

# **Novel Image Processing and Deep Learning Methods for Liver Cancer Delineation from CT Data**



Ming Gong

Department of Electronic and Electrical Engineering

University of Strathclyde

Glasgow, United Kingdom

A thesis submitted for the degree of

Doctor of Philosophy

2025

# Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in or derived from this thesis.

Ming Gong

2025

## **Acknowledgements**

Throughout my graduate studies and the process of writing this thesis, I have received invaluable support and assistance from many individuals. I would like to express my deepest gratitude to all those who have contributed to this journey.

First and foremost, I am deeply grateful to my supervisor, Professor John Soraghan, for his continuous encouragement and guidance throughout the PhD project. His critical thinking, enthusiasm, and knowledge provide me with strong help for my PhD time and further life.

Second, I would like to express my heartfelt gratitude to my second supervisor, Dr Gaetano Di-Caterina, for his exceptional support through my PhD life. His humour and humility gave me a lot of support while writing my thesis, just like my friend. I sincerely thank Dr Derek Grose, from Beatson West of Scotland Cancer Centre, Glasgow, for his help with clinical knowledge, and contributions to this research.

I am grateful to my friends and colleagues at the Centre of Signal and Image Processing, especially Dr Baixiang Zhao, Weijie Ke, Dr Yannan Xing, Dr Amlan Basu, Dr Paul Kirkland and Dr Keerati Kaewrak for their friendship and knowledge exchange.

I sincerely thank my family for their unconditional love and support. I appreciate everything my parents have done for me so I can pursue my studies. I am also grateful to my girlfriend Ms Xuelian Huang for her love in these years.

## **Abstract**

One of the important significances of medical image segmentation is its key role in personalized medicine and precision treatment. With the advancement of medical imaging technology, three-dimensional visualization and quantitative analysis of the liver and its lesion areas have been achieved. By accurately segmenting and identifying anatomical structures and lesion areas, the processed images can provide high-resolution and quantitative anatomical information to help clinicians develop more effective and personalized treatment plans. In liver surgery or tumour radiotherapy, accurate image segmentation can help plan surgical paths or ensure that drugs are accurately concentrated in the tumour area while avoiding healthy tissues and organs at risk (OARs), thereby improving treatment efficacy and reducing side effects. In the traditional liver segmentation process, radiologists need to manually depict the liver contour, abnormal lesions in the liver (such as cancerous tumour areas), and other key anatomical structures. However, this manual depiction process is easily affected by differences between observers and the same observer at different times, which may affect the accuracy of treatment. Therefore, with the help of advanced imaging and computing technologies, it is of practical significance and value to achieve automated and high-precision segmentation of the liver and liver cancer areas in clinical applications. This thesis designed a series of innovative deep learning methods to achieve automatic segmentation of the liver and liver cancer areas in computed tomography (CT) images, focusing on improving segmentation accuracy, robustness, and the ability to handle small tumour areas. The datasets used in this study are from the public liver segmentation dataset, The Liver Tumour Segmentation (LiTS).

In view of the liver CT segmentation problem, this study proposed an improved U-Net model based on multi-scale feature fusion, which significantly improved the segmentation performance. The model combines dilated convolution and pyramid pooling modules to enhance the ability to capture multi-scale features. Among them, dilated convolution improves the model's sensitivity to fine-grained features without

increasing the number of parameters by expanding the receptive field; the pyramid pooling module enhances the recognition ability of complex liver anatomical structures by aggregating multi-scale global features. Experimental results show that the multi-scale feature fusion U-Net performs well in the liver segmentation task, with an average Dice coefficient of 0.95, and still shows stability and robustness when dealing with cases with blurred boundaries.

In the liver cancer segmentation task, this study proposed two cascaded U-Net networks based on the attention mechanism, which solved the challenges in this field with different strategies. The first method introduces a custom attention mechanism module based on the cascaded U-Net to simulate the characteristics of humans automatically focusing on cancerous areas during the segmentation process, which significantly improves the accuracy of segmentation. The model can automatically focus on the key features of the tumour in the decoding stage while suppressing the interference of background noise, providing effective support for accurate segmentation.

The second method proposes an end-to-end trained cascade hybrid attention network, whose main goal is to directly complete the liver cancer segmentation task through a single network, avoiding the process of using two independent networks in traditional methods. The network achieves precise focus on the liver region by adding a Bounding Box module based on the hard attention mechanism between two U-Nets. At the same time, the channel attention module and the spatial attention module are introduced in the feature extraction stage, so that the model can effectively capture the key features of the tumour region and further improve the segmentation performance. The improved U-Net model performs well in dealing with tumour regions with complex shapes and irregular boundaries. Experimental data show that both cascade U-Nets based on the attention mechanism show excellent performance in the liver cancer segmentation task, maintaining high accuracy even in the case of small tumours and blurred boundaries. The average Dice coefficient of the first method reaches 0.78, while

the second method is further optimised to 0.80, which is significantly better than the traditional U-Net of 0.70.

In addition, this study developed a novel boundary optimisation method to improve the accuracy of the segmentation boundary by extracting the boundary area of the initial segmentation result and refining it. The optimisation module uses a deep learning-based boundary refinement network (BRN) to fine-tune the boundaries of the liver and tumour. Through block processing, this method effectively reduces false positive and false negative areas and significantly improves the continuity and accuracy of the segmentation results. Multiple case verification results show that the optimised model has improved boundary-related indicators (such as average symmetric distance and boundary overlap error) by about 10%.

## Contents

<b>Acknowledgements .....</b>	<b>iii</b>
<b>Abstract .....</b>	<b>iv</b>
<b>List of acronyms.....</b>	<b>x</b>
<b>List of figures.....</b>	<b>xii</b>
<b>List of tables.....</b>	<b>xvi</b>
<b>Chapter 1. Introduction.....</b>	<b>1</b>
1.1 Preface .....	1
1.2 Motivation and Challenges .....	2
1.2.1 Motivation .....	2
1.2.2 Challenges .....	3
1.3 Research objectives .....	5
1.4 Summary of contribution.....	6
1.5 Author's publications .....	7
1.6 Thesis Outline.....	8
<b>Chapter 2. Liver and Hepatic disease and Computed Tomography .....</b>	<b>11</b>
2.1 Introduction .....	11
2.2 Liver Anatomy and Physiology .....	12
2.3 Liver cancers.....	15
2.3.1 Overview and Epidemiology of Liver Cancer.....	15
2.3.2 Causative factors.....	16
2.3.3 Diagnostic methods for liver diseases .....	18
2.3.4 Treatment Options and Clinical Interventions.....	19
2.4 Computed Tomography (CT) Technology .....	25
2.4.1 Overview of Medical Imaging Modalities for Liver Diagnosis .....	25
2.4.2 Overview of CT .....	27
2.4.3 Imaging principle.....	30
2.4.4 CT artefacts.....	32
2.5 Conclusion .....	43
<b>Chapter 3. Image segmentation and quantification techniques based on CT Data</b>	<b>44</b>
3.1 Introduction .....	44
3.2 Image preprocessing techniques .....	44
3.3 Classical image segmentation techniques.....	47
3.3.1 Interactive Segmentation Method.....	49
3.3.2 Fully Automatic Segmentation .....	51
3.4 Deep Neural Network for image segmentation .....	57
3.4.1 Deep Neural Network.....	58
3.4.2 Deep Convolutional Neural Network .....	63
3.4.3 Fully Convolutional Network.....	68
3.4.4 U-Net .....	73

3.4.5	Application of deep learning in liver and tumour segmentation ....	76
3.5	Conclusion .....	85
<b>Chapter 4.</b>	<b>CT dataset and evaluation parameters .....</b>	<b>87</b>
4.1	The Liver Tumour Segmentation (LiTS) Dataset.....	87
4.2	Evaluation Parameters .....	90
4.3	Conclusion .....	97
<b>Chapter 5.</b>	<b>Liver segmentation method using an improved Multi Scale Feature U-Net (MSF U-Net) .....</b>	<b>98</b>
5.1	Introduction .....	98
5.2	Liver CT scans segmentation using Modified MSF U-Net.....	99
5.2.1	Overview of Modified MSF U-Net .....	99
5.2.2	Dilated Convolution .....	100
5.2.3	Pyramid Pooling Module.....	102
5.3	Visualisation and statistical analysis of liver segmentation results .....	104
5.3.1	Network Implementation.....	104
5.3.2	Statistical Analysis.....	105
5.4	Conclusion .....	111
<b>Chapter 6.</b>	<b>Novel CT liver tumour segmentation method based on Attention Module 113</b>	
6.1	Introduction .....	113
6.2	A novel Cascade U-Net with customised attention mechanism (CAU-Net) for liver tumour segmentation.....	114
6.2.1	Overview of the liver tumour segmentation process .....	114
6.2.2	Details of proposed Customised Attention U-Net (CAU-Net).....	115
6.3	A novel end-to-end Cascade Hybrid Attention U-Net (CHAU-Net) for liver tumour segmentation.....	119
6.3.1	Overview of proposed end-to-end Cascade Hybrid Attention U-Net .....	119
6.3.2	Details of end-to-end cascade attention U-Net.....	119
6.4	Visualisation and Statistical analysis of tumour segmentation results ....	126
6.4.1	CAU-Net and CHAU-Net Implementation for tumour segmentation .....	126
6.4.2	Discussion with segmentation results.....	127
6.5	Conclusion .....	136
<b>Chapter 7.</b>	<b>A Boundary Optimisation Scheme for Liver Tumours from CT Images 138</b>	
7.1	Introduction .....	138
7.2	Boundary Optimisation Process for liver tumour scans .....	139
7.2.1	Overview of Proposed Optimisation Workflow .....	139
7.2.2	Boundary Patch Extraction .....	140
7.2.3	Boundary Refinement Network (BRN).....	142
7.2.4	Reassembling Process.....	144



7.3	Visualisation and Statistical analysis of segmentation results.....	145
7.4	Conclusion.....	150
<b>Chapter 8.</b>	<b>Conclusion .....</b>	<b>151</b>
8.1	Conclusion.....	151
8.2	Future Work.....	153
<b>References</b>	<b>.....</b>	<b>156</b>

## List of acronyms

AFB1	Aflatoxins B1
AFP	Alpha-fetoprotein
ADH	Alcohol Dehydrogenase
ALPPS	Associating liver partition & portal vein ligation for staged hepatectomy
ASSD	Average symmetric surface distance
BCLC	Barcelona Clinic Liver Cancer
BS U-Net	Bottleneck Supervised U-Net
CNN	Convolutional Neural Network
CT	Computed Tomography
DCNN	Deep Convolutional Neural Networks
DICOM	Digital Imaging and Communications in Medicine
DNA	Deoxyribonucleic acid
EASL	European Association for the Study of the Liver
ECOG	Eastern Cooperative Oncology Group
FBP	Filtered Back Projection
FCN	Fully Convolutional Network
FEM	Finite Element Modelling
FLR	Future Liver Remnant
FN	False Negative
FP	False Positive
GAN	Generative Adversarial Network
HAIC	Hepatic Arterial Infusion Chemotherapy
HBV	Hepatitis B virus
HCC	Hepatocellular Carcinoma
HCV	Hepatitis C virus

HU	Hounsfield Unit
IARC	International Agency for Research on Cancer
LiTS	Liver Tumour Segmentation
MRI	Magnetic Resonance Imaging
MSD	Maximum symmetric distance
NAFLD	Non-alcoholic fatty liver disease
NCI	National Cancer Institute
NIFTI	Neuroimaging Informatics Technology Initiative
OARs	Organs at Risks
PCA	Principal Component Analysis
PLB	Percutaneous Liver Biopsy
PVE	Partial Volume Effect
RFA	Radiofrequency Ablation
RT	Radio Therapy
RVD	Relative Volume Distance
SGM	Statistical Shape Model
SGD	Stochastic Gradient Descent
SSC	Sparse Shape Composition
SVM	Support Vector Machine
TACE	Transcatheter Arterial Chemoembolization
TLV	Total Liver Volume
TN	True Negative
TNM	Tumour–node–metastasis
TP	True Positive
TJLB	Trans-jugular Liver Biopsy
VOE	Volume Overlap Error
WHO	World Health Organization

## List of figures

Figure 1.1 Required functional liver ratio (FLR) after liver resection in patients with normal liver, moderate liver disease and cirrhosis.....	3
Figure 1.2 Abdominal CT scan showing: (a) normal liver and stomach, (b) normal liver and heart, (c) liver tumour and surrounding tissues.....	4
Figure 2.1 The anatomical illustration of the liver, both anterior and posterior views[30] . .....	14
Figure 2.2 Couinaud's liver segments [5]. (a) Anterior surface view; (b) visceral surface view. I, caudate/Spigel's lobe; II, left, posterolateral segment; III, left anterolateral segment; IVa, left superomedial segment; IVb, left inferomedial segment; V, right anteroinferior segment; VI, right posteroinferior segment; VII, right posterosuperior segment; and VIII, right anterosuperior segment.....	14
Figure 2.3 Four types of major hepatectomy. White segments are planned for surgical resection. a: Complete right hepatectomy, keep I,II,II and IV part. B: Extended right hepatectomy, keep I,II and III part. C: Complete left hepatectomy, keep I,V,VI,VII and VIII part. D: Extended left hepatectomy, keep VI and VII part[10]. .....	23
Figure 2.4 Components of a modern third generation CT system. The fan beam covering a whole-body scan field of view with a diameter of 50 cm is indicated.....	28
Figure 2.5 Original CT image and CT image after window transformation.....	32
Figure 2.6 Cupping artefacts (A) in the skull model and artefact-corrected image (B) [85].....	34
Figure 2.7 Streaks artefacts located in the posterior fossa region of the brain [86]. ...	35
Figure 2.8 A, Beam hardening causing streak artefact at the level of the thyroid isthmus results in decreased soft-tissue contrast and obscures the vascular (arrow) B, clear vascular without artefact [87]. ....	35
Figure 2.9 Partial volume artefact (circled in yellow) decreased with decreased voxel size [89].....	37
Figure 2.10 CT image of a shoulder phantom shows streaking artefacts by photon starvation [82]. ....	38
Figure 2.11 Motion causes blurring and double images (left), as well as long range streaks (right) [25].....	39
Figure 2.12 Clinical benefits of increasing CT scanner speed by increasing the number of rows. (a) 16-slice coronal image showing inconsistent cardiac motion. These inconsistencies are greatly reduced in (b) the coronal image obtained from the 64-slice scan [91]. ....	39
Figure 2.13 a. Streak artefacts due to motion in abdominal scanning and b. Abdominal scanning results under ideal conditions [82].....	40
Figure 2.14 A. Dark streak between hip replacements is mostly due to beam hardening and scatter. B. Sharp thin alternating streaks surrounding an aneurysm coil are	

mostly due to motion and under sampling. C. Smoothly undulating streaks around cholecystectomy clips are due to windmill artefact [25].	42
Figure 2.15 Series of CT images from a helical scan of the abdomen shows helical artefacts (arrows) [25].	43
Figure 3.1 (a) and (b) Abdominal CT images before and (c) and (d) after window level adjustment, the first column is the original image, and the second column is the adjusted image.	47
Figure 3.2 Segmentation process of data sets in statistical liver models. As the number of shape modes increases, the results of the optimised combination of position and shape gradually tend to the true contour [125].	53
Figure 3.3 History of AI research from 1958 to the present [149, 153, 154].	58
Figure 3.4 A typical neuron structure.	59
Figure 3.5 Sigmoid Function and its gradient expression.	60
Figure 3.6 Tanh Function and its gradient expression.	61
Figure 3.7 ReLU Function and its gradient expression.	61
Figure 3.8 The structure of one-layer neural network and a 6-layer deep neural network.	63
Figure 3.9 The structure of a typical convolution neural network.	65
Figure 3.10 Example of Max Pooling and Average Pooling[164].	68
Figure 3.11 The simple structure of a Fully Convolutional Network [159].	71
Figure 3.12 Up-sampling process in FCN, including 8 times and 16 times up-sampling prediction [159].	72
Figure 3.13 The architecture of a classical U-Net.	74
Figure 3.14 The encoder-decoder architecture of the Transformer model [147]. The core structure of the Transformer model includes the encoder module on the left and the decoder module on the right. Each module consists of multiple layers of stacked substructures: multi-head attention mechanism, feedforward network, residual connection and layer normalization.	83
Figure 4.1 Four CT scans with different HU values, which can be seen from the depth of the liver.	89
Figure 4.2 Confusion matrix in the liver tumour segmentation.	92
Figure 4.3 Dice score in segmentation problem.	94
Figure 4.4 Example of calculating the ASSD and MSD of two graphs.	96
Figure 5.1 The architecture of proposed modified U-Net for automatic liver segmentation.	101
Figure 5.2 Normal Convolution calculation vs Dilated Convolution with a dilated ratio of 2.	101
Figure 5.3 Architecture of Pyramid Pooling Module.	103
Figure 5.4 Statistical analysis of segmentation performance across seven random training-testing splits using six evaluation metrics.	106
Figure 5.5 Visualisation results of liver segmentation from three patients, each row corresponding to a different patient, and columns representing different stages of	

the analysis process.....	110
Figure 5.6 A challenging case where the liver segmentation model produces significant errors. (a), (b), (c), and (d) represent the original image, liver prediction result, liver label, and overlapped image, respectively. The green contour shows the predicted liver region, while the red represents the ground truth. ....	111
Figure 6.1 The flow chart for proposed liver tumour segmentation process. ....	114
Figure 6.2 The architecture of CAU-Net with attention block. ....	116
Figure 6.3 The structure of Customised Attention Block. ....	117
Figure 6.4 The architecture of proposed modified cascade attention U-Net .....	121
Figure 6.5 The structure of channel attention module and spatial attention module respectively. ....	124
Figure 6.6 Visualization of tumour segmentation results using the proposed CAU-Net method across patients with small and medium tumours. The first row shows the original CT slices, followed by the ground truth annotations (in red), model predictions (in green), and the overlay of prediction and ground truth. ....	132
Figure 6.7 Visualization of tumour segmentation results using the proposed CAU-Net method across patients with multiple medium tumours and large tumours. Overlay images confirm high spatial correspondence, with minor discrepancies on smaller lesions. ....	133
Figure 6.8 Visualization of tumour segmentation results using the proposed CHAU-Net method across patients with small and medium tumours. The first row shows the original CT slices, followed by the ground truth annotations (in red), model predictions (in blue), and the overlay of prediction and ground truth. ....	134
Figure 6.9 Visualisation of tumour segmentation results using the proposed CAU-Net method across patients with multiple medium tumours and large tumours. Overlay visualizations show good alignment for multiple and medium-sized tumours, with slight bound.....	135
Figure 6.10 Visualisation results after hard attention stage. The second column is the cropped liver label, and the third column is the corresponding cropped input image .....	136
Figure 7.1 Proposed optimisation workflow, (a) liver CT image and corresponding ground truth, (b) extracted coarse segmentation (c) image and (d) mask boundary patches from the coarse segmentation (e) mask boundary patches from the ground truth (f) all patches pass the boundary refinement network (g) final refined segmentation result. ....	140
Figure 7.2 The architecture of proposed boundary refinement segmentation network. ....	142
Figure 7.3 Patches are captured with a step size of 40 on the tumour boundaries. ...	144
Figure 7.4 Visualisation results of four different patients. Red line presents the ground truth of tumour boundary, yellow line present segmentation results need to be refined, green line presents refined segmentation boundary.....	148
Figure 7.5 Comparison of boundary optimisation segmentation network with and	

without mask patch. Red: Ground Truth; Yellow: Coarse Segmentation; Green: Refined Segmentation. Yellow boxes show the Dice score between the coarse result and ground truth, while green boxes show the Dice score between the refined result and ground truth. .... 149

## List of tables

Table 2.1 ECOG Physical Status Rating Criteria.....	20
Table 2.2 Child-Pugh Score System Criteria. ....	21
Table 2.3 BCLC Staging System Criteria. ....	21
Table 2.4 Overview of Common Medical Imaging Techniques for Liver Disease Diagnosis.....	26
Table 2.5 Introduction to CT scanners from the first to the fifth generation, including scanning methods, scanning times and working principle [79, 80]. ....	29
Table 2.6 HU values of different tissues .....	32
Table 2.7 Artefacts in CT scanning according to different sources .....	33
Table 3.1 Comparison between NIFTI file and DiCOM file commonly used in medical imaging. ....	45
Table 3.2 Summary of medical image segmentation techniques. ....	48
Table 3.3 Summary of Model-based segmentation algorithm. ....	52
Table 3.4 Comparative Analysis of Liver and Tumour Segmentation Methods. ....	85
Table 4.1 Confusion Matrix. ....	90
Table 5.1 Comparison of the proposed model with other nine liver segmentation methods based on six measured metrics. The symbol “N/A” represents unreported result. Bold font represents the highest score on each measurement.....	107
Table 6.1 Comparison of proposed model with other seven tumour segmentation methods based on six measured metrics. The symbol “N/A” represents unreported result. Bold font represents the highest score on each measurement.....	128
Table 7.1 Comparison of proposed model with other seven tumour segmentation methods based on six measured metrics. The symbol “N/A” represents unreported result. Bold font represents the highest score on each measurement.....	146



# Chapter 1. Introduction

## 1.1 Preface

Hepatocellular Carcinoma (HCC) refers to malignant tumours that mainly occur in the liver, including subtypes such as hepatocellular carcinoma and cholangiocarcinoma [1]. According to the World Health Organization (WHO) [2], there will be approximately 1 million [3] new cases of liver cancer and approximately 870,000 cancer deaths worldwide in 2025. Liver cancer is more common in male patients, accounting for approximately 75% of all new cases. In addition, the global incidence of liver cancer has increased over the past few decades, especially in Asia and Africa [4]. In the United States, according to the National Cancer Institute (NCI) [4], new cases of liver cancer accounted for approximately 2.4% of all cancer cases in 2020, and liver cancer deaths accounted for approximately 5.3% of all cancer deaths. In the United Kingdom, the incidence of liver cancer is also rising, with approximately 6,000 new cases and approximately 4,000 deaths each year from 2017 [5].

