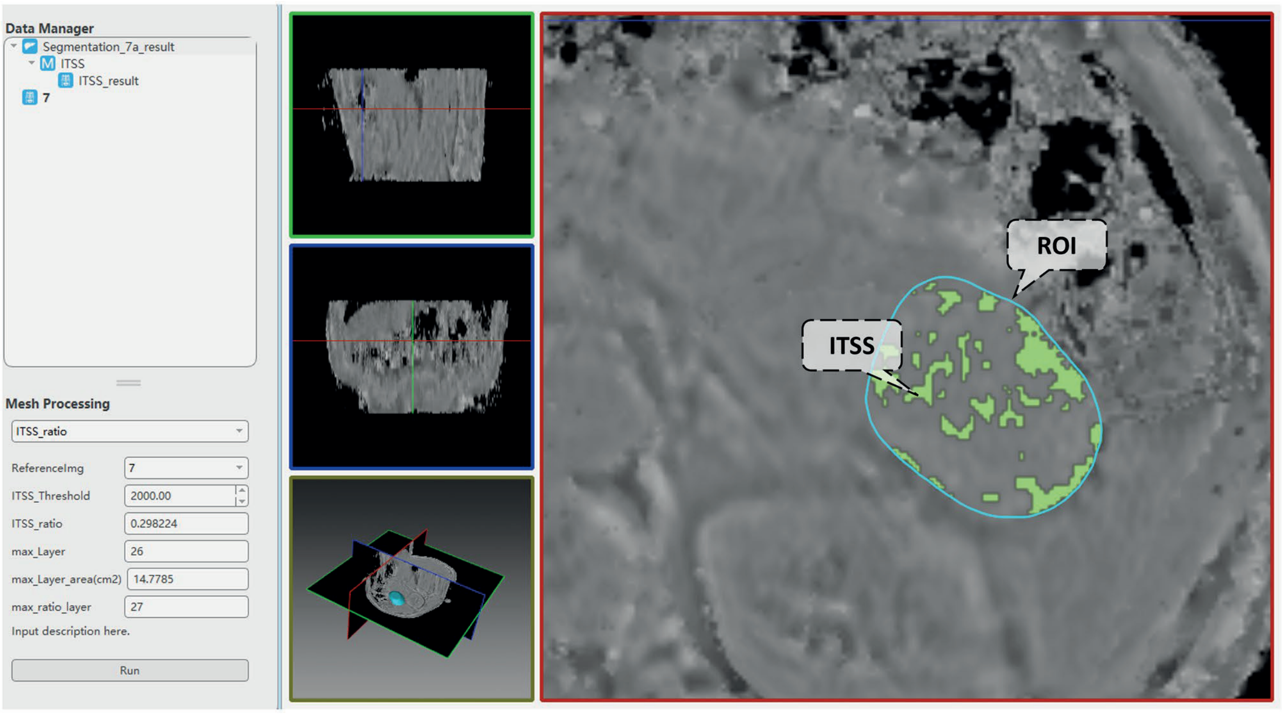
In the first example, AnatomySketch was used in a medical research project for analyzing intratumoral susceptibility signal intensities (ITSS) in enhanced T2 angiography of hepatocellular carcinoma. Because this study uses special MRI pulse sequences for ITSS imaging, customized software needs to be developed for the data analysis. As a collaborator of this study, our group took only 1 day to develop a plugin module and created a GUI-based software prototype for the ITSS MR image analysis. With our software prototype, the doctors contoured the region of interest (ROI) in a few interleaved axial slices and interpolated the 3D surface from the 2D contours. The proofreading tool was used to adjust the 3D surface to precisely fit the tumour boundary. The software automatically removed the MR imaging artefact and extracted the high-ITSS voxels via thresholding. The algorithm parameters such as ITSS threshold can be adjusted in the customized widget panel. Figure 6 shows the software GUI. The blue contour represents the ROI boundary and the green pixels represent extracted high-ITSS voxels. This tool has been used in a series of published studies [20–23]. Although the development of this module is simple and straightforward in the AS platform, similar extension may require time-consuming software recompilation in some existing tools (e.g. for MITK and ITKsnap).