The main risk factors for HCC include chronic hepatitis B (HBV) and hepatitis C (HCV) infection, cirrhosis, excessive alcohol consumption, non-alcoholic fatty liver disease (NAFLD), and aflatoxin exposure [6]. Treatments for early-stage HCC include surgical resection, radiofrequency ablation, and radiotherapy (RT). For advanced HCC, systemic therapies such as targeted therapy and immunotherapy are also important treatment options [7]. RT can be applied to both primary tumours and metastatic lesions. As an alternative to surgery, RT can achieve similar oncological outcomes while preserving organs at risk (OARs) [8].

Ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI) are commonly used tools for auxiliary diagnosis. At the same time, quantitative analysis and three-dimensional visualization of scan results are also important bases for clinical diagnosis, pathological grading, and treatment decisions for liver cancer.

However, manual segmentation of the liver and tumours from medical images not

only relies on clinical expertise and available facilities but is also time-consuming and susceptible to subjective differences. To address these issues, it is particularly important to develop automatic tumour delineation methods. Electronic computer-based automation technology can reduce subjective differences and provide objective and consistent support to oncologists, thereby reducing their workload and improving the effectiveness of treatment [9].

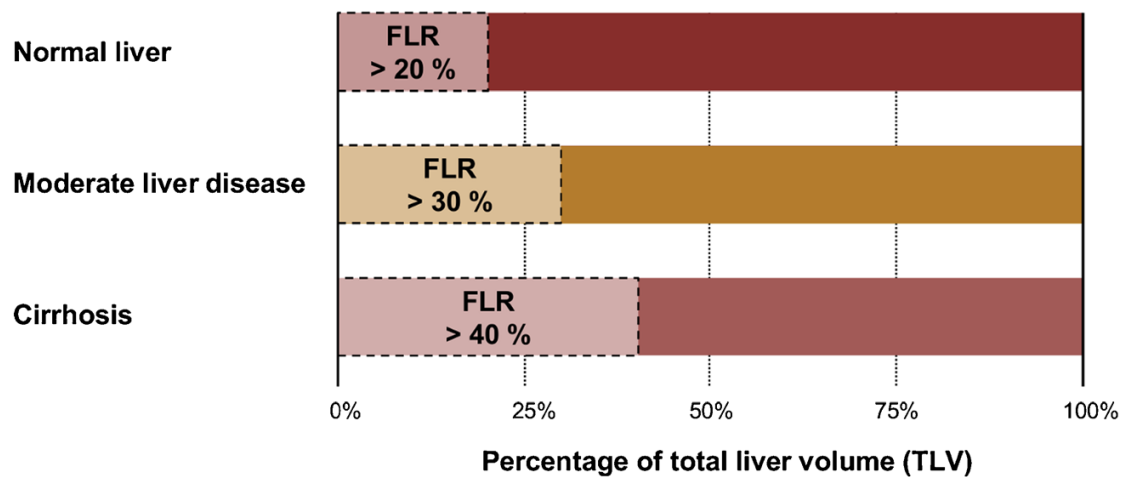
## **1.2 Motivation and Challenges**

### **1.2.1 Motivation**

The precise identification and quantification of liver structures through CT segmentation is crucial for surgical planning and postoperative evaluation. It allows for a detailed understanding of the liver's anatomy and pathological changes, essential for determining the surgical approach and techniques. Accurate segmentation helps in assessing the size and location of tumours, planning resections, and ensuring that enough healthy liver tissue remains post-surgery to maintain liver function. The most intuitive function of liver CT image segmentation is to calculate the total liver volume (TLV) and future liver remnant (FLR). These two indicators are crucial for most liver surgeries, such as major hepatectomy, portal vein embolization, associating liver partition and portal vein ligation for staged hepatectomy (ALPPS) and liver transplant surgery [10]. As illustrated in Figure 1.1, for patients with normal liver function, postoperative FLR should usually be at least 20% of TLV; in patients with moderately diseased liver, the ratio should be higher than 30%; if the patient has cirrhosis or impaired liver function, this ratio may need to be higher, usually 40% or more [11].

Liver CT tumour segmentation plays a key role in surgical planning. It provides surgeons with a basis for decision-making by accurately delineating the tumour's location, size, morphology, and its spatial relationship to surrounding critical structures, such as blood vessels and healthy liver tissue. This allows surgeons to develop optimal surgical paths and strategies to minimise damage to healthy tissue and potential surgical

risks. In addition, precise segmentation of the tumour allows doctors to evaluate the complete resection of the tumour and whether the remainder of the liver can maintain its normal function after surgery. This is critical in the selection of treatment options, monitoring for potential recurrence and whether further treatment is needed.



**Figure 1.1 Required functional liver ratio (FLR) after liver resection in patients with normal liver, moderate liver disease and cirrhosis.**

### 1.2.2 Challenges

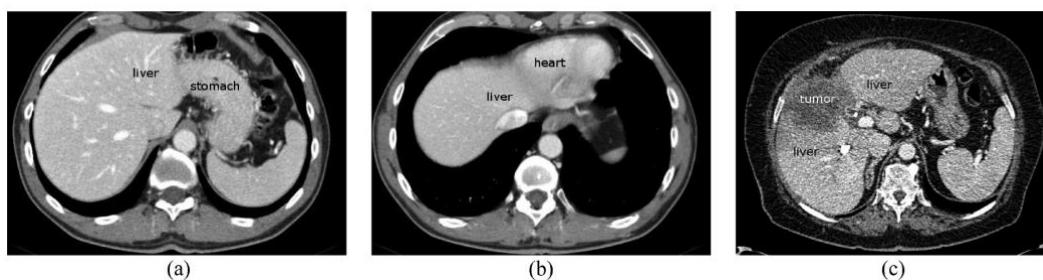
CT segmentation of liver and liver tumours is a complex task, which faces various challenges and difficulties in the field of medical image processing. Here are some of the main challenges:

- (1) Whether it is liver segmentation or tumour segmentation, motion artefacts, beam hardening artefacts and noise in CT images may affect the image quality, thus interfering with the accurate segmentation of targets [12].
- (2) The contrast between the liver and other surrounding tissues (e.g., kidneys, stomach) may be low, making it more difficult to accurately distinguish boundaries. Similarly, the contrast between a tumour (in its early stages) and surrounding normal liver tissue may not be apparent, especially if the tumour has blurred edges or is similar to the surrounding tissue. In addition, changes in the contrast between the target and the background are affected by many factors, such as different contrast agents, different injection times and acquisition times [13].

(3) The shape, size, and structure of the liver can vary greatly from patient to patient. This diversity makes it difficult to design a universal segmentation algorithm for all types of liver images. This problem is even more challenging for tumours, where the diversity in size, shape, borders, and internal structure of liver tumours (including benign and malignant tumours) complicates segmentation. In the case of chronic liver disease, the background information of the liver tissue may change significantly making tumour segmentation more difficult [14, 15].

(4) Small tumor segmentation in CT images is particularly challenging. Due to their limited volume, small tumours often occupy only a few pixels or slices, making them difficult to distinguish from surrounding liver tissue. Their intensity distribution may be very similar to that of healthy tissue, resulting in low contrast and blurred boundaries. In addition, the presence of imaging noise may completely cover these small lesions. These factors make small tumor detection extremely susceptible to false negatives, thus requiring segmentation algorithms to be both sensitive and spatially accurate.

(5) Medical images are usually three-dimensional, high-resolution images. Segmentation algorithms need to process a large amount of volume data, which inevitably increases computational complexity. This requires the algorithm to be efficient while maintaining accuracy.



**Figure 1.2 Abdominal CT scan showing: (a) normal liver and stomach, (b) normal liver and heart, (c) liver tumour and surrounding tissues.**

Figure 1.2 show some examples to illustrate why liver segmentation is a challenging task. In the first two images, the liver tissue has to be separated from adjacent organs stomach (Figure 1.2a) and heart (Figure 1.2b). The gray-values in all structures are highly similar, which makes boundary detection difficult without a-priori

information about the expected shape in these regions. In the third image (Figure 1.2c), the tumour should be segmented as part of the liver. However, there is a considerable intensity difference between both structures, which often leads to misclassification of the tumour as non-liver tissue [16].

Furthermore, the morphology of lesions is difficult to predict due to the irregular growth of tumours and their changes after medical intervention such as surgical resection, undermining the effectiveness of segmentation strategies that rely on predefined lesion shapes. This is especially true for shape-based segmentation technology, which is already well established in the field of organ segmentation [17].

### 1.3 Research objectives

The aim of this thesis is to design an automatic liver and liver tumour segmentation framework based on image processing techniques and deep neural networks, which can obtain similar results when compared to the professional oncologist, but with less time. The framework is structured around three main goals:

1. **Liver Segmentation with an Improved U-Net:** The first goal is to achieve precise segmentation of the liver from abdominal CT scans using an improved version of the U-Net architecture. The aim is to enhance the segmentation accuracy to above 90%, approaching the current state-of-the-art methods, which typically achieve accuracy in the range of 90% to 95%, thereby ensuring that the delineated liver regions are highly accurate and reliable.
2. **Tumour Segmentation:** The second goal is to accurately segment the tumour regions within the liver. This involves developing a robust algorithm capable of identifying and marking all cancerous areas with high sensitivity. The focus will be on maximizing the detection of tumour regions to ensure comprehensive coverage of all malignant tissues.
3. **Optimisation Algorithm for Refinement:** The third goal is to develop an optimisation algorithm that can further refine the segmentation results of both the liver and the tumours. This algorithm will enhance the initial segmentation

outputs, improving the overall performance and accuracy. The refined results are expected to exhibit superior quality, facilitating better diagnosis and treatment planning.

## **1.4 Summary of contribution**

The main contributions of this thesis are described below:

1. The first major contribution is an automatic liver segmentation method based on a multi-scale feature fusion U-Net (MSFU-Net) architecture. By using a two-stage U-Net network to first performs coarse segmentation of the liver and then utilising the MSFU-Net for fine segmentation. It is shown that this method significantly improves the accuracy of segmentation compared to the other leading approaches. In order to enhance network performance, residual modules and dense connections are introduced, and multi-dimensional information fusion technology is adopted. The network performs an average Dice score of 0.961 on the Liver Tumour Segmentation (LiTS) dataset [17], outperforming most existing methods. In addition, the class imbalance problem is effectively solved through image cropping and data augmentation. At the same time, the method has a simple structure and reduces computational complexity and overfitting risk.
2. The second contribution is the design of a concise improved U-Net network to achieve automatic segmentation of liver tumours. By introducing a custom attention module, the sensitivity of the deep neural network to extracting features is increased to achieve accurate tumour segmentation. In addition, this method is also applicable to liver segmentation.
3. The third contribution is an automatic segmentation method of liver and liver tumours based on an improved cascaded U-Net network. This method combines a hybrid attention mechanism and effectively enhances feature extraction and tumour detection capabilities by embedding hard attention modules and soft attention mechanisms. Evaluation on the LiTS dataset shows

that the new method achieves an average Dice score of 0.762, which is significantly improved compared to other methods. This method not only reduces computational complexity in automatic cropping of the liver region, but also improves the class imbalance problem, especially in small tumour detection.

4. The fourth contribution in the research is a new liver tumour segmentation optimisation algorithm to improve the segmentation accuracy of tumour boundaries in CT images. This method significantly improves the segmentation accuracy by extracting boundary patches from the coarse segmentation results and optimizing them, with the average Dice coefficient reaching 0.805. A multi-layer information fusion network combines the original image and coarse segmentation results to enhance the accuracy of semantic and location information. Experimental results show that this method is superior to existing methods in multiple evaluation indicators, demonstrating the ability to effectively optimise tumour boundaries and providing an important reference for future research and clinical applications.

## **1.5 Author's publications**

The following have been published and presented at technical conferences:

1. M. Gong, J. Soraghan, G. Di-Caterina & D. Grose, "A U-Net based multi-scale feature extraction for liver tumour segmentation in CT images" in 2021 Communications, Signal Processing, and Systems (CSPS), China.
2. M. Gong, J. Soraghan, G. Di-Caterina, B. Zhao & D. Grose, "Hybrid attention mechanism for liver tumour segmentation in CT images" in 2022 European Workshop on Visual Information Processing (EUVIP), Portugal.
3. M. Gong, J. Soraghan, G. Di-Caterina, X. Li & D. Grose, "A boundary optimisation scheme for liver tumours from CT images" in 2023 European Signal Processing Conference (EUSIPCO), Finland.

## 1.6 Thesis Outline

This thesis is structured as follows:

Chapter 2 provides a comprehensive overview of liver physiology, anatomy, and various hepatic diseases, with a particular focus on liver cancer. It begins with an introduction to the liver's anatomical structure and functions, detailing how the liver plays a crucial role in metabolic regulation, detoxification, and protein synthesis. The chapter then explores the different types of liver cancer, such as hepatocellular carcinoma, and discusses the main risk factors, including chronic hepatitis B and C infections, alcohol consumption, and non-alcoholic fatty liver disease. The chapter also covers the various diagnostic methods used to detect liver cancer, including blood tests, ultrasound, computed tomography, and magnetic resonance imaging. Each method's advantages and limitations are discussed to highlight their relevance in clinical practice. The chapter further delves into treatment options for liver cancer, ranging from surgical resection and radiofrequency ablation to chemotherapy and targeted therapies. The section on computed tomography includes a historical overview of its development, the principles of CT imaging, and a discussion of common CT artefacts that can affect image quality.

Chapter 3 focuses on the various image segmentation and quantification techniques used in processing CT data. It begins with an introduction to image pre-processing techniques that prepare CT images for segmentation by enhancing image quality and removing noise. The chapter then examines classical image segmentation methods, including interactive and fully automatic segmentation techniques. Interactive methods, such as region growing and watershed algorithms, require user input to guide the segmentation process, while fully automatic methods use predefined criteria to segment images without human intervention. The chapter also provides an in-depth analysis of deep neural networks, particularly their application in image segmentation. It explores the architecture and functionality of deep convolutional neural networks (CNNs) and fully convolutional networks (FCNs). A significant portion of the chapter is dedicated



to the U-Net, a popular deep learning model for biomedical image segmentation.

Chapter 4 describes the CT datasets used in this thesis, specifically the Liver Tumour Segmentation dataset, which includes CT scans from multiple patients annotated by professional radiologists. The chapter details the dataset composition, including the number of images, the variety of tumour types, and the labelling process. It also discusses the challenges associated with the dataset, such as class imbalance, where certain types of tumours are underrepresented, making the segmentation task more difficult. The chapter then outlines the evaluation parameters used to assess the performance of the segmentation algorithms developed in this thesis. These metrics provide a comprehensive evaluation framework to compare the proposed methods against existing techniques and ensure robust performance assessment.

Chapter 5 introduces a novel liver segmentation method based on a modified U-Net architecture designed to improve the accuracy of liver segmentation from CT scans. The chapter begins with an overview of the modified U-Net, highlighting its key components, such as dilated convolutions and pyramid pooling modules, which enhance the network's ability to capture multi-scale features and improve segmentation accuracy. The implementation details of the network are provided as well, including the training process, hyperparameter selection, and optimisation techniques. Experimental results demonstrate that the proposed method achieves superior performance compared to traditional U-Net and other existing segmentation methods. Visualization of the segmentation results is provided to illustrate the method's effectiveness in accurately delineating liver boundaries in CT images.

Chapter 6 introduces two advanced liver tumour segmentation methods based on the cascaded U-Net architecture with integrated attention modules. This chapter first introduces the cascaded U-Net and explains how the network processes CT images in multiple stages to gradually refine the segmentation results. The designed attention module aims to enhance the network's ability to focus on relevant features and improve segmentation accuracy by selectively emphasizing important regions in the input image.

Experimental results are provided, showing significant improvement in segmentation performance compared to the traditional U-Net model.

Chapter 7 introduces a novel framework to enhance the segmentation accuracy of liver tumours in CT images. This chapter first introduces the various stages of optimisation scheme and explains how these stages (including boundary patches extraction, boundary optimisation network and restoration) process CT images in multiple stages to gradually refine the segmentation results. The proposed boundary optimisation network will be described in detail, and how to use a multi-scale information fusion strategy to enhance the details and accuracy of the segmentation will be discussed. The results of liver segmentation and tumour segmentation of this scheme will be listed, analysed and discussed separately.

The final chapter summarises the main contributions and findings of the thesis, reflects on the research objectives and discusses the limitations encountered during the research process. It provides a comprehensive overview of how the proposed methods advance the field of liver tumour segmentation on CT images and highlights their potential impact on clinical practice.

# **Chapter 2. Liver and Hepatic disease and Computed Tomography**

## **2.1 Introduction**

In modern medicine, understanding the functions of human organs and related diseases is the cornerstone for improving diagnostic accuracy and treatment effectiveness. The liver, as one of the largest internal organs in the human body, plays multiple vital physiological roles. The liver is mainly responsible for metabolic regulation, detoxification, protein synthesis and bile secretion [18], and is crucial for maintaining metabolic balance and internal environment stability throughout the body. However, due to its central position and multifunctionality, the liver is also highly susceptible to various pathological conditions, liver cancer being one of the most lethal.

Liver cancer, especially hepatocellular carcinoma, is the fourth leading cause of death and the sixth most common cancer worldwide. Its development is often associated with chronic liver diseases such as hepatitis B and C virus infection, alcoholic liver disease, non-alcoholic fatty liver disease, and long-term exposure to liver carcinogens. Liver cancer develops through multiple stages, and early diagnosis is crucial to improving patient prognosis [19, 20].

The process of diagnosing and treating liver cancer involves multiple stages. Commonly used clinical liver cancer staging systems, such as the Barcelona Clinic Liver Cancer (BCLC) [21] staging system, can guide treatment selection and predict patient prognosis. There are various methods to treat liver cancer, including surgical resection, local ablation, chemotherapy, radiotherapy, and targeted therapy and immunotherapy [8, 22-24]. Which treatment is chosen depends on the stage of the cancer, the patient's overall health, and treatment availability.

Medical imaging plays a central role in diagnosis and treatment planning. CT is a widely used imaging technology that synthesises X-ray measurement data to produce detailed cross-sectional images of the inside of the body. The high resolution and

scanning speed of modern CT scans make them ideal for identifying, localizing and evaluating liver tumours. However, although CT technology is extremely useful in diagnosing liver cancer, its image quality can be affected by various artefacts, such as motion artefacts, beam hardening, scattering, and noise, which can interfere with image interpretation [25].

Section 2.2 of this chapter will review the anatomy and functions of the liver. An accurate understanding of the liver's anatomy is critical to planning the surgical path, minimising damage to healthy tissue, and maintaining blood supply to the surgical area. Section 2.3 introduces the distribution and main causes of liver cancer around the world and introduces corresponding treatment methods. Section 2.4 then introduces the widely used medical imaging technology called CT, including its development, and working principles. In addition, various artefacts that may appear in CT imaging data are discussed in this section. Section 2.5 gives the motivation and challenges of this research.

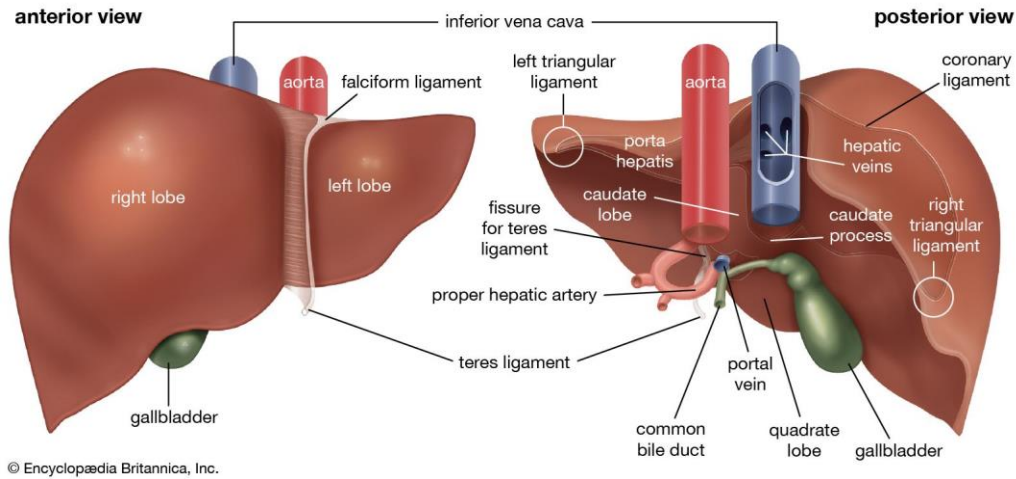
## **2.2 Liver Anatomy and Physiology**

In humans, liver is the largest internal organ in the body, it is usually in the right upper quadrant of the abdomen and partially extended to upper left, below the diaphragm and above the stomach. From the perspective of body surface projection, liver is generally located between the 5th and 11th ribs. The shape of the liver is an irregular wedge, blunt and thick on the right side and narrow on the left side. Generally, the left and right diameter (length) is about 25 cm, the upper and lower diameter (width) is about 15 cm, and the front and rear diameter (thickness) is about 6 cm. The liver is a reddish-brown colour and has a smooth, shiny surface due to the high amount of blood and haemoglobin [26-28].

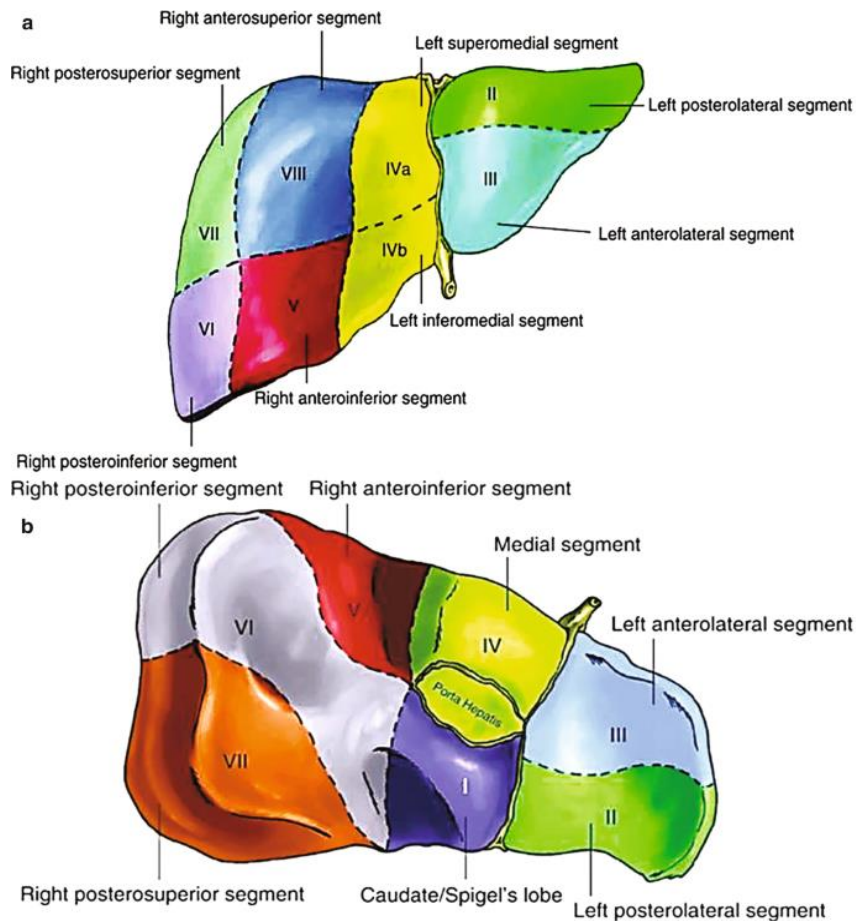
Figure 2.1 is an anatomical illustration of the liver, showing both anterior and posterior views. The liver is divided into two primary lobes, the right and left lobes. The right lobe is larger, and the left lobe is situated to the left of the gallbladder and the falciform ligament, which separates the two lobes. Additionally, the liver features the

caudate and quadrate lobes; the caudate lobe is positioned near the inferior vena cava on the posterior aspect, while the quadrate lobe is adjacent to the gallbladder. The liver is a highly circulatory organ with a complex pipeline structure, including hepatic arteries, portal veins, hepatic sinusoids, and common bile ducts. The aorta, a major artery of the body, is shown behind the liver. The blood flow in the liver is provided by the portal vein and hepatic artery, the nutrient-rich portal blood from the visceral circulation is perfused through the portal vein, and the hepatic arterial blood with high oxygen content perfused by the hepatic artery, which forms a perfusion circuit of visceral-sine-systemic circulation. Blood undergoes metabolic processing in the hepatic sinusoids and hepatocytes before being transported out of the liver via the hepatic sinusoids and subsequently entering the inferior vena cava. The inferior vena cava runs along the back of the liver and carries deoxygenated blood from the lower body to the heart. The common bile duct is the conduit for bile produced in the liver to the gallbladder and eventually to the small intestine for digestion. The gallbladder, shown in green, is situated beneath the liver, storing and concentrating bile. The liver is also connected to other structures by various ligaments: the falciform ligament, the left and right triangular ligaments, and the coronary ligament, which help to stabilize the liver's position within the abdominal cavity.

In terms of liver anatomy, Couinaud's liver division divides the liver into eight independent functional segments based on the blood supply and bile duct structure of the liver. Each segment has its own hepatic artery branches, portal vein branches, and bile ducts [29]. Figure 2.2 illustrates an intuitive partition pattern. This method of partitioning is important for precise liver surgery and disease diagnosis, because it allows surgeons and radiologists to understand the liver's anatomy in detail, allowing for more refined resections or treatments.



**Figure 2.1** The anatomical illustration of the liver, both anterior and posterior views[30] .



**Figure 2.2** Couinaud's liver segments [5]. (a) Anterior surface view; (b) visceral surface view. I, caudate/Spigel's lobe; II, left, posterolateral segment; III, left anterolateral segment; IVa, left superomedial segment; IVb, left inferomedial segment; V, right anteroinferior segment; VI, right posteroinferior segment; VII, right posterosuperior segment; and VIII, right anterosuperior segment.

As the largest internal organ in the human body, the liver performs a variety of important functions, including filtering blood, producing bile to aid in digestion, storing nutrients, and detoxifying harmful substances [18, 31-34]. It regulates carbohydrate metabolism by converting monosaccharides absorbed from the intestine into glycogen, which helps stabilize blood glucose levels. In protein metabolism, the liver synthesises proteins and converts ammonia, a byproduct of amino acid breakdown, into urea for excretion. Lipid metabolism in the liver involves the breakdown of fatty acids for energy production via the citric acid cycle and the synthesis of fats, cholesterol, and phospholipids. These lipids are then transported to tissues for storage and utilization. Additionally, the liver detoxifies harmful substances, such as drugs and environmental toxins, transforming them into water-soluble metabolites for excretion. The liver also stores essential vitamins (A, B12, D), minerals (iron, copper), and glycogen, and it plays a key role in hormone regulation by metabolizing excess hormones. Overall, the liver is a crucial organ that performs a wide range of essential functions necessary for maintaining good health.

## **2.3 Liver cancers**

### **2.3.1 Overview and Epidemiology of Liver Cancer**

Liver cancer refers to malignant tumours that affect the liver. There are two main types of liver cancer: primary liver cancer, which starts in the liver, and secondary liver cancer, which starts in another part of the body and spreads to the liver [5]. Primary liver cancer can be further classified into several types, with the most common being hepatocellular carcinoma, which accounts for about 75% of cases. Other types of primary liver cancer include intrahepatic cholangiocarcinoma, which starts in the bile ducts within the liver, and hepatoblastoma, which is a rare type of liver cancer that usually affects children. Secondary liver cancer is cancer that starts elsewhere in the body such as the pancreas, colon, stomach, breast, or lungs and spreads (metastasizes) to the liver. The liver is a common site for multiple tumour metastases, and the cancer

most likely to spread to the liver is bowel cancer. This is because the blood supply to the gut is connected to the liver via the portal vein. Other cancers including skin melanoma, lung cancer, breast cancer and neuroendocrine tumours have a chance of metastasizing to the liver [35-37].

The liver cancer is the sixth commonly cancer the third most deadly malignant tumour in the world. According to statistics, around 564,000 new cases of liver cancer occurred worldwide in 2000 [38], and around 548,000 patients have already died from the disease. And these data have risen to 905,700 and 830,200 respectively by 2020 [1]. In most regions, the incidence and mortality rates for males are 2 to 3 times that of females. Liver cancer ranks fifth (6.3%) and second (10.4%) in global incidence and death rates among men, respectively; while in women, it is only <3% and 5.7% [39]. Transitioning countries have a higher incidence rate compared to transitioned countries, and this disease is the most common cancer in 11 geographically diverse countries in East Asia, South-Eastern Asia, and Northern and Western Africa [39]. The age distribution of liver cancer incidence is related to region. In areas with high liver cancer risk, such as Southeast Asia or West Africa, the incidence of liver cancer increases after the age of 20 and peaks or stabilizes over the age of 50. In contrast, in more developed high-risk countries, age-specific incidence rates increased steadily with age, similar to the pattern in low-risk areas, with most cases occurring in people aged 55 to 74 years [40].

### **2.3.2 Causative factors**

Liver cancer is primarily associated with liver cirrhosis, with 70-90% of cases occurring in patients with cirrhosis [20, 41]. The leading causes of liver cancer include chronic Hepatitis B and C infections, responsible for 56% and 20% of global cases, respectively, especially in regions like sub-Saharan Africa and parts of Asia [19]. These viruses act synergistically in promoting liver cancer development.

Unhealthy lifestyle habits can also increase the risk of liver cancer, especially smoking and heavy drinking. Globally, approximately 18% of the liver cancer burden



may be related to smoking, and an estimated 17% may be related to alcohol consumption [42]. Tobacco contains various carcinogens (such as nicotine, tar, benzopyrene, etc.) that can impair liver function, inhibit the regeneration and repair of liver cells, and these products are mainly metabolized and detoxified by the liver [43]. Excessive alcohol intake increases the risk of developing alcoholic hepatitis, cirrhosis, and fatty liver [20, 41, 44], which in turn increases the risk of developing liver cancer. According to research by Hiroshi et al. [45], excessive drinking of alcohol will increase the risk of liver cancer by 3-10 times. Alcohol is metabolized by alcohol dehydrogenase (ADH) in the liver into acetaldehyde, a toxic compound that is highly toxic to cells and carcinogenic. Furthermore, acetaldehyde can combine with deoxyribonucleic acid (DNA) to form DNA adducts, which interferes with the normal replication and repair process of DNA, leading to gene mutations and cell dysfunction, thereby promoting the development of cancer.

Non-alcoholic fatty liver disease and diabetes are also important causes of liver cancer [46]. According to statistics, 10%-20% of HCC cases are related to NALFD globally [47], and the proportion of the elderly is more common than that of young patients. Liver cancer caused by NALFD accounts for a relatively high proportion in economically developed areas. The United States, Europe, and China are expected to witness rapid increases in the incidence and prevalence of NAFLD-related HCC by 2030 [48].

In addition, some toxins such as Aflatoxins B1 (AFB1) can also increase the risk of cancer if ingested into the body due to improper dietary habits. AFB1 is the most common in humans and animals and is classified as a Group 1 carcinogen by the International Agency for Research on Cancer (IARC). When AFB1 is metabolized by the liver in the body, it is converted into highly active metabolites that can cause DNA breakage, gene mutations, and other genetic material damage, thereby interfering with normal cell function and promoting the formation and development of cancer cells [49, 50]. Aflatoxins are particularly common in tropical and subtropical regions, including

parts of Africa, Southeast Asia, South America, and Central America. The incidence of liver cancer in these areas with high AFB1 exposure is 3 times higher than in normal areas, and in synergy with hepatitis B virus, the risk of liver cancer is 30 times higher than with single exposure to AFB1 [51].

Liver cancer often has no obvious symptoms in the early stage, so it is difficult to find early liver cancer. As the disease progresses, the following symptoms will appear, such as fatigue, loss of appetite, weight loss and other systemic symptoms; upper abdominal discomfort or pain; jaundice, ascites, abnormal liver function; and liver cirrhosis-related complications, such as splenomegaly, hypersplenism, esophageal varices, etc.

### **2.3.3 Diagnostic methods for liver diseases**

There are various methods for screening and examination of HCC, including blood tests, medical imaging tests and pathological tissue examinations [52]. Among these methods, performing pathological tissue examination on liver tissue samples is the most accurate way to determine if a person has cancer.

Blood tests are one of the simplest and most convenient methods for screening liver cancer. Alpha-fetoprotein (AFP) in serum can serve as a tumour marker for liver cancer. The normal reference value for serum AFP concentration ranges from 0 $\mu$ g/L to 40 $\mu$ g/L [53]. When a patient's blood AFP level is higher than the normal value, the possibility of having cancer should be noted. However, this cannot be used as the sole criterion for diagnosing cancer [54]. A study in [55, 56] highlighted that AFP's sensitivity in diagnosing HCC is found to be around 54%, indicating that AFP levels may not be elevated in all HCC cases.

Medical imaging examinations are also widely used in the detection of liver cancer, including ultrasound examination, computed tomography (CT) scans and magnetic resonance imaging (MRI) scans. Each of these methods has its advantages and disadvantages. Ultrasound is non-invasive, inexpensive, and has no radiation exposure, but its specificity in diagnosing benign and malignant lesions is not as good as enhanced

CT scans or MRI. CT scans can clearly display the size, number, morphology, location, boundaries, and intrahepatic ductal system of liver cancer, but they come with some radiation exposure. MRI can provide more information from cross-sectional, coronal, and sagittal images, and its sensitivity in detecting small HCCs and its ability to distinguish lesions are better than CT. However, MRI examinations take longer and have limited ability to display intrahepatic calcification lesions, such as intrahepatic bile duct stones and post-interventional iodized oil deposition [57]. In general, CT or enhanced CT scans are the most common medical imaging techniques used in clinical diagnosis.

Liver biopsy is usually the last resort used to diagnose liver cancer because surgery is required, such as percutaneous liver biopsy (PLB) and trans-jugular liver biopsy (TJLB). Liver biopsy is a procedure that can directly understand the pathological changes of liver tissue, and it is the gold standard for the diagnosis of liver diseases [58, 59].

### **2.3.4 Treatment Options and Clinical Interventions**

#### **2.3.4.1 BCLC staging system.**

Under normal circumstances, when a patient is diagnosed with cancer, the cancer is staged to guide subsequent treatment and prognosis. The staging of cancer is determined based on the development and spread of the tumour. Internationally recognised methods for staging liver cancer include the Barcelona Clinic Liver Cancer (BCLC) staging system and the tumour–node–metastasis (TNM) staging system, etc. Next, the BCLC staging system [60] introduced in 1999 by European Association for the Study of the Liver (EASL) will be described in detail, which is widely used in Europe and North America. The BCLC system takes into consideration the patient's performance status (via Eastern Cooperative Oncology Group Physical Status Rating system), liver cirrhosis situation (via Child-Pugh Score System), and condition of liver tumour, providing corresponding treatment recommendations for each stage of liver

cancer [61].

The patient's performance status system is introduced by the Eastern Cooperative Oncology Group (ECOG), which is used to quantify a patient's general health and activities of daily living. It is an important tool in assessing the progression of a patient's disease, the impact of the disease on the patient's ability to perform daily living and determining appropriate treatment and prognosis. Table 2.1 gives the corresponding scoring standards, with scores ranging from 0 to 5, with 0 indicating complete normality and no activity restrictions, and 5 indicating death. Scores of 1 to 4 correspond to mild to severe activity limitations respectively.