In another example, AnatomySketch was used to annotate lung lobes in CT images for training a lobe segmentation network of commercial software. Engineers from a local company created an interactive lung lobe annotation plugin for AnatomySketch. The contours of lung fissures are manually sketched in a few coronal slices and a complete fissure surface was interpolated from the contours using radial basis function (RBF) interpolation. The proofreading tools of AnatomySketch were used to adjust the interpolated fissure surfaces. Figure 7 shows the GUI of the plugin module; VIDEO 2 attached to this paper exhibits the working process of this plugin module. Using this module, the annotation of five lung lobes took less than 20 min per image. Two engineers from the company annotated 100 CT images in 3 days, thanks to the stylus support of our software. In contrast, stylus interaction is not specially optimized in any other existing tool.

Deep Learning Supports

Through the plugin interface, user-developed deep network models are integrated into AnatomySketch as extension modules. As shown in the example of Fig. 8, a dense V-Net model [24] was trained to segment the intervertebral disc and the surrounding nerves and vessels from lumbar CT images. Due to the lack of enough training data, the network occasionally produced inaccurate segmentation at the fuzzy boundary of the herniated discs (Fig. 8a). Thanks to the 3D FFD proofreading tool, a human expert was able to correct the inaccurate segmentation within 5 min per image (Fig. 8b). In this way, the AI model and human expert collaborate with each other to achieve efficient and accurate segmentation of anatomical objects with weak boundaries. As a comparison, none of the existing medical image processing tools facilitate such efficient and direct proofreading of the DL segmentation results. Both MITK and 3D Slicer provide the AI-assisted annotation plugins, but they require Internet connection to the NVIDIA AI-Assisted Annotation Server for data transfer, which is inconvenient for the applications without Internet connections or with data privacy concerns.

Another example of human-AI interaction is the realization of AID workflow. We trained a DeepSnake network [25] to segment abdominal organs from CT images. The network generates 2D contours surrounding the target organs and deforms the contours to fit the organ boundaries. We first used a small set of expert-labelled training images to train a preliminary network, then used the preliminary network to generate the organ contours of more images. The automatically generated contours are proofread by human experts using the 2D FFD tool of AnatomySketch, and the images with proofread contours are supplemented to the training set to finetune the network. Figure 9 displays the predicted contours of the preliminary and retrained models for two representative slices, respectively. It is obvious that the retrained network (trained with



ITSS_ratio

ReferenceImg 7

ITSS_Threshold 2000.00

ITSS_ratio 0.298224

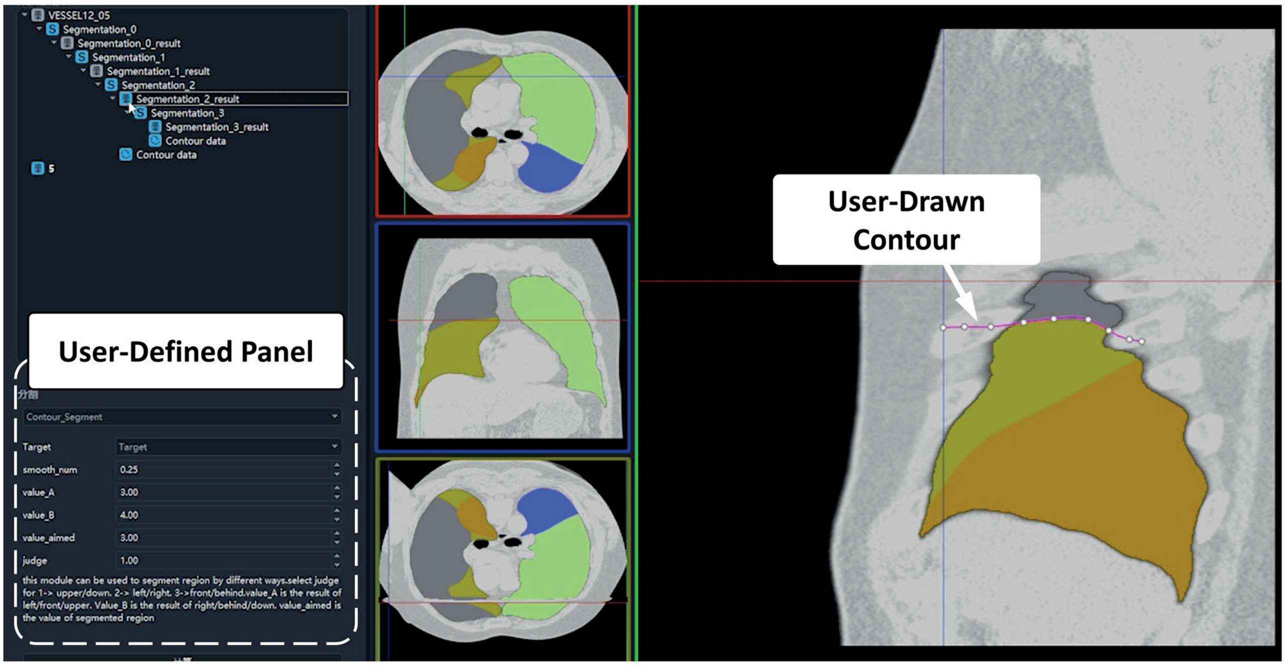
max_Layer 26

max_Layer_area(cm2) 14.7785

max_ratio_layer 27

Input description here.