**Table 2.1 ECOG Physical Status Rating Criteria.**

ECOG Physical Status Rating Criteria	
PS rate	Physical Condition Description
0	Fully active, able to carry on all pre-disease performance without restriction.
1	Restricted in physically strenuous but ambulatory and able to carry out of work of a light or sedentary nature.
2	Ambulatory and capable of all selfcare but unable to carry out any work and activities. Up and about more than 50% of waking hours.
3	Capable of only limited selfcare, confined to bed or chair more than 50% of waking hours.
4	Completely disabled. Cannot carry on selfcare. Totally confined to bed or chair.
5	Dead

The Child-Pugh classification system is a scoring system used by medical professionals to assess the severity and prognosis of chronic liver disease, particularly cirrhosis. The system measures five clinical indicators including bilirubin, serum albumin, prothrombin time, ascites condition and hepatic encephalopathy condition. Table 2.2 indicates the corresponding scoring standards, each indicator is scored as 1-3 according to different situations, where 1 represents the best situation and 3 represents the worst situation. The sum of the scores for the patient's five clinical indicators results in the Child-Pugh score. A score of 5-6 is classified as Class A, representing normal liver function; a score of 7-9 is classified as Class B, representing mild to moderate damage; a score of 10-15 implies severe liver damage, at which point the patient is not

suitable to undergo aggressive treatment.

**Table 2.2 Child-Pugh Score System Criteria.**

Child-Pugh Scoring System			
Biochemical Indicators	Score		
	1	2	3
Total bilirubin (mg/dl)	<2	2-3	>3
Serum albumin (mg/ml)	>3.5	2.8-3.5	<2.8
Prothrombin time(INR)	<1.7	1.7-2.3	>2.3
Ascites	None	mild	Moderate or severe
Hepatic encephalopathy	None	I-II	III-IV
Class A = 5–6 point, Class B = 7–9 point, Class C = 10-15 point			

The final step of the BCLC staging involves the evaluation of the tumour, including the size of the tumour, the number of tumours, the presence of vascular invasion, and the status of regional lymph node metastasis and extrahepatic spread. Combining the three scoring systems is the complete Barcelona staging system. Table 2.3 indicates the BCLC scoring standards.

**Table 2.3 BCLC Staging System Criteria.**

BCLC staging system					
Stage	PS system	Child-Pugh level	Tumour situation		
			Tumour number	Tumour size	Vascular invasion lymph node metastasis
Very Early (0)	0	A	Single	<2cm	None
Early (A)	0	A-B	Single	>2cm	None
	0	A-B	Less than 3	<3cm	None
Intermediate(B)	0	A-B	Multiple	Arbitrarily	None
Advanced(C)	1-2	A-B	Arbitrarily	Arbitrarily	Yes
Terminal(D)	3-4	C	Arbitrarily	Arbitrarily	Yes

BCLC provides a prognostic prediction and treatment recommendation strategy based on different cancer stages [21]. When the tumour occurs in BCLC-0 (very early

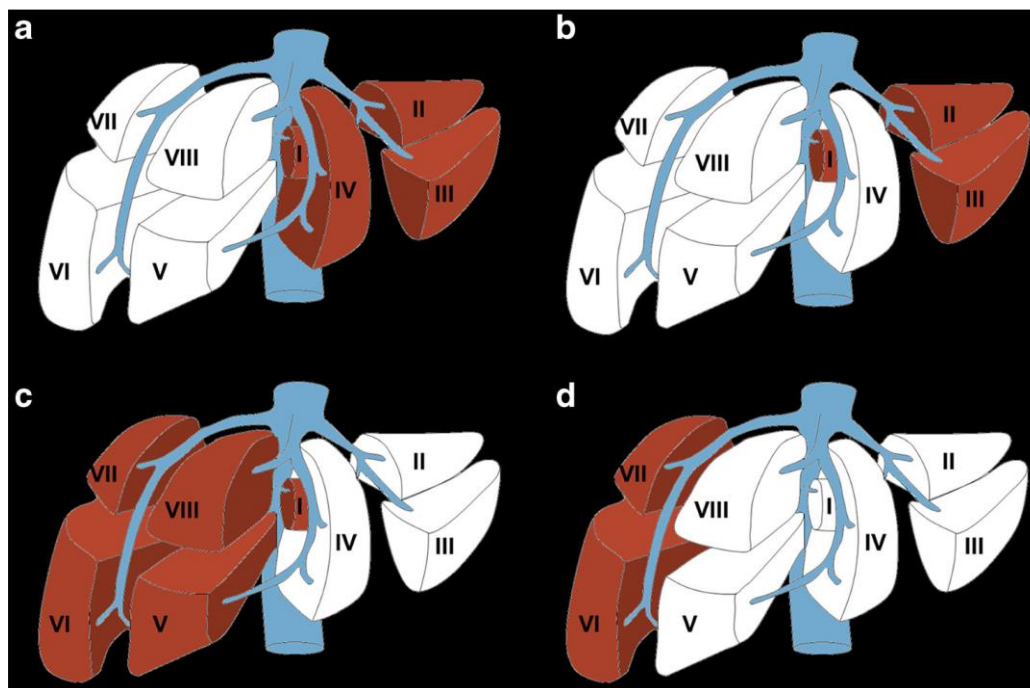
stage) or BCLC-A (early stage), the patient has a single or a small number (less than 3) of tumours, without vascular invasion and cancer-related symptoms. The recommended treatment currently is local ablation or surgery like resection and transplant. In BCLC-B (intermediate stage), multiple tumours appear but there is no extrahepatic metastasis or vascular invasion, and liver function may be damaged. For patients at this stage, transcatheter arterial chemoembolization (TACE) is the treatment of choice [62]. In BCLC-C (advanced stage), there are multifocal tumours with vascular invasion or extrahepatic metastasis, or poor liver function, but the patient has a good general condition. Systemic therapy, such as targeted drugs such as sorafenib, should be a treatment option. In BCLC-D (end stage), the patient's liver function is severely impaired, and there are obvious symptoms and poor quality of life. At this stage, treatment focuses on relieving symptoms and improving quality of life, such as using best supportive care.

#### **2.3.4.2 Surgery Treatment**

Liver resection is mainly used to treat liver tumours, including benign tumours such as hepatocellular adenomas, hepatic hemangiomas, and focal nodular hyperplasia, as well as malignant tumours such as liver metastases (commonly such as metastases from colorectal cancer), primary hepatocellular carcinoma and cholangiocarcinoma [63]. Hepatectomy may also be used to treat intrahepatic gallstones or parasitic cysts in the liver. In living donor liver transplantation, partial hepatectomy is also used to remove part of the liver from a living donor for transplantation. Figure 2.3 shows four types of major hepatectomy. Anatomic resection is generally preferred due to the lower risk of bleeding and biliary fistulas.

During liver resection surgery, major risks include bleeding, biliary fistulas (leakage of bile), and liver failure, particularly in patients with cirrhosis. The risk of these complications varies on an individual basis, and preoperative evaluation and postoperative management are critical to reducing risk [64]. In recent years, advances

in liver surgery have reduced blood loss, infectious morbidity, and operative mortality after extensive hepatectomy. Although advances in liver resection techniques have improved treatment outcomes, liver transplantation remains the preferred option for patients with underlying chronic liver disease [65].



**Figure 2.3 Four types of major hepatectomy. White segments are planned for surgical resection. a: Complete right hepatectomy, keep I,II,II and IV part. B: Extended right hepatectomy, keep I,II and III part. C: Complete left hepatectomy, keep I,V,VI,VII and VIII part. D: Extended left hepatectomy, keep VI and VII part[10].**

### 2.3.4.3 Radiofrequency ablation

Radiofrequency ablation (RFA) technology began to be widely used in the field of cardiology in the 1990s to treat cardiac arrhythmias [66]. Subsequently, this technology was extended to tumour treatment, especially in the treatment of liver, kidney and bone tumours. RFA is a kind of minimally invasive treatment that can be used for early liver cancer. It is effective for tumours that are smaller in size, typically less than 3cm [67] in diameter. Prior to the procedure, imaging scans such as ultrasound, CT scan, or MRI are used to map out the exact location of the tumours in the liver, which helps the doctors to plan the procedure accurately. The fundamental principle of RFA is the use of high frequency alternating current to generate localised high temperatures within

tumour tissue, inducing protein denaturation and coagulative necrosis of the cells. Simultaneously, the thermal effect coagulates surrounding blood vessels, blocking the tumour's blood supply and ultimately eliminating the tumour.

Ablation therapy is already a mature clinical technology with the advantages of minimal invasiveness, safety, high operability, good repeatability, and fast postoperative recovery. However, RFA also has some limitations and risks [68, 69]. For example, it may not be very effective for larger tumours or tumours close to large blood vessels because the heat dissipates too quickly. There is also a risk of extrahepatic complications in nearby structures, such as the bile ducts or intestines, and potential complications such as bleeding, infection, and liver damage could occur. Another major disadvantage of RFA is the high rate of disease recurrence after treatment. Tumours located deep in the liver, close to important blood vessels, or larger than 3 cm may be more difficult to completely eliminate by RFA [70], which increases the risk of recurrence.

#### **2.3.4.4 Chemotherapy**

For patients with intermediate or advanced liver cancer, a single surgery treatment may not be able to prevent the metastasis and spread of cancer cells. In this case, patients are usually treated with chemotherapy. This method uses drugs to kill or inhibit the growth and replication of cancer cells [71]. Common drugs used for treating advanced HCC cancer like Sorafenib, Regorafenib, and Lenvatinib [21], can inhibit the growth of tumour cells and the formation of new blood vessels.

Chemotherapy methods for treating liver cancer mainly include systemic chemotherapy and local chemotherapy. Systemic chemotherapy is usually administered orally or intravenously to allow chemotherapy drugs to enter the blood circulation and reach all parts of the body. It is suitable for cancer cells that have spread to other parts of the body. Local chemotherapy is more concentrated in the area where the cancer is located, such as Hepatic Arterial Infusion Chemotherapy (HAIC) [22, 23] and



Transcatheter Arterial Chemoembolization (TACE) [72, 73], which injects chemotherapy drugs directly into the arteries of the liver to reduce the impact on other parts of the body. TACE is a minimally invasive procedure primarily used for treating liver cancer by blocking the blood supply to the tumour. Since liver cancer nodules receive most of their blood supply from the hepatic artery, TACE involves injecting chemotherapy drugs and embolic agents directly into the hepatic artery to induce tumour necrosis and shrinkage. However, tumour hypoxia from TACE can stimulate new blood vessel formation, potentially leading to recurrence. Compared with TACE alone, TACE combined with Sorafenib can prolong the overall survival of patients with liver cancer [74]. Additionally, combining TACE and RFA in the treatment of hepatocellular carcinoma has been shown to improve outcomes compared with TACE or RFA alone. In [75] this combined approach resulted in better overall survival and recurrence-free survival. Studies have shown that for tumours larger than 3 cm, combination therapy achieves better results than RFA alone without significantly increasing major complications [76].

In summary, which method to use to treat liver cancer needs to be decided according to the specific situation of the patient, including the patient's condition (such as the size, number, and location of the tumour), liver function, physical condition, and whether he can bear possible side effects. This decision should be made jointly by the patient and the healthcare team.

## **2.4 Computed Tomography (CT) Technology**

### **2.4.1 Overview of Medical Imaging Modalities for Liver Diagnosis**

Medical imaging plays a vital role in the diagnosis, monitoring, and treatment planning of liver diseases. Among the numerous imaging modalities, computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound are the most widely used in clinical hepatology. These methods have their own advantages and disadvantages in terms of image resolution, soft tissue contrast, acquisition time, cost,

and availability.

CT imaging is widely used in liver disease diagnosis due to its high spatial resolution, fast acquisition time, and widespread availability in clinical settings. It provides detailed anatomical information and can precisely visualise the liver parenchyma, vasculature, and tumor location. CT is also the primary image source for several public datasets, such as the Liver Tumor Segmentation (LiTS) Challenge, making it a practical choice for algorithm development and benchmarking.

On the other hand, MRI has superior soft tissue contrast and functional imaging capabilities, including diffusion-weighted imaging (DWI) and dynamic contrast-enhanced scanning. MRI is particularly valuable in detecting small lesions and differentiating benign from malignant tumours. However, MRI is more expensive, difficult to obtain, and more susceptible to motion artefacts, which limits its widespread application in large-scale automated analysis.

Ultrasound is commonly used for initial screening and interventional guidance due to its low cost, portability, and real-time imaging properties. However, ultrasound suffers from low reproducibility, operator dependence, and limited image quality, which makes it less suitable for automatic segmentation tasks.

**Table 2.4 Overview of Common Medical Imaging Techniques for Liver Disease Diagnosis**

Modality	Strength	Limitations	Usage in Liver Tumor Segmentation
CT	High spatial resolution. widely available; fast	Radiation exposure. less soft-tissue contrast.	Preferred for automated segmentation.
MRI	Better soft-tissue contrast; no radiation	Expensive. longer acquisition time; motion sensitivity	Better for characterization but less used in public datasets.
Ultrasound	Real-time; low cost; portable	Operator-dependent; low image quality; poor reproducibility	Not suitable for automated segmentation pipelines.

A summary of the comparison of these imaging modalities is presented in Table

2.4. Based on this evaluation, CT is selected as the primary imaging modality for this study because it strikes a balance between imaging quality, clinical usability, and compatibility with large, annotated datasets. The next section will focus on CT imaging techniques and their application in liver cancer detection.

#### **2.4.2 Overview of CT**

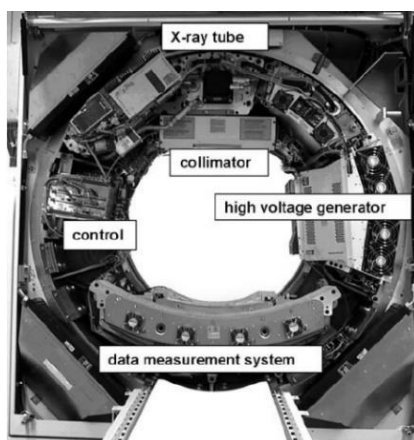
X-ray computed tomography is a method of obtaining structural information about biological tissues and engineered materials without destroying the sample, by using the sample's absorption of radiation energy, and it has the advantage of providing very detailed images of internal structures. CT imaging provides a very detailed image of the internal structure of the sample, which makes it an important tool in the medical field, especially in the diagnosis of diseases. CT imaging is irreplaceable in the diagnosis of diseases of the central nervous system, head and neck, chest, heart and abdomen. Particularly in tumour screening, CT has become an indispensable tool in the diagnosis and treatment planning of tumours because it can clearly show the tumour and its relationship to the surrounding tissue.

The basic components of a modern spiral CT imaging system include X-ray sources, detectors, scanners, and computer systems. The X-ray sources emit X-rays, which pass through the object to be inspected and are then received by the detector on the opposite side. The X-ray source and detector are fixed on the scanning frame, and they rotate at high speed around the axis of the object being scanned. This rotation allows the system to capture X-ray images from multiple angles, providing data for subsequent three-dimensional reconstruction. The X-ray attenuation information collected by the detector is transmitted to the computer system. The computer system processes these data and uses image reconstruction algorithms (such as back-projection algorithms) to reconstruct the two-dimensional projection data into a three-dimensional image to show the internal structure of the scanned object. The reconstructed images can be displayed directly on a computer screen for analysis by doctors or researchers. Alternatively, images can be recorded via specific output devices such as multi-frame cameras or laser

cameras for archiving or further analysis.

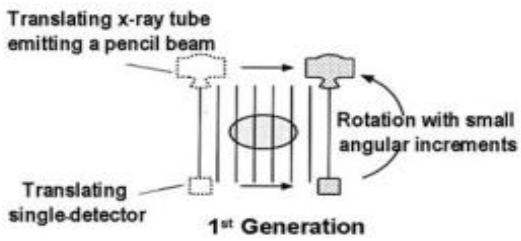
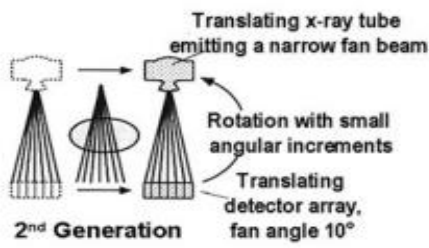
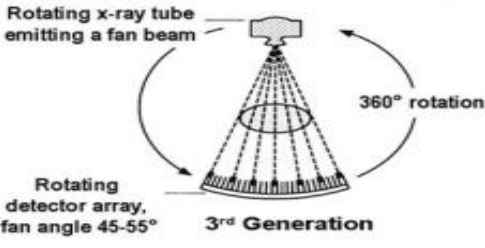
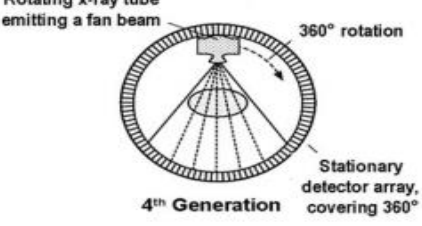
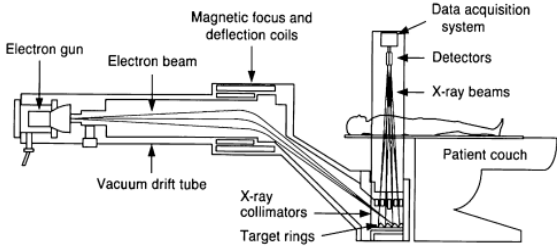
The development of CT technology has undergone significant advancements since its inception, with each generation introducing innovations that improved image quality, scanning efficiency, and clinical applications. Table 2.5 summarises the evolution of scanning methods, scanning time, and working principles from the first to the fifth generation of CT scanners.

Early CT technology used single-row detectors and single-beam rotational scanning, which had long scanning times and low image quality, and was mainly used for head imaging. Subsequently, multi-row detectors and fan-shaped beams were introduced, which increased the scanning speed and improved the image quality, and CT began to be used in more clinical scenarios. With the advancement of technology, the third-generation CT technology with rotating X-ray tubes and full detector arrays emerged, achieving full 360-degree scanning, greatly improving image resolution, and further expanding the scope of application. The fourth-generation CT reduces artefacts and enhances imaging accuracy through the design of a fixed detector ring. The subsequent electron beam CT is widely used in cardiac imaging due to its extremely high scanning speed [14, 77, 78]. The third generation CT technology is currently widely used in clinical practice. From , it can be seen that the key components of a CT scanner include an X-ray tube, a high-voltage generator, a collimator, a control system, and a data measurement system.



**Figure 2.4 Components of a modern third generation CT system. The fan beam covering a whole-body scan field of view with a diameter of 50 cm is indicated.**

**Table 2.5 Introduction to CT scanners from the first to the fifth generation, including scanning methods, scanning times and working principle [79, 80].**

Generation	Scanning method	Scan time	Working Principle
1 <sup>st</sup>	Single beam Scanning	5min	 <p>Translating x-ray tube emitting a pencil beam</p> <p>Translating single-detector</p> <p>Rotation with small angular increments</p> <p><b>1<sup>st</sup> Generation</b></p>
2 <sup>nd</sup>	Narrow angle fan beam Scanning	20~60s	 <p>Translating x-ray tube emitting a narrow fan beam</p> <p>Rotation with small angular increments</p> <p>Translating detector array, fan angle 10°</p> <p><b>2<sup>nd</sup> Generation</b></p>
3 <sup>rd</sup>	Wide Angle Fan Beam Scanning	5~10s	 <p>Rotating x-ray tube emitting a fan beam</p> <p>360° rotation</p> <p>Rotating detector array, fan angle 45-55°</p> <p><b>3<sup>rd</sup> Generation</b></p>
4 <sup>th</sup>	Wide Angle Fan Beam Scanning	1~5s	 <p>Rotating x-ray tube emitting a fan beam</p> <p>360° rotation</p> <p>Stationary detector array, covering 360°</p> <p><b>4<sup>th</sup> Generation</b></p>
5 <sup>th</sup>	Electron Beam Scanning	<1s	 <p>Electron gun</p> <p>Electron beam</p> <p>Magnetic focus and deflection coils</p> <p>Vacuum drift tube</p> <p>X-ray collimators</p> <p>Target rings</p> <p>X-ray beams</p> <p>Detectors</p> <p>Data acquisition system</p> <p>Patient couch</p> <p><b>5<sup>th</sup> Generation</b></p>

### 2.4.3 Imaging principle

CT technology is an imaging technique that enables non-invasive visualisation and assessment of subtle structures within the human body by scanning the body using X-rays. As X-rays pass through the body, they encounter different types of tissue (e.g. bone, muscle, fat, etc.), each of which absorbs X-rays to a different degree (attenuation coefficient). This attenuation is caused by three main physical processes: the photoelectric effect, the Compton scattering effect and the electron pair effect [25, 81].

This different results in the grey values of different areas on the CT image. Higher grey values correspond to lower X-ray attenuation and typically represent less dense tissues, whereas lower grey values correspond to higher attenuation and are generally associated with denser tissues. By scanning the body with X-rays at different angles, two-dimensional projections of the body at different angles can be obtained. These 2D projections contain information about the attenuation inside the body from all angles and implicitly about the spatial localisation. Subsequently, these 2D projections are reconstructed into 3D images using complex algorithms (e.g., inverse projection algorithms), which are able to show the detailed structure of the human body's interior.

Raw CT values are based on the intensity of the detected X-rays and can range from a few thousand to tens of thousands, depending on the scanner manufacturer and scanning parameters. In order to provide a standardized method in medical imaging to quantify the density of different tissues or substances on CT images. Raw CT values are usually converted into Hounsfield Unit (HU) values to facilitate medical diagnosis and analysis. HU is the quantitative scale unit of Hounsfield scale opacity, which is a linear mapping of the attenuation coefficient of the measured medium. Among them, the radiodensity of distilled water and air at standard temperature and standard pressure is defined as 0HU and -1000HU respectively. Its mathematical expression is

$$U = \frac{\mu - \mu_{water}}{\mu_{water} - \mu_{air}} \quad (2.1)$$

where  $\mu_{water}$  and  $\mu_{air}$  are the linear attenuation coefficients of water and air respectively.

In CT imaging, HU are a measurement used to represent the CT number of each pixel on the image, with value range (-3024, +1410) or (-1024, +1410) according to different CT machines. Table 2.6 shows the HU values of some common tissues.

In order to facilitate doctors to see different tissues and structures more clearly, the window width and window centre are adjusted to optimise the observation of different types of tissues. For instance, window settings specifically optimised for liver imaging enable better visualization of the liver and adjacent abdominal organs.

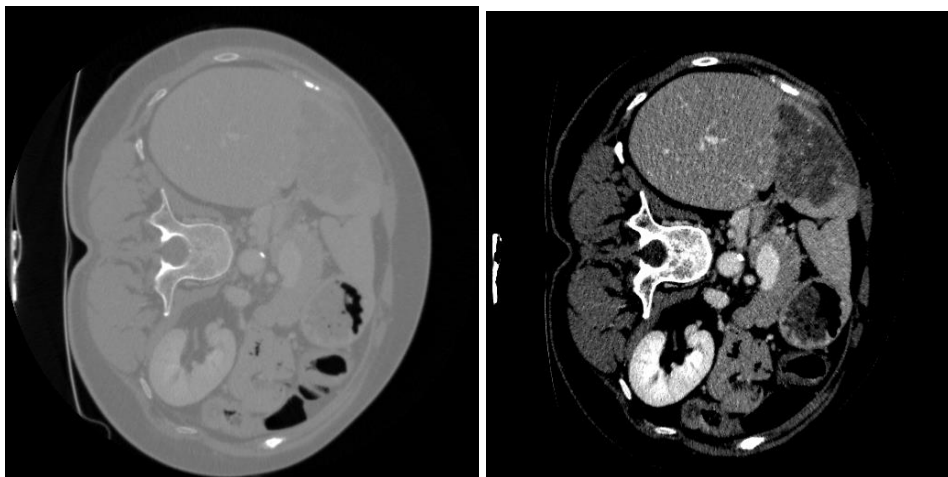
Window width in CT imaging defines the range of CT values displayed on the image. Within this set range, different tissues and lesions are represented in varying shades of gray. When the CT values of a tissue exceed this range, they appear as pure white on the image, regardless of how much they exceed, and the subtle differences in grayscales are no longer visible. Similarly, if the CT values are below this range, they will appear as pure black, with no variation in grayscales either. Increasing the window width will include a wider range of CT values on the image, allowing for the display of a greater variety of tissue structures, but the grayscale contrast between different tissues will be reduced. Conversely, decreasing the window width will narrow the range of CT values, resulting in fewer tissue structures displayed on the image, but the contrast between them will become more pronounced.

The window centre defines the midpoint of the window width range. The setting of the window centre determines the brightness of the image, since adjusting the window centre essentially changes the range of CT numbers mapped to grayscale display on the monitor. Usually, the CT value of the tissue to be observed is used as the window centre. By adjusting the window centre, the structures of interest (such as soft tissue, liver tissue, or bone) can be displayed with the correct brightness on the grayscale image, allowing for optimal visualization. Figure 2.5 shows the original CT image and the CT image after window transformation. It can be clearly observed that after window transformation, the differences between different tissues are more obvious, which is convenient for doctors to observe and perform data analysis using deep

learning.

**Table 2.6 HU values of different tissues**

Tissue	CT Number (HU)
Bone	+1000
Liver	40 - 60
White mater	-20 to -30
Blood	40
Muscle	10-40
Kidney	30
Water	0
Fat	-50 to -100
Air	-1000



**Figure 2.5 Original CT image and CT image after window transformation.**

#### **2.4.4 CT artefacts**

CT scan artefacts can originate from CT system hardware [25], patient interaction with the hardware, and fundamental physical limitations of CT imaging. Hardware-related artefacts in CT can be caused by issues such as X-ray beam hardening, changes in detector sensitivity, and mechanical inaccuracies in the scanner's moving parts. Patient-related artefacts can arise from patient motion, differences in tissue density, streaking or starburst effects due to metallic implants in the body, and the presence of



contrast media. Finally, the inherent physical limitations of CT imaging technology produce artefacts including beam hardening effects, scattering, partial volume effects, and quantum speckle, among others. If not handled properly, these artefacts can seriously affect the quality and analysis of medical images. Table 2.7 summarises the main CT artefacts according to different sources[25, 82, 83], including physics-based artefacts, patient-based artefacts and scanner-based artefacts.

**Table 2.7 Artefacts in CT scanning according to different sources**

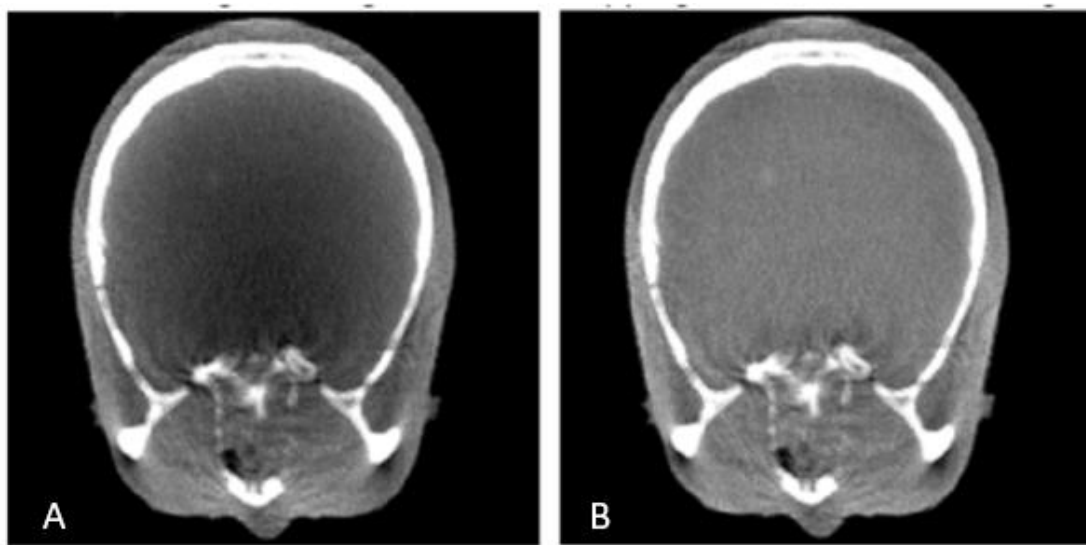
Physics-based Artefacts	Patient-based Artefacts	Scanner-based Artefacts
Wire harness hardening	Motion artefacts	Ring artefact
Streak artefacts	Clothing artefacts	Tube discharge artefact
Cupping artefact	Transient interruption of contrast	Bubble artefact
Metallic artefacts		Out of field artefact
Partial volume effect		Stair step artefact
Quantum spots		Zebra stripes
Under sampling artefacts		Cone beam artefact

#### **2.4.4.1 Beam Hardening Effect**

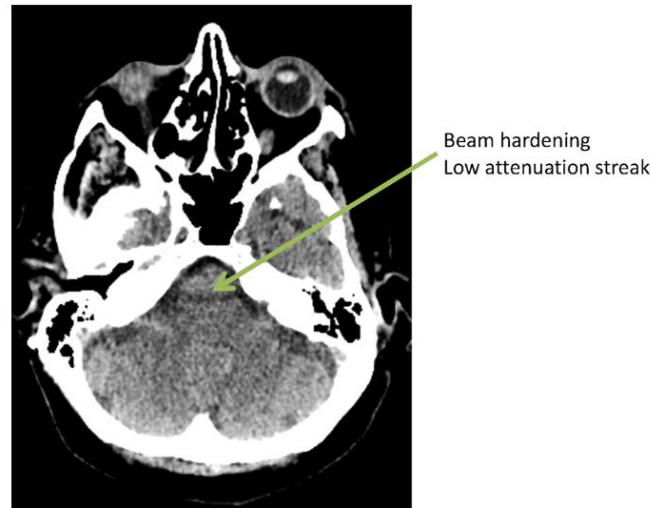
The beam hardening effect is a common physical artefact in CT imaging that occurs when X-rays pass through the object being scanned. Since X-rays of different energies pass through objects at different absorption rates, low-energy X-rays are more easily absorbed than high-energy X-rays, resulting in uneven distribution of remaining X-ray energy after penetrating the object. This uneven energy distribution will cause unrealistic image density in imaging, which is usually manifested as bright or dark bands near the edges of the image, especially when penetrating thick or dense objects [25]. Beam hardening effects have a negative impact on the quality and accuracy of diagnostic images and need to be corrected and mitigated through software algorithms or scanning techniques.

In CT imaging, the beam-stiffening effect caused by high-density materials (such as bone or iodine contrast media) may lead to two characteristic artefacts: streak (or dark band) artefact and cupping artefact. Streak artefact appears as dark stripes on the image, while cupping artefact causes the central area of the image to appear darker than the edges. These effects are particularly noticeable in brain scans, especially in the posterior fossa area due to dense the petrous bone structure is more prominent [84].

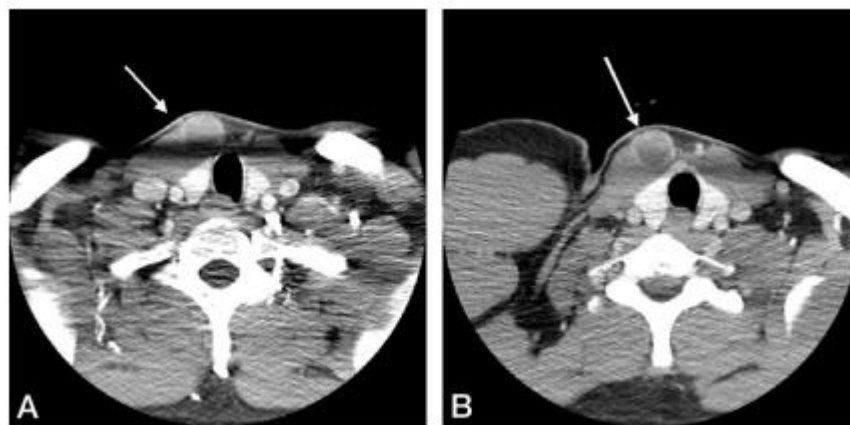
Figure 2.6 shows the cupping artefact in a skull model. It can be seen that due to the presence of cupping artefacts, the central area of the scan is darker than without artefacts. Figure 2.7 and Figure 2.8 show the streak artefact caused by beam hardening effect in the posterior fossa region and thyroid isthmus respectively. The appearance of stripe artefacts will obscure the existence of some tissues and directly affect the doctor's judgment.



**Figure 2.6 Cupping artefacts (A) in the skull model and artefact-corrected image (B)**  
[85]



**Figure 2.7 Streaks artefacts located in the posterior fossa region of the brain [86].**



**Figure 2.8 A, Beam hardening causing streak artefact at the level of the thyroid isthmus results in decreased soft-tissue contrast and obscures the vascular (arrow) B, clear vascular without artefact [87].**

#### **2.4.4.2 Partial Volume Effect**

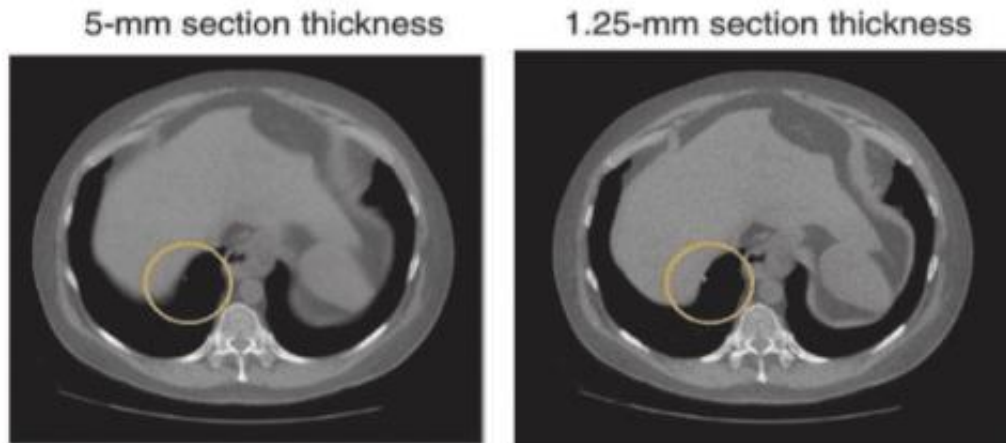
As described in Section 2.4.2, CT scanning is used to reconstruct cross-sectional images of structures in the body. Each cross-sectional image can be regarded as a slice, and the thickness of this slice is called the "layer thickness". Layer thickness is an important factor in CT imaging parameters that directly affects the quality of the image and the level of detail in the image. Thinner layer thicknesses provide higher image resolution and finer anatomical detail, helping to better identify and evaluate small

structures and lesions. However, thin layer thickness also means that longer scan times and higher radiation doses are required, while potentially increasing image noise. Therefore, while using thinner slice thickness can improve spatial resolution and increase the sensitivity of detecting smaller anatomical structures, it also increases image noise due to the reduction of photon statistics per voxel. It is crucial to strike a proper balance between image quality, scan time and radiation dose in CT imaging protocols.

The partial volume effect (PVE) is caused primarily by the inclusion of signals from multiple different tissues within a single voxel during imaging, resulting in imaging results that do not accurately reflect the true nature of any single tissue. When voxels are large or scanned with insufficient resolution, a voxel may contain many different tissue structures within it. Since the imaging process is performed based on voxels, if a voxel contains tissues of different natures within it, the signal reflected by this voxel is a mixture of the signals from these different tissues and does not accurately reflect the true nature of any one tissue. This results in blurring on the image, creating undesired artefacts.

The partial volume effect is very likely to occur in skull imaging, e.g., in cross-sectional CT reconstructions of the skull, where dark streaks are often connected to the rocky bone, and these artefacts can only be slightly reduced, even with double-overhardening correction. However, these artefacts can be significantly reduced by decreasing the slice thickness [88]. In addition, for aliasing artefacts in helical cone-beam CT (helical CT) image reconstruction, an algorithm was developed to understand these artefacts by simulating nonlinear partial volume averaging, and an interim scheme aimed at mitigating nonlinear partial volume and aliasing artefacts was proposed [89].

In liver CT imaging, partial volume effect produces artefacts mainly at the soft tissue-hard tissue or air interface, which include the bone-air and bone-fat interfaces. Figure 2.9 shows the improvement of artefacts in liver scans due to the partial volume effect by reducing the scan slice thickness.



**Figure 2.9** Partial volume artefact (circled in yellow) decreased with decreased voxel size [89].

#### **2.4.4.3 Photon Starvation**

During CT imaging, photon starvation [90] refers to an insufficient number of photons reaching the detector because of X-rays travelling through high-density areas or longer paths in the scanned object, due to the fact that CT scans are much more absorbent of X-rays than soft tissue when they involve areas that contain metals or other materials with high atomic numbers. The photon starvation effect can lead to projection data with large statistical errors in the reconstructed image, resulting in the formation of fine bright and dark bars along the direction of maximum attenuation, which severely affects the quality of the image reconstruction. They reduce the ability to recognise fine structures, increase uncertainty in the diagnostic process, and sometimes may even lead to misdiagnosis. For example, the photon starvation effect is particularly pronounced when evaluating patients containing metal implants, as metal attenuates X-rays much more than human soft tissue. Figure 2.10 shows a CT image of a shoulder phantom exhibiting streaking artefacts due to photon starvation.

To reduce the effects of photon starvation artefacts, several strategies can be employed. Some approaches include optimizing scanning parameters (e.g., increasing the X-ray dose or adjusting the scanning geometry), using advanced image reconstruction algorithms (e.g., iterative reconstruction techniques), and developing

specialised artefact reduction techniques that are designed to improve the quality of the image and thus the accuracy and reliability of the diagnosis. By these methods [83], the photon starvation effect can be compensated for to a certain extent and its impact on CT image quality can be reduced.