Fig. 6 An example of user-defined extension module for ITSS analysis in MR images



Contour_Segment

Target Target

smooth_num 0.25

value_A 3.00

value_B 4.00

value_aimed 3.00

judge 1.00

this module can be used to segment region by different ways.select judge for 1-> upper/down. 2-> left/right. 3->front/behind.value_A is the result of left/front/upper. Value_B is the result of right/behind/down. value_aimed is the value of segmented region

Fig.7 An example of user-developed plugin module for lung lobe annotation in CT images

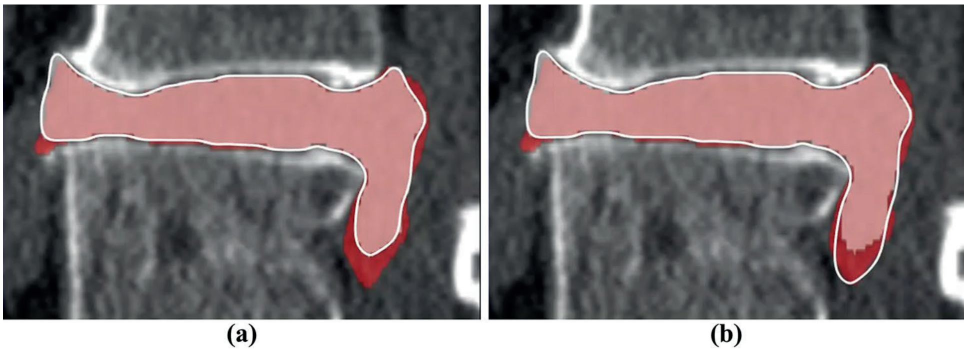


Fig. 8 An example of 3D FFD proofreading. **a** The pink area in white contour is the automatic segmentation result of V-Net model. The red area and the white contour depict the under-segmented part. **b** Human expert proofreading result (the adjusted white contour) using the FFD tool

about 1800 slices of 15 CT series) yields more accurate contour prediction than the preliminary network (trained using about 600 slices of five CT series).

Discussions

AnatomySketch is developed with the goal of accelerating MIA algorithm development and bridging the gap between laboratory research and clinical application. The software is designed to meet many specific needs of MIA algorithm development:

(i) *Software prototyping.* During the past decade, many MIA algorithms were developed as command-line tools due to the difficulty of GUI construction. This problem also hampers the promotion of DL-based algorithms. Many DL models are published as source codes that cannot be tested by clinical users. AnatomySketch tackles this problem with a flexible plugin interface. As demonstrated in the examples of Figs. 6 and 8, the prototyping of a new software tool for special image modality (e.g. the ITSS MR) or specific application task (e.g. the segmentation of lumbar disc herniation) took less than 1 day. With a friendly GUI, the promotion of new MIA algorithms in the clinical environment becomes faster and more convenient.

(ii) *Data annotation.* The fast popularization of deep learning techniques proposes strong needs for data annotation. AnatomySketch provides annotation

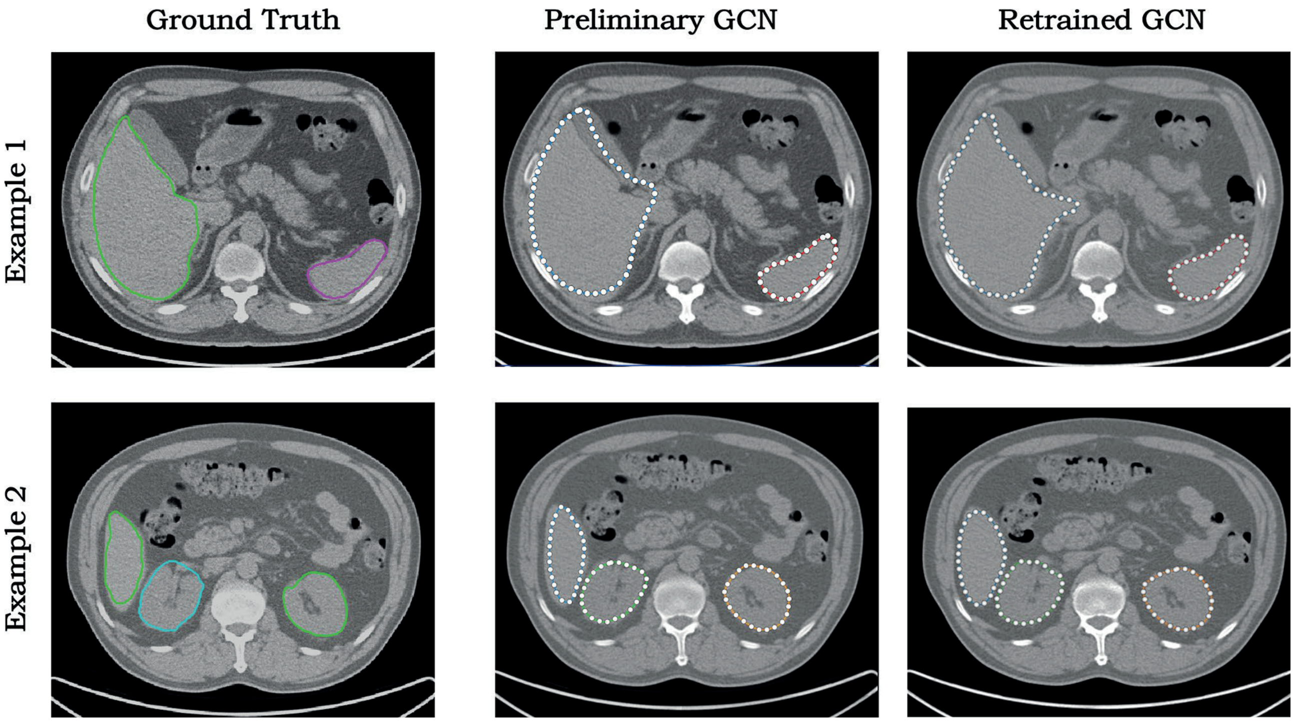


Fig. 9 AID annotation results of two exemplar CT slices, showing that the retrained network yield more accurate results than the preliminary network. The ground truth comes from human expert manual labelling

tools and the tablet mode for efficient image annotation. It also realizes the AID workflow for iterative annotation and training. An AID workflow is developed for abdominal organ segmentation based on the DeepSnake network. Thanks to the flexible plugin interface, AnatomySketch is versatile enough for realizing task-specific AID workflow based on different network models.

(iii) *Human–computer interaction.* Because fully automatic algorithms may not guarantee robustness in complex clinical scenario, semi-automatic algorithms with human guidance and corrections become practical choices. As shown in the example of Fig. 7, the plugin module interpolates the lung fissure surfaces using human-annotated curves as the guidance. Figure 9 demonstrates another example in which human interaction is introduced to correct the DL model output. AnatomySketch supports both input guidance and posterior correction to the DL models, facilitating flexible human-AI collaboration for clinical image analysis.

When comparing AnatomySketch with existing MIA software tools, we find that some key features of AnatomySketch have been integrated into the existing tools. The extension module is also available in the Slicer software [15, 16]. The integration of DL models has been realized by the MITK NVIDIA clara plugin[3] and the RIL-contour software [18]. The AID workflow is also supported by the RIL-contour software. However, these software tools were only designed to assist with a certain step of the entire algorithm development workflow. The advantage of AnatomySketch is the supporting of the complete workflow, including data visualization, image annotation, algorithm integration and software prototyping.

As a newly established software, the number of user-developed plugin modules for AnatomySketch is still growing, especially for specific clinical applications. We will keep maintaining the web community to help the developers sharing their plugin modules and gain potential users from universities and hospitals. We also plan to add online crowdsourcing tools for multi-rater annotation and proofreading. Moreover, because AnatomySketch is increasingly used by the doctors who do not share the medical images, we will incorporate federated learning [26] ability for multi-centre model training without sharing confidential medical data. A plugin module will be developed to allow AS software to communicate with multi-centre client databases and to invoke models for inference. In this scenario, physicians at each centre can use the tools provided in AS to annotate images and invoke network models with AS interface.

Conclusion

We developed a medical image analysis software platform named AnatomySketch to assist with MIA algorithm development. The software is specially designed for efficient image annotation and convenient integration of user-developed algorithm modules including deep neural networks. The AID workflow can also be realized to accelerate the training of DL models. For the next step, we will construct a web community for sharing user-developed extension modules and incorporate federated learning to facilitate mutual learning between DL models from multi-centres.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10278-022-00660-5.