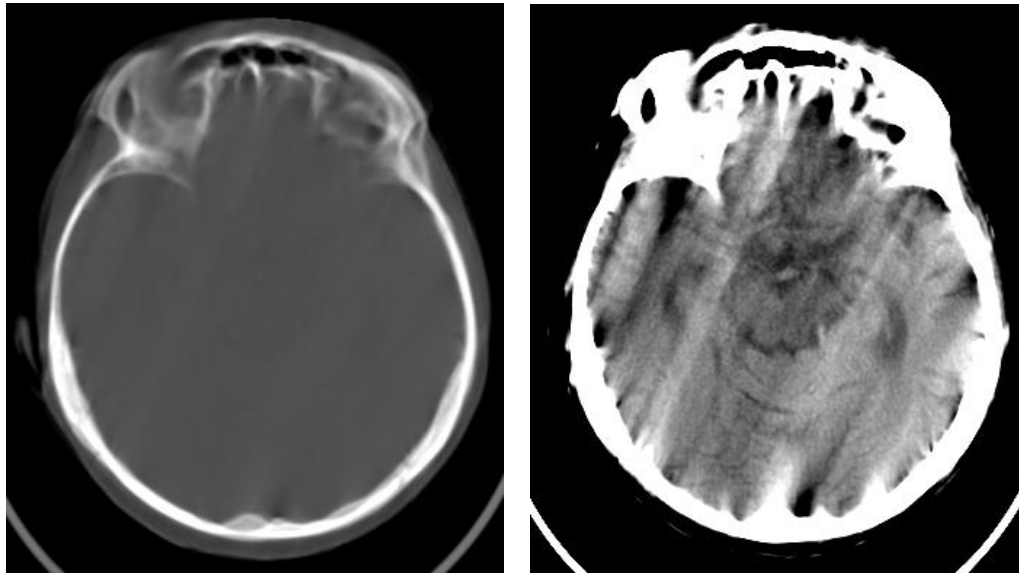


**Figure 2.10** CT image of a shoulder phantom shows streaking artefacts by photon starvation [82].

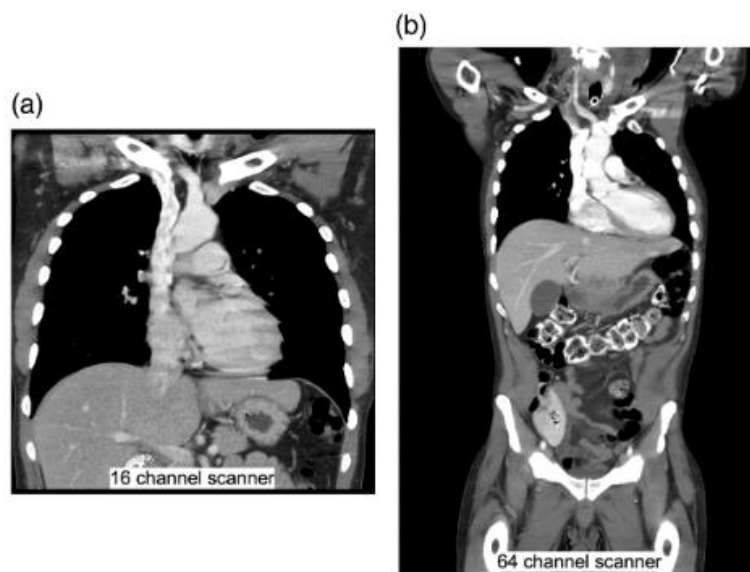
#### **2.4.4.4 Motion artefact**

Motion artefacts are a common problem during CT scanning, especially when imaging abdominal organs such as the liver. These artefacts can be caused by factors such as patient breathing, small movements of body parts or heartbeat, affecting the quality of the image and thus interfering with the diagnosis and assessment of disease.

Motion artefacts during liver scans are of particular concern because the liver is positioned close to the lungs and is susceptible to breathing movements. Motion artefacts can occur during scanning if the patient is unable to maintain sufficient respiratory stillness, or if the scanning time is long making it difficult for the patient to maintain stillness. These artefacts appear as blurring, streaking or ghosting on the image and may mask lesions or simulate non-existent lesions, making diagnosis difficult. Figure 2.11 to Figure 2.12 show some examples of motion artefacts in CT.



**Figure 2.11** Motion causes blurring and double images (left), as well as long range streaks (right) [25].

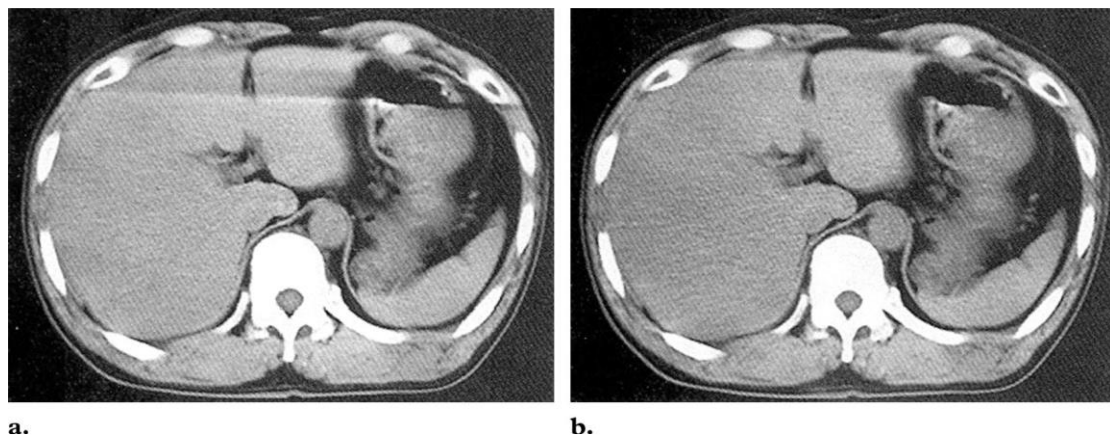


**Figure 2.12** Clinical benefits of increasing CT scanner speed by increasing the number of rows. (a) 16-slice coronal image showing inconsistent cardiac motion. These inconsistencies are greatly reduced in (b) the coronal image obtained from the 64-slice scan [91].

The use of high-speed scanning equipment is effective in reducing the occurrence of motion artefacts due to the fact that the patient only moves a small amount during image acquisition. This goal can be reached by accelerating the gantry rotation speed or increasing the number of X-ray sources [92]. Moreover, CT scanners that use more

detector rows can cover larger volumes in less time, thereby reducing artefacts caused by patient motion. For example, 256-slice or 320-slice CT can provide faster scanning speed than traditional 64-slice CT, which is especially important for cardiac imaging. Figure 2.13 shows the clinical advantages of increasing the number of detector rows for CT scanning. As the number of detector rows increases, CT scans gradually transform into a true volumetric imaging mode, allowing images to be reconstructed into sagittal and coronal views, minimising motion artefacts.

Patients are trained and instructed on breathing control prior to the scan and are taught how to maintain smooth breathing or hold their breath at specific moments during the scan to minimise movement and consequent artefacts [93]. This process is often combined with respiratory gating techniques which used external equipment to track and record the patient's breathing pattern in real time. These devices monitor the patient's breathing in real time and predict the phases of the respiratory cycle, allowing the imaging device to capture images at the optimal time. Imaging during the breath-hold phase minimises image blurring due to respiratory movements and improves diagnostic accuracy [24, 94].



**Figure 2.13 a. Streak artefacts due to motion in abdominal scanning and b. Abdominal scanning results under ideal conditions [82].**

On the other hand, reconstruction techniques for CT images have a direct impact on the effectiveness of scanning, from filtered back projection (FBP) proposed in the 1970s to today's iterative reconstruction algorithms in image space. With the increase in computational power and advances in algorithms, iterative reconstruction techniques



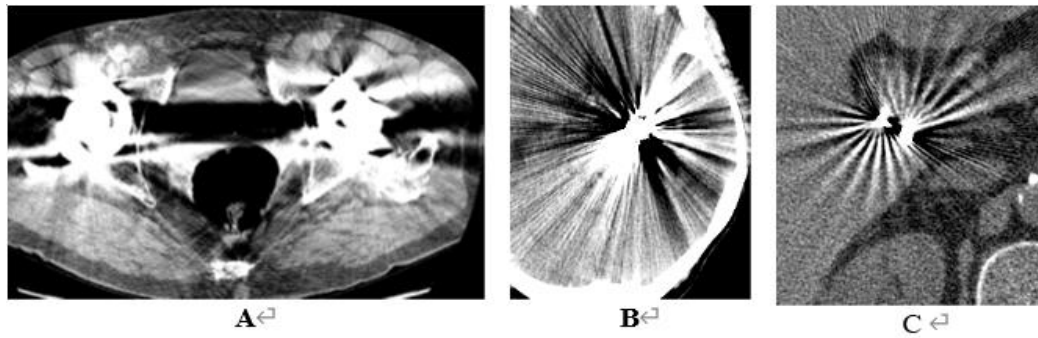
have become the key to improving image quality at low radiation doses [95, 96].

#### **2.4.4.5 Metal Artefact**

In liver CT scans, metal artefact is a common problem, occurring in 21% of a series of scans [97], which can significantly affect image quality and interfere with correct diagnosis. The generation mechanisms of metal artefacts are complex and diverse, including problems caused by the metal itself and multiple factors related to the metal edge. Figure 2.14 shows some examples of metal artefacts in different organs.

Artefacts produced by the metal itself are primarily caused by metallic substances within the patient's body, such as surgical implants, dental fillings, or metal clips. The absorption and scattering of X-rays by these metal objects in CT scans is much greater than that of human tissues, resulting in stripe-like artefacts caused by the beam hardening effect or the photon starvation effect, which has been clearly introduced in previous sections.

The high contrast at the edges of metallic objects requires very high spatial resolution to be accurately captured. In actual CT scans, such high contrast variations cannot be fully recorded due to the limited sampling rate. As a result, metal edges cannot be accurately reproduced in image reconstruction, leading to significant streaking artefacts. For example [98], the high-contrast sharp edges characteristic of metal crown borders may often appear in the oral cavity. In addition, when using a rotating CT scanner, image discontinuities created by the projection of metal edges at different angles can lead to streaking artefacts similar to the rotation of windmill blades in the image reconstruction [99].

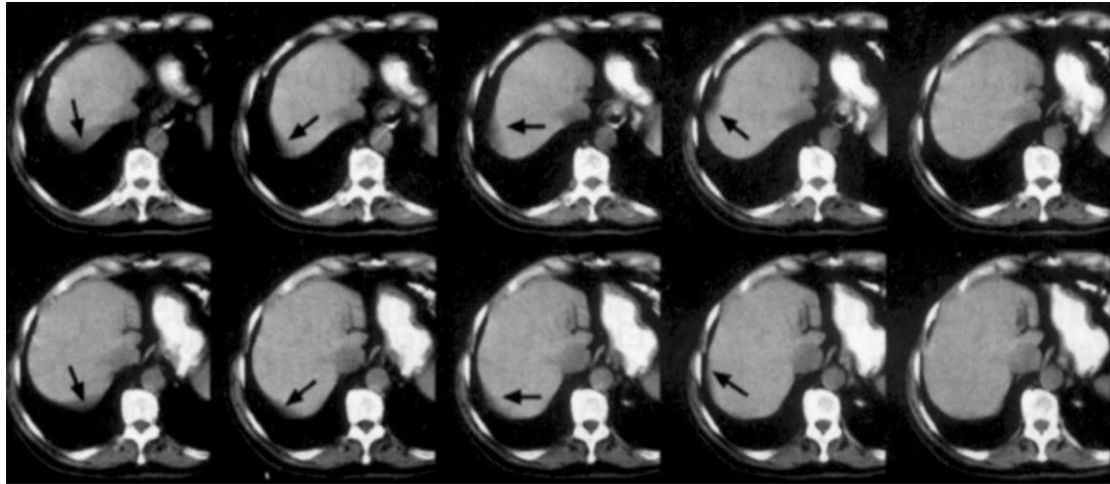


**Figure 2.14** A. Dark streak between hip replacements is mostly due to beam hardening and scatter. B. Sharp thin alternating streaks surrounding an aneurysm coil are mostly due to motion and under sampling. C. Smoothly undulating streaks around cholecystectomy clips are due to windmill artefact [25].

#### 2.4.4.6 Scanner Artefact

Ring artefacts are a common type of artefact in CT images. This type of artefact is usually caused by the uneven response to incident X-rays when the detector unit of a CT scanner is inaccurately calibrated or damaged. Ringing artefact appears as a series of concentric rings on the image, which may affect image quality and interfere with doctors' diagnosis according to the image. Methods to address ring artefacts typically include hardware maintenance and calibration, such as replacing or recalibrating damaged detector units, and image post-processing using iterative reconstruction algorithms to reduce or eliminate these artefacts.

In helical CT scans, some unique artefacts may appear due to the unique scanning method and data reconstruction process. The X-ray source and bed rotate and advance simultaneously during scanning. This scanning method allows for continuous and rapid collection of volumetric data. Collected data come from different planes need to be integrated into the same plane through interpolation algorithms for subsequent image reconstruction. Specific weighting functions used in helical interpolation algorithms may distort the shape of the reconstructed image. This causes the direction of the artefact to change with the position of the X-ray source at the centre of the image plane. In clinical images, this helical artefact can sometimes be misinterpreted as pathological changes, as shown in Figure 2.15.



**Figure 2.15** Series of CT images from a helical scan of the abdomen shows helical artefacts (arrows) [25].

## 2.5 Conclusion

In summary, this chapter provides an overview of the anatomical and physiological structure of the liver, as well as the pathological mechanisms, etiological factors, clinical staging systems, and current treatment options for liver cancer. A thorough understanding of these medical foundations is essential for the development of clinically meaningful image analysis techniques. In addition, the role of CT in liver disease diagnosis is examined, including its imaging principles, acquisition techniques, and the common artefacts encountered in abdominal imaging. These insights lay the foundation for the development of automated liver and tumour segmentation algorithms based on medical and imaging knowledge. In general, early and accurate diagnosis of liver cancer, along with effective treatment planning supported by advanced imaging technologies, plays a critical role in improving patient outcomes.

## **Chapter 3. Image segmentation and quantification techniques based on CT Data**

### **3.1 Introduction**

The previous chapter introduced the physiological characteristics of the liver and the causes and treatments of related cancers. Appropriate diagnosis and treatment of liver cancer at an early stage can provide patients with a higher chance of successful treatment, the possibility of minimally invasive surgery, and an overall improved prognosis. Early liver cancer can usually be treated with surgery to remove the tumour (resection) or with a liver transplant if the cancer is localised to the liver and has not spread. Accurate identification and quantification of liver structures through CT segmentation is critical for surgical planning and postoperative evaluation.

This chapter is organised as follows,

Section 3.2 reviews some conventional medical image preprocessing methods and emphasize the preprocessing methods of liver CT images. Section 3.3 describes the progression of segmentation methods from labour-intensive manual methods to sophisticated automated techniques. Section 3.4 reviews the development history of deep neural networks and its achievements in semantic segmentation.

### **3.2 Image preprocessing techniques**

Medical image data is usually stored in NIFTI (Neuroimaging Informatics Technology Initiative) format or DICOM (Digital Imaging and Communications in Medicine) format files. Table 3.1 gives the information contained in both files. DICOM files have more detailed data and general acceptability in clinical settings, while NIFTI provides simplicity and ease of use for research and analysis.

Before performing image processing operations such as image segmentation, the data needs to be converted into an appropriate image format in order to obtain visualization results. A suitable library, such as SimpleITK, is used to read the data and

convert it into common image formats such as JPG or PNG. During this period, the size, orientation or contrast of the image can be adjusted to facilitate subsequent image processing operations. For liver and tumours data, CT images are taken from the patient's abdominal scan results, and the data also need to be appropriately selected, sorted and organised to ensure their coherence during processing and reduce data redundancy. The selected label data must undergo noise reduction processing to eliminate artefacts introduced during image saving or compression. These artefacts may result in non-binary pixel values along the label boundaries. To ensure training consistency and model interpretability, all label masks are cleaned to contain only two discrete values (0 and 1, or 0 and 255).

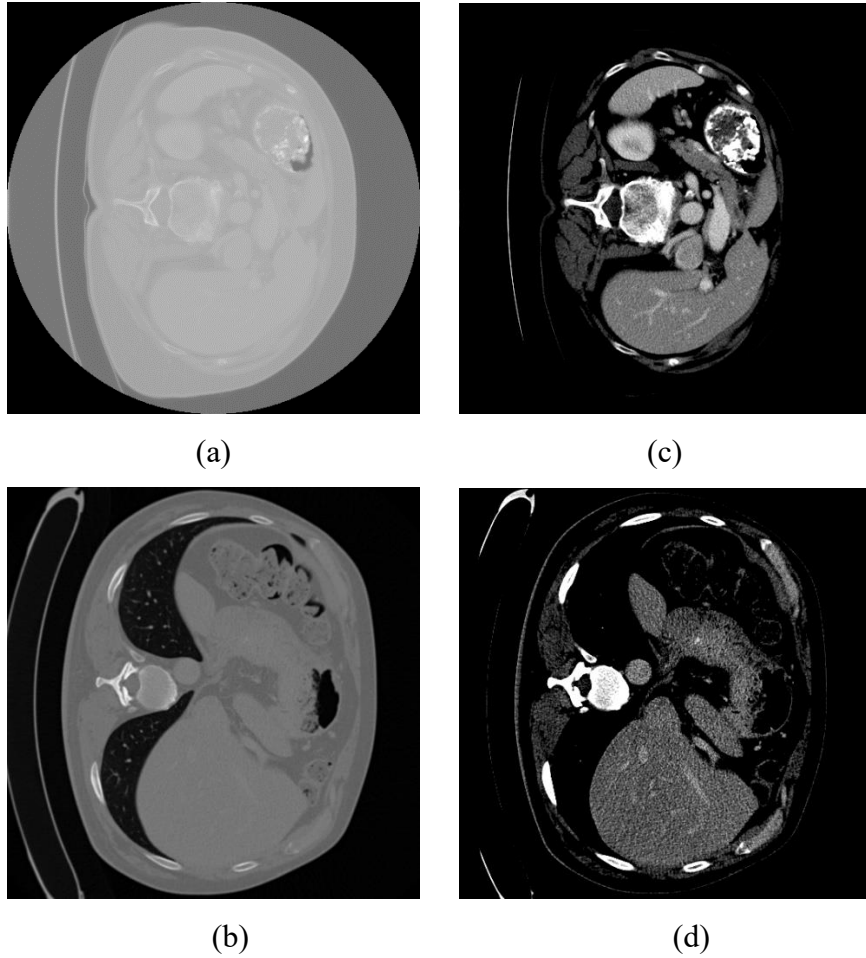
**Table 3.1 Comparison between NIFTI file and DiCOM file commonly used in medical imaging.**

Feature	NIFTI File	DICOM File
Primary Use	Neuroimaging	Clinical radiology, PACS systems
Data Format	.nii or .nii.gz file	.dcm file
Metadata Support	Limited metadata, including image dimension, voxel size and spatial orientation	Rich metadata (patient/equipment info, medical records)
Time Series Support	Yes	Yes
Patient Data	Not included	Included
Image Orientation	Fully supported	Fully supported
Interoperability	Research-focused tools	Widely supported in clinical settings

Image normalization [100] is an important image preprocessing technique that aims to improve the quality of image data and make it more suitable and consistent for further analysis and processing. Normalization can improve the contrast of images and make details clearer, especially in medical imaging, which helps to better observe and analyse structures and lesions. The consistency of normalized image data is critical to measuring the accuracy and efficiency of deep learning algorithms. Common image

normalization operations include histogram equalization operations, non-linear changes in pixels (gamma changes), image whitening, as well as other related methods. [101, 102].

For CT images, it is necessary to perform a unified standardization operation on the image by adjusting the window width and window level to achieve the best visualization standard. As given in Eq. 2.1, CT values are usually expressed in Hounsfield Units (HU). Depending on the scanner model, the scan results fluctuate within the range of (-3024, 1410). Due to inherent differences in CT values between different anatomical structures and pathological changes, appropriate window widths and window centres need to be selected for specific tissue structures or pathological observations. Window width is a parameter that determines the range of CT values displayed in the image. Tissues and lesions within this range are displayed in different grayscales. Increasing the window width will reveal more tissue structures of different densities, but the grayscale differences between these structures will be reduced. This is useful when you need to view an entire structure, such as viewing a large area of tissue in a chest or abdominal scan. Reducing the window width will reduce the displayed tissue structure, but improve the grayscale difference between each structure, which is beneficial to detailed observation of specific tissues or lesions, such as tumours or small blood vessels. The window level is a parameter that sets the grayscale centre of the image. It determines which CT values will be mapped to the midtones of the grayscale image. Adjusting the window level can help doctors see details of specific tissues more clearly. Figure 3.1 shows the abdominal CT images before and after window level adjustment. It can be seen that after window level adjustment, the liver tissue is clearer and has high contrast.



**Figure 3.1** (a) and (b) Abdominal CT images before and (c) and (d) after window level adjustment, the first column is the original image, and the second column is the adjusted image.

### 3.3 Classical image segmentation techniques

The segmentation technology of liver CT images has developed over time from manual segmentation to semi-automatic segmentation to fully automatic segmentation. These advancements aim to improve the accuracy, efficiency and versatility of liver tumour detection and delineation. Table 3.2 summarises the key segmentation techniques used in medical image analysis, highlighting their respective advantages and disadvantages.

**Table 3.2 Summary of medical image segmentation techniques.**

<b>Techniques</b>	<b>Description</b>	<b>Advantages</b>	<b>Disadvantages</b>
Manual Segmentation	Manually delineate the boundaries of the liver and lesions, relying on experienced radiologist[10, 16]	The highest level of precision.	Time-consuming and labour-intensive. Subjective results.
Interactive Segmentation	Combines manual initialization and automatic segmentation algorithms.	Segmentation efficiency is improved while still retaining a degree of user control.	Requires a certain degree of manual intervention. Sensitive to the user's initialization quality.
Automatic Segmentation	Utilise advanced computer vision and machine learning techniques.	Improve segmentation speed and consistency. Reduce manual intervention	Face the challenges of data quality, algorithm generalization ability, and handling complex situations.

Manual segmentation was the earliest method, with the operator using computer software tools to manually outline the boundaries of the liver and lesions. This method requires significant time and effort while being highly reliant on the operator's knowledge and experience. The precision offered by this technique is seen as gold standard, but the process is subjective and labour-intensive [103]. In order to improve segmentation efficiency and accuracy, interactive segmentation methods were developed. Interactive segmentation reduces the degree of manual intervention but still requires some manual effort. Common interactive techniques include threshold segmentation, region growing, graph cut, etc [104]. These methods usually require user initialization (such as selecting seed points or providing a rough area), and then the algorithm automatically performs segmentation. Interactive segmentation improves efficiency while still retaining some user control. With the development of computer



vision and machine learning technology, fully automatic segmentation methods have become a research hotspot. Early fully automatic methods relied on traditional image processing techniques such as edge detection, texture analysis, and morphological operations. In recent years, deep learning, especially convolutional neural networks, has achieved remarkable results in fully automatic segmentation. The U-Net architecture [105] and its variants are widely applied in medical image segmentation.

### **3.3.1 Interactive Segmentation Method**

In terms of image processing, region-based segmentation techniques [106, 107] were among the earliest methods proposed to effectively delineate different regions in an image by utilising the similarity or continuity between pixels. Among them, threshold segmentation appears as a basic method. Its technical simplicity, small computational requirements [104] and good performance make it a typical and widely used technology in the field of image segmentation. Pixels belonging to different targets or regions often exhibit obvious grayscale differences, which can be clearly observed in the image histogram. The division of regions usually corresponds to different histogram peaks, and the optimal threshold selection lies in the valleys that separate these peaks. Region growing is another commonly used segmentation method based on intensity. The algorithm starts from a user-specified seed point (or an automatically selected seed point) and iteratively incorporates neighbouring pixels that are similar in intensity to the seed point into the growing region. This method has been effectively applied in liver and tumour segmentation, where a seed point is often selected in the centre of a tumour and the region expands based on similar intensities in the surrounding tissue [108, 109].

The significant advantages of intensity-based segmentation lie in its simplicity of implementation and its efficacy in segmenting images when different objects exhibit significantly different grayscale values or other characteristics. On the contrary, the method has obvious limitations, especially in its applicability to multi-channel images or images with small differences in grayscale intensities or other local features.

Accurate segmentation of images that lack obvious grayscale differences or in which the grayscale value ranges of different objects overlap significantly remains a difficult challenge. Notably, in the context of liver CT segmentation, liver regions and other adjacent organ regions or tumours exhibit distinct parenchymal heterogeneity, with minimal overlap among their respective tissue characteristics, making threshold segmentation a feasible strategy [110]. However, the inevitable contact between liver tissue and adjacent muscle tissue, and the consequent similarity in boundary pixel values, often leads to unsatisfactory segmentation results [111].

Beyond basic intensity-based segmentation method, more advanced methods have been developed to address the limitations of simple thresholding. Wang et al. [112] proposed a liver segmentation method that calculate multiple thresholds based on the slope difference distribution. Based on these thresholds, the method segmented CT slices into meaningful regions and delineated the liver boundaries by incorporating surrounding tissue constraints, energy minimization, and morphological operations, ultimately enabling accurate three-dimensional reconstruction of the liver. Similarly, Sangeeta et al. [113] also employed slope difference distribution to determine multiple appropriate thresholds for organ differentiation. Subsequently, seed points for boundary expansion were identified using Sethian's Fast Marching Method, a numerical algorithm that simulates the monotonic propagation of boundaries.

Semi-automatic segmentation methods based on graph cuts are a powerful tool to address the limitations of traditional intensity-based methods, especially when dealing with complex boundaries and ambiguous intensity regions. The graph cut algorithm was originally proposed by Yuri and his colleagues [114]. This algorithm models image segmentation as an energy minimization problem on a graph, where pixels (or super pixels) are regarded as nodes connected by weighted edges. The user labels some pixels as background or target to provide hard constraints for segmentation, while soft constraints include boundary smoothness and regional intensity consistency [115]. Based on these constrains, an appropriate energy function can be established as well as using the maximum

flow or minimum cut algorithm to find the best segmentation result of the CT image, which is usually better than local intensity-based methods. Based on this, many researchers have proposed effective and reliable algorithms for liver segmentation [116-118].

In the field of liver segmentation, many studies have proposed effective improvements on this algorithm. For example, the method proposed by Liao et al. [119] effectively excluded complex background and highlights the liver region by combining an intensity-based model and an appearance model based on principal component analysis (PCA)[120]. Using intensity information, local context and spatial correlation of adjacent slices, combined with the position information of adjacent slices, the liver in each CT slice was automatically segmented through graph cutting technology. Ahmed et al. [121] proposed a method that based on case-specific knowledge, which estimated the shape and intensity information of the liver through the spatial relationship between adjacent CT slices. This information was integrated into a shape-based graph cut framework to segment the entire CT volume without the need for pre-built models or training data. The method was evaluated on ten CT scans containing various liver abnormalities and showed high accuracy, low user interaction requirements, and short processing time. In addition, Lu et al. [122] introduced a method that integrates multi-dimensional features and shape constraints into a graph cut framework. The method first estimated the initial liver shape through multi-atlas segmentation and then automatically constructed a graph based on unsigned distance fields without manually annotating seed points. Combining shape constraints and multi-feature information, the method can effectively refine the liver boundary and achieve excellent segmentation accuracy on public CT datasets.

### **3.3.2 Fully Automatic Segmentation**

Fully automatic segmentation technology reduces user intervention to increase efficiency and objectivity. However, they may be deficient in handling pathological or unusual situations and require manual tuning to ensure accuracy in these complex scenarios. The next part will introduce the application of automated segmentation

algorithms on medical images from two different directions.

### 3.3.2.1 Model-Based Segmentation

Deformable models have proven to be effective and powerful in many medical applications [123]. These models can be broadly divided into three categories as described in Table 3.3: Statistical model based, Geometric model based, and Physical model based [124-128].

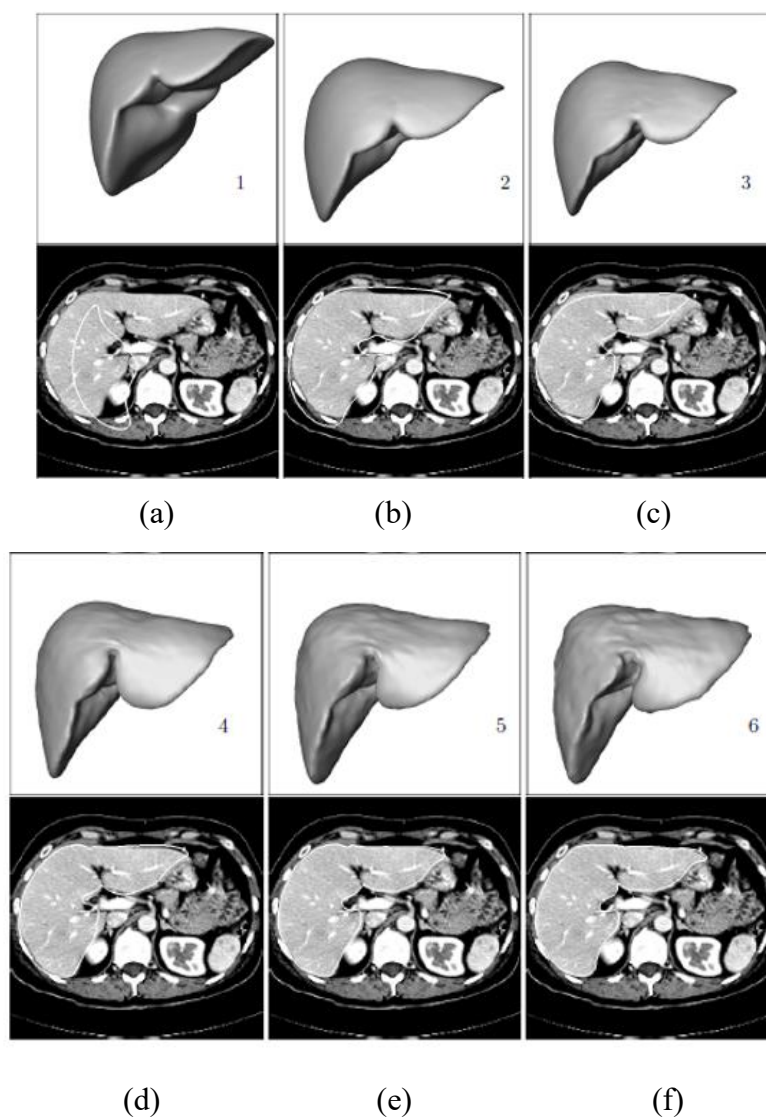
Statistical shape models (SSMs) [129] are mainly used to analyse and understand the shape variation of anatomical structures between different individuals. By learning the statistical properties of shape from a set of training data, SSM can accurately capture and simulate the normal range of variation in target anatomy.

**Table 3.3 Summary of Model-based segmentation algorithm.**

Model	Description	Application
Statistical model based	Learn the statistical distribution of shape variability and appearance features	Statistical Shape Models. Active Appearance Models. Point Distribution Models
Geometric model based	Adjust geometric properties to fit the target edge.	Active Contour Models; Snakes. Level Set.
Physical model based	Simulate physical and mechanical properties to adapt to the target shape.	Elastic model. Fluid model.

Liver CT image segmentation using SSM can utilise prior shape information to guide the segmentation process, even when the image quality is poor, or the boundaries are unclear [130]. Although there are inter-individual variations in the shape of the liver, these variations can be quantified through statistical methods. By analysing shape variation in a series of training images, SSM is able to capture the main changing patterns of liver shape. SSM exploits this variability to generate a model that predicts organ shape in new images. During the segmentation process, the model adapts to the

organs in a specific image based on statistics derived from the training dataset. Figure 3.2 shows how to apply SSM for liver CT segmentation. First, an average model is located in the CT image data-show in Figure 3.2(a, b, c), usually based on a previously constructed statistical shape model. The goal of initial positioning is to achieve approximate alignment between the model and the target anatomy. To this end, an initial optimisation of the model's positional parameters is performed to improve the alignment. By progressively increasing the number of shape patterns involved in the optimisation process, the model's fit to the target structure is gradually refined until optimal alignment is achieved, as shown in Figure 3.2(d, e, f).



**Figure 3.2 Segmentation process of data sets in statistical liver models. As the number of shape modes increases, the results of the optimised combination of position and shape gradually tend to the true contour [125].**

Figure 3.2 illustrates the progressive refinement process in liver segmentation using a statistical shape model. The top row (1–6) displays the 3D liver models generated at different stages of the optimisation, while the bottom row ((a)–(f)) shows their corresponding alignments with 2D CT slices. The process begins with an initial coarse model, often based on the mean liver shape derived from the training set. As the number of shape modes incorporated into the optimisation increases (from model 1 to model 6), the shape and position of the liver model are incrementally adjusted. This leads to progressively better alignment with the anatomical boundary observed in the CT images. By stage (f), the model closely approximates the true liver contour.

Active contour model is an important method for image segmentation, especially suitable for liver segmentation. This method uses curve evolution to detect targets and obtain accurate edge information. The basic principle is to define an initial curve, then construct an energy function based on the image data and change the curve by minimising this energy function so that it gradually approaches the target edge. The advantage of this method is that it produces closed, smooth edge curves.

The "Snake" model proposed by Kass et al. [131] is a classic parametric active contour model, which moves the curve to the target edge through the joint action of internal and external forces. It has the advantages of strong interactivity and fast implementation, but it is difficult to handle topological changes [132]. The geometric active contour method using the level set [133] framework effectively solves this problem by allowing the contour to naturally split or merge, making it more flexible for handling complex shapes or multiple objects in medical images. Applying these models in liver segmentation allows you to flexibly select the constraint force, initial contour, and scope to obtain better segmentation results. This makes active contour models have received a lot of attention in medical image processing tasks such as liver segmentation. For instance, Li et al. [134] used a region-based level set method for liver segmentation in CT images, which effectively integrated region and boundary information to handle low-contrast images. In the work of Li et al. [135], a liver segmentation framework was

developed that integrates a level set method with intensity bias correction and sparse shape composition (SSC), followed by a graph cut refinement step to address under-segmentation in pathological livers.

Physics-based liver segmentation methods incorporate biomechanical models or physical constraints to simulate tissue deformation, organ motion, or inter-organ interactions. These methods are beneficial in respiratory motion, surgical planning, or image registration across different phases or modalities. Common approaches include finite element modelling (FEM), mass-spring systems, and elastic or viscoelastic deformation models [136, 137]. Elastic models [127] treat the liver as a deformable solid, enabling the simulation of elastic tissue responses under mechanical forces, while fluid-based models represent the liver as a viscoelastic medium, allowing for continuous deformation under external constraints. By integrating prior knowledge of organ mechanics, these methods aim to enhance segmentation robustness and anatomical plausibility, particularly in complex clinical scenarios. It is worth noting that many physics-based methods in liver imaging focus primarily on registration or surgical guidance, rather than direct segmentation. While these approaches incorporate biomechanical models to simulate tissue deformation, their primary role is often to align preoperative and intraoperative images rather than delineate organ boundaries.

### **3.3.2.2 Learning-Based Segmentation**

The development of learning-based methods in the field of image segmentation is a typical manifestation of the progress of machine learning technology. The core of these methods is to automatically recognise and exploit complex patterns in image data to distinguish different samples, thereby achieving effective image segmentation.

Clustering algorithms, such as K-means, perform segmentation by grouping pixels or image regions into clusters with similar characteristics. These algorithms are often based on characteristics such as intensity, colour, or texture of pixels. Ramin et al. [138] used Fuzzy C-means to segment liver and tumour combined with Kirsch filter and mean

averaging algorithm to extract and smooth organ edges. Shraddha et al. [139] used K-means method with a special localised contouring algorithm to process liver CT segmentation.

However, clustering methods still face some limitations when applied to liver segmentation. Clustering algorithms depend on the initial selection of clusters, and their performance can vary significantly based on this initialization. Furthermore, the traditional fuzzy C-means algorithm, which relies on Euclidean distance, is not optimal for non-spherical structures like the liver as described in [140]. Additionally, the presence of noise in the image data can further impact the accuracy of these clustering methods [141].

Bayesian networks are probabilistic graphical models used to simulate complex relationships between pixels or regions in an image. They can effectively perform image segmentation by building a probabilistic model that describes the dependencies between pixel or region attributes. This enables them to distinguish pixels with similar gray levels that belong to different tissue types and to accurately assign them to their respective anatomical regions. Bayesian networks can integrate prior clinical and imaging knowledge into the segmentation process, improving the robustness of the segmentation results. In addition, they allow for probabilistic reasoning, offering not just segmentation results but also quantifiable confidence levels, which can be crucial for clinical decision-making [142].

The algorithm proposed by RG Mohamed [143] repeatedly applies multiresolution smoothed Bayesian classification followed by adaptive morphological operations and active contours refinement. M Freiman [144] proposed an adaptive fully automatic liver segmentation method for MRI images based on thresholding and Bayesian classification. However, as stated by Kyrimi et al. [145], Bayesian networks are rarely used in routine clinical practice despite hundreds of relevant papers, so there is still a large gap in implementation.

Support vector machine (SVM) is a powerful classifier that can be used for image



segmentation, especially when the number of samples is small. SVM distinguishes different categories of pixels or regions by finding an optimal boundary in multi-dimensional space. Suhuai Luo et al. [146] demonstrated that the combination of morphological operations and pixel-level support vector machine classifiers can accurately delineate liver volume. However, with the rise of deep learning, deep learning models can process more images in a short time and have higher accuracy compared to single SVM models [147].

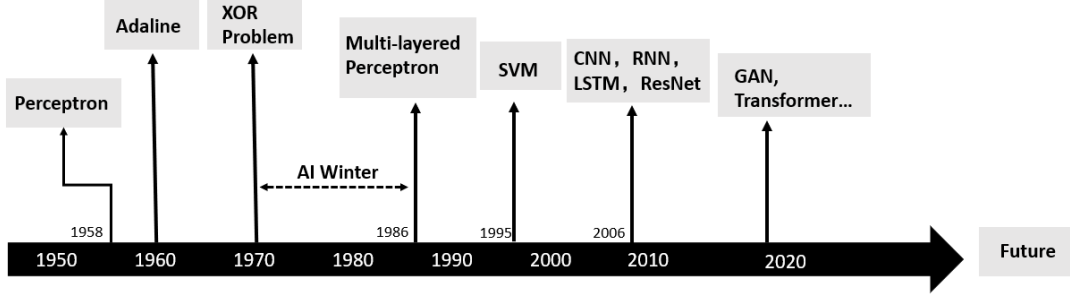
With the continuous advancement of machine learning technology, these learning-based image segmentation methods are becoming more and more mature and capable of handling more complex and diverse image segmentation tasks. These methods not only improve the accuracy of segmentation but also improve the efficiency of automated image analysis to a great extent.

### **3.4 Deep Neural Network for image segmentation**

Research in the field of AI has experienced many ups and downs, from the perceptron in the 1950s to the introduction of the multi-layer perceptron and back propagation algorithm in the 1980s [148-150]. The invention of the perceptron was one of the earliest algorithms designed to simulate the function of biological neurons. However, due to its limitations in solving nonlinear problems, the field stagnated in the 1970s and 1980s, known as the "AI winter". It was not until the 1980s that the introduction of multilayer perceptron and the application of backpropagation algorithms enabled complex nonlinear problems to be solved, which revitalized the field. In the 1990s, the development of support vector machines (SVMs) provided a powerful alternative for classification tasks, greatly enhancing people's confidence in machine learning to cope with diverse real-world challenges.