Author Contribution Mingrui Zhuang and Zhonghua Chen contributed equally to this paper.

Funding This work was supported in part by the National Key Research and Development Program No. 2020YFB1711500, 2020YFB1711501 and 2020YFB1711503; the general program of National Natural Science Fund of China (No. 81971693, 61971445 and 61971089); Dalian City Science and Technology Innovation Funding (No. 2018J12GX042); the Fundamental Research Funds for the Central Universities (No. DUT19JC01 and DUT20YG122); the funding of Liaoning Key Lab of IC & BME System and Dalian Engineering Research Center for Artificial

Availability of Data and Material The Declarations

Code Availability The Anatomy Sketch software is opened at https://github.com/DlutMedimgGroup/AnatomySketch-Software.

Declarations

Conflict of Interest The authors declare no competing interests.

---

[3] MITK NVIDIA clara plugin: https://www.mitk.org/wiki/MITK_ReleaseNotes_2018.04.2

References

1. Yoo, T.S.; Ackerman, M.J.; Lorensen, W.E.; Schroeder, W.; Chalana, V.; Aylward, S.; Metaxas, D.; Whitaker, R. Engineering and algorithm design for an image processing Api: a technical report on ITK--the Insight Toolkit. *Stud Health Technol Inform* **2002**, *85*, 586-592.
2. McCormick, M.; Liu, X.; Jomier, J.; Marion, C.; Ibanez, L. ITK: enabling reproducible research and open science. *Frontiers in neuroinformatics* **2014**, *8*, 13, https://doi.org/10.3389/fninf.2014.00013.
3. Schroeder, W.; Martin, K.; Lorensen, B. *The Visualization Toolkit (4th ed.)*; Kitware: 2006.
4. Klein, S.; Staring, M.; Murphy, K.; Viergever, M.A.; Pluim, J.P.W. elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Transactions on Medical Imaging* **2010**, *29*, 196-205, https://doi.org/10.1109/TMI.2009.2035616.
5. Shamonin, D.; Bron, E.; Lelieveldt, B.; Smits, M.; Klein, S.; Staring, M. Fast Parallel Image Registration on CPU and GPU for Diagnostic Classification of Alzheimer's Disease. *Frontiers in Neuroinformatics* **2014**, *7*, https://doi.org/10.3389/fninf.2013.00050.
6. Avants, B.; Tustison, N.; Song, G. Advanced normalization tools (ANTS). *Insight J* **2008**, *1–35*.
7. van Griethuysen, J.J.M.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillion-Robin, J.-C.; Pieper, S.; Aerts, H.J.W.L. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research* **2017**, *77*, e104-e107, https://doi.org/10.1158/0008-5472.Can-17-0339.
8. Yushkevich, P.A.; Gao, Y.; Gerig, G. ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 16–20 Aug. 2016, 2016; pp. 3342–3345.
9. Wolf, I.; Vetter, M.; Wegner, I.; Böttger, T.; Nolden, M.; Schöbinger, M.; Hastenteufel, M.; Kunert, T.; Meinzer, H.-P. The Medical Imaging Interaction Toolkit. *Medical Image Analysis* **2005**, *9*, 594-604, https://doi.org/10.1016/j.media.2005.04.005.
10. Poon, K.; Hamarneh, G.; Abugharbieh, R. *Segmentation of complex objects with non-spherical topologies from volumetric medical images using 3D livewire*; SPIE: 2007; Volume 6512.
11. Dong, Q.; Luo, G.; Haynor, D.; O'Reilly, M.; Linnau, K.; Yaniv, Z.; Jarvik, J.G.; Cross, N. DicomAnnotator: a Configurable Open-Source Software Program for Efficient DICOM Image Annotation. *J Digit Imaging* **2020**, *33*, 1514-1526, https://doi.org/10.1007/s10278-020-00370-w.
12. Lacalle, D.; Castro-Abril, H.A.; Randelovic, T.; Dominguez, C.; Heras, J.; Mata, E.; Mata, G.; Mendez, Y.; Pascual, V.; Ochoa, I. SpheroidJ: An Open-Source Set of Tools for Spheroid Segmentation. *Comput Methods Programs Biomed* **2021**, *200*, 105837, https://doi.org/10.1016/j.cmpb.2020.105837.
13. Moreau, J.T.; Saint-Martin, C.; Baillet, S.; Dudley, R.W.R. MNI SISCOM: an Open-Source Tool for Computing Subtraction Ictal Single-Photon Emission CT Coregistered to MRI. *J Digit Imaging* **2021**, *34*, 357-361, https://doi.org/10.1007/s10278-021-00422-9.
14. Zhang, L.; Xiang, D.; Jin, C.; Shi, F.; Yu, K.; Chen, X. OIPAV: an Integrated Software System for Ophthalmic Image Processing, Analysis, and Visualization. *J Digit Imaging* **2019**, *32*, 183-197, https://doi.org/10.1007/s10278-017-0047-6.
15. Pieper, S.; Halle, M.; Kikinis, R. 3D Slicer. In Proceedings of the 2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821), 18–18 April 2004, 2004; pp. 632–635 Vol. 631.
16. Fedorov, A.; Beichel, R.; Kalpathy-Cramer, J.; Finet, J.; Fillion-Robin, J.C.; Pujol, S.; Bauer, C.; Jennings, D.; Fennessy, F.; Sonka, M.; et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging* **2012**, *30*, 1323-1341, https://doi.org/10.1016/j.mri.2012.05.001.
17. Pinter, C.; Lasso, A.; Wang, A.; Sharp, G.C.; Alexander, K.; Jaffray, D.; Fichtinger, G. Performing radiation therapy research using the open-source SlicerRT toolkit. In *World Congress on Medical Physics and Biomedical Engineering, June 7–12, 2015, Toronto, Canada*; IFMBE Proceedings; 2015; pp. 622–625.
18. Philbrick, K.A.; Weston, A.D.; Akkus, Z.; Kline, T.L.; Korfiatis, P.; Sakinis, T.; Kostandy, P.; Boonrod, A.; Zeinoddini, A.; Takahashi, N.; et al. RIL-Contour: a Medical Imaging Dataset Annotation Tool for and with Deep Learning. *J Digit Imaging* **2019**, *32*, 571-581, https://doi.org/10.1007/s10278-019-00232-0.
19. Sederberg, T.W.; Parry, S.R. Free-form deformation of solid geometric models. In Proceedings of the Proceedings of the 13th annual conference on Computer graphics and interactive techniques, 1986; pp. 151–160.
20. Cui, D.; Liu, A.; Wang, H.; Zhuang, M.; Zhao, Y.; Song, Q. Value of intratumoral susceptibility signal intensities in quantitatively and automatically evaluating histological grading of hepatocellular carcinoma using enhanced T2 star-weighted angiography. In Proceedings of the ISMRM 2021, Vancouver, Canada, 2021; p. 4364.
21. Cui, D.; Liu, A.; Wang, H.; Zhuang, M.; Song, Q. The combination of ITSS and R2* in quantitatively and automatically evaluating histological grade of HCC using ESWAN: A feasibility study. In Proceedings of the ISMRM 2021, Vancouver, Canada, 2021; p. 4645.
22. Hu, W.; Liu, A.; Li, Y.; Wang, H.; Zhuang, M.; Song, Q. Evaluation of R2* and automatically quantitative ITSS in diagnosis of malignant ovarian tumor. In Proceedings of the ISMRM 2021, Vancouver, Canada, 2021; p. 3758.
23. Liu, Y.; Wang, H.; Zhuang, M.; Chen, L.; Song, Q.; Meng, S.; Liu, A. Differential diagnosis of PCa and BPH using intratumoral susceptibility signal intensities based on ESWAN. In Proceedings of the ISMRM 2021, Vancouver, Canada, 2021; p. 3723.
24. Gibson, E.; Giganti, F.; Hu, Y.; Bonmati, E.; Bandula, S.; Gurusamy, K.; Davidson, B.; Pereira, S.P.; Clarkson, M.J.; Barratt, D.C. Automatic Multi-Organ Segmentation on Abdominal CT With Dense V-Networks. *IEEE Trans Med Imaging* **2018**, *37*, 1822-1834, https://doi.org/10.1109/TMI.2018.2806309.
25. Peng, S.; Jiang, W.; Pi, H.; Li, X.; Bao, H.; Zhou, X. Deep Snake for Real-Time Instance Segmentation. 2020, pp 8530–8539.
26. Pfitzner, B.; Steckhan, N.; Arnrich, B. Federated Learning in a Medical Context: A Systematic Literature Review. *ACM Transactions on Internet Technology* **2021**, *21*, 1-31, https://doi.org/10.1145/3412357.