In the 21st century, the rise of deep learning has brought about rapid innovation in algorithms and technologies. From convolutional neural networks (CNNs) to recurrent neural networks (RNNs), to generative adversarial networks (GANs) [151] and Transformers [152]. They have made breakthrough progress in many fields such as

image recognition, natural language processing, game strategy, medical diagnosis, and autonomous driving. The success of deep learning is due in part to the availability of big data, significant improvements in computing power, and continued innovation in algorithms. These factors jointly promote the rapid development and widespread application of deep learning technology, leading a new wave of artificial intelligence.



**Figure 3.3 History of AI research from 1958 to the present [149, 153, 154].**

### 3.4.1 Deep Neural Network

Artificial neuron is the basic unit that constitutes a neural network. It receives a set of input signals and generates output by simulating the structure and characteristics of biological neurons. Figure 3.4 depicts a typical neuron structure diagram. The neuron first multiplies each input with its corresponding weight and adds the bias term, then performs a nonlinear transformation through the activation function and finally outputs the result.

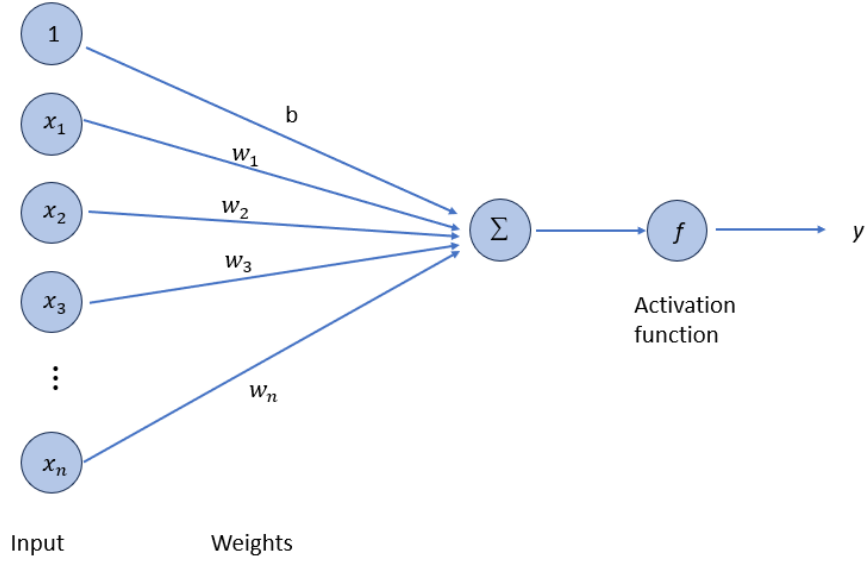
Suppose a neuron receives  $N$  inputs  $x_1, x_2, x_3, \dots, x_n$ , let the vector  $x = [x_1, x_2, x_3, \dots, x_n]$  represent this set of inputs, and let the net input  $z \in \mathbb{R}$  represent the weighted sum of the input signals  $x$  obtained by a neuron,

$$z = \sum_{n=1}^N w_n x_n + b = w^T + b \quad (3.1)$$

Where  $w = [w_1, w_2, w_3, \dots, w_n] \in \mathbb{R}$  is the weight vector of  $N$  dimension,  $b \in \mathbb{R}$  is the bias. After passing through a nonlinear function  $f(\cdot)$ , the value  $z$  obtains the activation value of the neuron  $y$ .

$$y = f(z) \quad (3.2)$$

Where the non-linear function  $f(\cdot)$  is called the activation function.



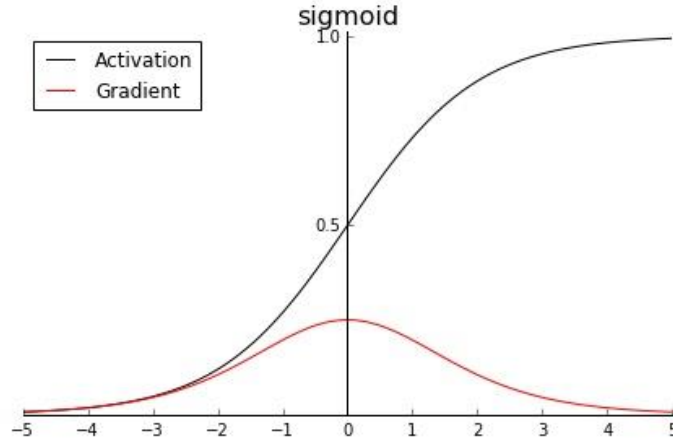
**Figure 3.4 A typical neuron structure.**

The activation function introduces nonlinearity into the neural network, allowing the network to learn and represent more complex relationships. It is a continuous and differentiable nonlinear function that enables the direct use of numerical optimisation methods to learn network parameters. On the other hand, the activation function and its derivative function should be as simple as possible, and their value range should be within a suitable interval, otherwise it will affect the efficiency and stability of training. The following introduces several activation functions commonly used in neural networks.

### 3.4.1.1 Activation functions

The sigmoid function (Logistic function) is a non-linear function with an “S” curve shape between 0 and 1. Its mathematical expression is as follows.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (3.3)$$



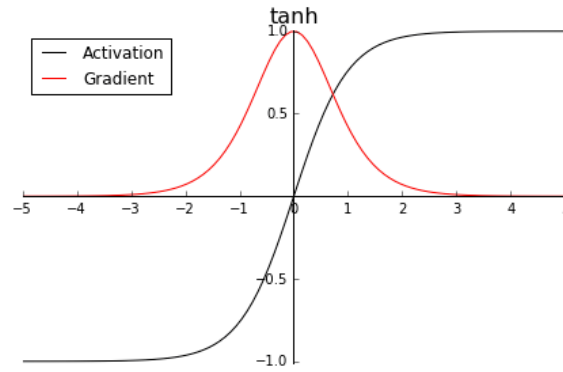
**Figure 3.5 Sigmoid Function and its gradient expression.**

From Figure 3.5, it can be seen clearly that the output range of the logistic function is between 0 and 1. When the input value is near 0, the sigmoid function is approximately a linear function; when the input value is close to both ends, the input is suppressed. The smaller the input, the closer it is to 0; the larger the input, the closer it is to 1. In this case, the gradient of the Logistic function is close to 0, which may lead to the vanishing gradient problem and thus affect the training efficiency of the neural network based on the backpropagation algorithm.

Tanh function is a function that is very similar to the sigmoid function which both have “S” shape, which is shown in Figure 3.6. It can be regarded as an enlarged and translated sigmoid function, and its value range is  $(-1, 1)$ . Its mathematical expression is as follows.

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = 2\sigma(2x) - 1. \quad (3.4)$$

Compared with the sigmoid function, the output range of the Tanh function is -1 to 1. This feature makes the mean of its output value closer to 0 (zero centralization), which helps the data remain stable during the training process and accelerates the convergence speed. However, similar to the sigmoid function, the tanh function also suffers from the vanishing gradient problem when the input values are very high or very low, which may affect the update of the weights during backpropagation.

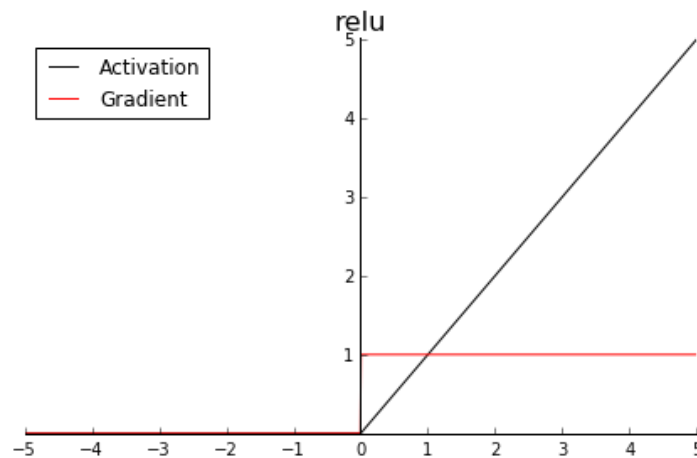


**Figure 3.6 Tanh Function and its gradient expression.**

ReLU (Rectified Linear Unit) is an activation function commonly used in deep neural networks. From Figure 3.7, it can be seen that ReLU is a typical ramp function with mathematical formula

$$ReLU(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} = \max(0, x) \quad (3.5)$$

Using ReLU as the activation function is computationally efficient because only comparison operations are required. In addition, the derivative is 1 if input is larger 0, which alleviates the vanishing gradient problem of the neural network that occurs when using the sigmoid function and the tanh function, and accelerates the convergence speed of gradient descent.



**Figure 3.7 ReLU Function and its gradient expression.**

### 3.4.1.2 Feedforward neural networks

Combining multiple neurons into a network through a certain connection method is a neural network. Up to now, researchers have developed a variety of neural network structures. The three commonly used neural networks include feedforward neural networks, memory networks and graph networks. In this study, the network structures proposed are all feedforward neural networks.

In feedforward neural networks, each neuron belongs to different layers. The neurons in each layer can receive signals from the neurons in the previous layer and generate signals to output to the next layer. The 0<sup>th</sup> layer is called the input layer and is responsible for accepting original data. The last layer is called the output layer and is used to produce the final result. The other intermediate layers are called hidden layers. The hidden layer processes data in a weighted manner and provides nonlinear transformation through the activation function. There is no feedback in the entire network, and the signal propagates in one direction from the input layer to the output layer. Figure 3.8 shows a simple structure diagram of a feedforward neural network.

Let  $a^{(0)} = x$ , the feedforward neural network propagates information by continuously iterating the following formula:

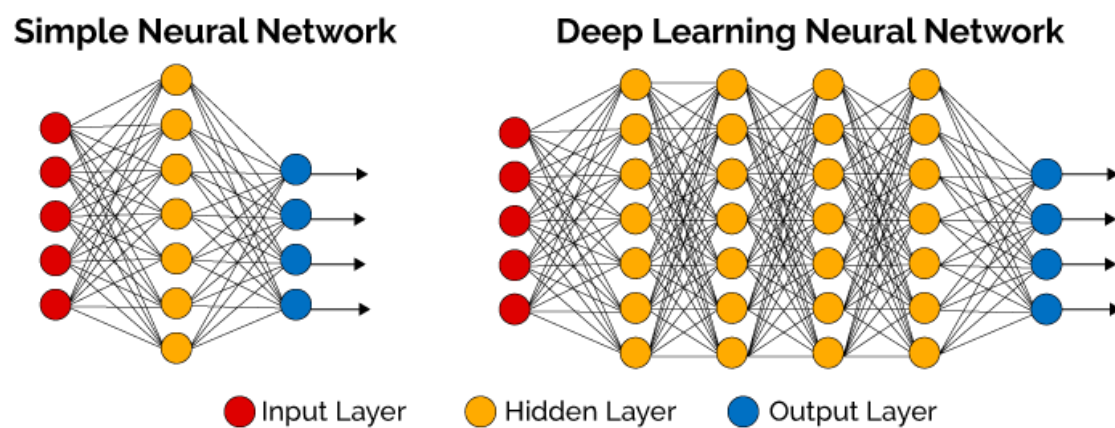
$$z^{(l)} = W^{(l)}f_{l-1}(z^{(l-1)} + b^{(l)}) \quad (3.6)$$

$$a^{(l)} = f_l(W^{(l)}a^{(l-1)} + b^{(l)}) \quad (3.7)$$

where  $z^{(l)}$  represents the input value of the current layer, and  $a^{(l)}$  represents the output value of the current layer. Calculate the input value  $z^{(l)}$  of the  $l^{\text{th}}$  layer neuron based on the output value  $a^{(l-1)}$  of the  $l-1^{\text{th}}$  layer neuron, and then obtain the output value  $a^{(l)}$  of the  $l^{\text{th}}$  layer neuron through an activation function, After multiple iterations,  $a^{(L)}$  is used as the output of the entire network.

Parameter learning is the core of deep feedforward networks, because the purpose of the neural network is to try to fit the required function with a series of suitable parameters. The network adjusts weights and biases to minimise the difference between the predicted and actual outputs. This process usually involves two key steps: forward

propagation and backward propagation. Input data is propagated forward through the network, and the output of each layer becomes the input to the next layer until the final output layer computes the final output. The error between the output and the target is calculated using a loss function (such as mean square error or cross entropy), and then this error is backpropagated through the network and used to calculate the gradient of each parameter. The training of neural networks is essentially an optimisation problem, and the goal is to minimise the loss function. Gradient descent method is the most commonly used optimisation method, which updates the weights in the gradient direction of the loss function. Since gradient descent can be very slow or get stuck in local minima, several variants of gradient descent have been developed, such as stochastic gradient descent (SGD), momentum methods, Adagrad, RMSprop, and Adam [155-158]. The backpropagation algorithm is used to compute the derivatives of the loss function with respect to each parameter. It applies the chain rule to recursively determine the influence of each weight on the final loss.



**Figure 3.8 The structure of one-layer neural network and a 6-layer deep neural network.**

### 3.4.2 Deep Convolutional Neural Network

Deep Convolutional Neural Networks (DCNN) have made significant breakthroughs in computer vision, natural language processing and other fields. Yann LeCun introduced the LeNet-5 model [159], which represents one of the earliest

convolutional neural networks designed for handwritten digit recognition. This architecture consists of multiple convolutional and pooling layers, followed by fully connected layers, aimed at performing classification tasks. One of the key innovations of this model is the combination of convolution layer and pooling layer to reduce the number of parameters and computational cost.

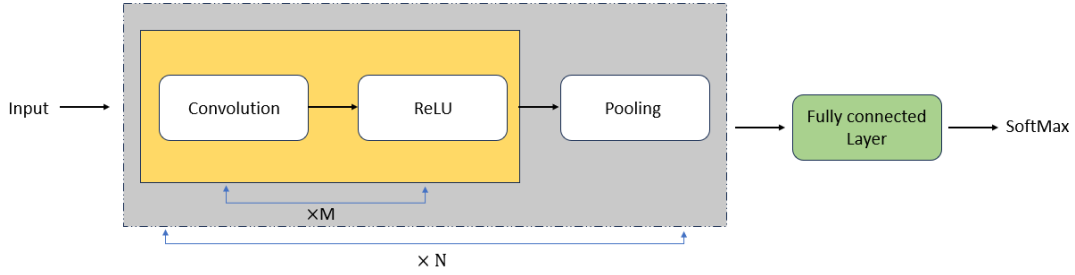
Building upon this foundation, Alex Krizhevsky and colleagues developed AlexNet [160], a deeper CNN that gained widespread attention by winning the 2012 ImageNet Large Scale Visual Recognition Challenge. AlexNet incorporates multiple convolutional and pooling layers with the ReLU activation function, and leverages GPU for efficient training. This network represented a major advancement in the field by demonstrating the scalability of deep learning models for large-scale visual recognition tasks. VGGNet [161] was proposed by Karen Simonyan and Andrew Zisserman. It is characterised by a unified architecture, using small-sized convolution kernels and deep network structures. Further advancing CNN architecture, Google-Net [162] introduced the Inception module, a novel approach allowing simultaneous application of multiple convolutional kernel sizes. This modular design significantly improves the network's capability for feature extraction, while also optimizing parameter efficiency. The Inception module has since been a cornerstone in designing large, wide and computationally efficient neural networks. Finally, Kaiming He [163] and his team proposed ResNet, which introduced the revolutionary concept of residual connections. These connections facilitate information bypassing across layers, effectively addressing the vanishing gradient problem prevalent in deep networks. With residual connections, networks comprising hundreds or even thousands of layers became feasible to train, significantly advancing the field of deep learning.

Recent advancements such as MobileNet [156] and DenseNet [157] have further enhanced CNN architectures by focusing on parameter efficiency and feature reuse, respectively. These improvements address the limitations of earlier models like AlexNet and VGGNet, especially in terms of computational cost and gradient propagation in



deep networks.

As shown in Figure 3.9, a typical convolutional network is composed of a convolutional layer, a pooling layer, and a fully connected layer. A convolutional block consists of consecutive  $M$  convolutional layers and 1 pooling layer. A convolutional network can be stacked with  $N$  consecutive convolution blocks, followed by 1 or 2 fully connected layers, and finally the output of the last layer of the neural network is converted into a probability distribution through the SoftMax function.



**Figure 3.9 The structure of a typical convolution neural network.**

### 3.4.2.1 Convolution layer

Convolution is a mathematical operation widely used in signal processing, image processing, and deep learning. One-dimensional convolution is often used in signal processing to calculate the delay accumulation of the signal. Its mathematical expression is as follows.

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau \quad (3.8)$$

For two-dimensional functions (such as images), given an image  $I \in \mathbb{R}^{M \times N}$  and a convolution kernel  $K \in \mathbb{R}^{U \times V}$ , the definition of 2-D convolution is:

$$(I * K)(i, j) = \sum_{u=1}^U \sum_{v=1}^V I(u, v)K(i - u, j - v) \quad (3.9)$$

Where  $i$  and  $j$  are the position coordinates on the image. The mean filter commonly used in image processing is a 2-D convolution. In addition, convolution is also widely used as a common feature extractor.

In the basic convolution operation, the stride and zero-padding can be added to

increase the diversity of the convolution. The stride refers to the interval at which the convolution kernel slides, and the zero-padding refers to the number of zeros padded at both ends of the input vector. Let the number of inputs of the convolution layer be  $M$ , the convolution kernel size be  $K$ , the step size be  $S$ , and  $P$  zeros are padded at both ends of the input, then the number of neurons  $N$  in the convolution layer is

$$N = \frac{(M - K + 2P)}{S} + 1 \quad (3.10)$$

When the amount of zero padding varies, convolution can be categorized as narrow, wide, or equal-width convolution, corresponding to valid padding, full padding, and same padding, respectively. In the current literature, convolution generally defaults to equal-width convolution.

Convolutional neural networks generally consist of convolutional layers, pooling layers and fully connected layers. The function of the convolutional layer is to extract features of a local area, and different convolution kernels are equivalent to different feature extractors. The convolution kernel slides horizontally and vertically on the input feature map according to the specified step size (stride). At each sliding position (window position), the convolution kernel performs an element-level product with the input feature map area it covers, and then all products are summed to obtain an output value, and this step is repeated until the convolution kernel covers all positions of the input feature map. Each output value forms a pixel of the output feature map.

The choice of kernel size, stride, and padding in convolutional layers significantly influences the network's ability to capture fine-grained details versus broader contextual information. Smaller kernels, such as the 3x3 kernels employed in VGGNet, focus on extracting local features, while larger kernels provide more global contextual understanding.

Since the image is a two-dimensional structure, in order to fully utilise the local information of the image, the neurons are usually constructed as layers with a three-dimensional structure, whose size is Height  $H \times$  Width  $W \times$  Depth  $D$ . For images, if it is a grayscale image, the depth of the input layer  $D = 1$ ; if it is a RGB image, the depth

of the input layer  $D = 3$ . The features extracted from an image (or other features) after convolution extraction are called feature maps. Each feature map can be used as a type of extracted image features. In order to improve the representation ability of the convolutional network, multiple different feature maps can be used in each layer to more fully represent the characteristics of the image.

### 3.4.2.2 Pooling Layer

Let the input feature map of the pooling layer be  $X \in \mathbb{R}^{H \times W \times D}$ , for each feature map  $X^d \in \mathbb{R}^{H \times W}$  divide it into many areas  $R_{m,n}^d$ . Pooling refers to down sampling each area to obtain a value as a representative of this area. There are usually two types of pooling methods, max pooling and average pooling.

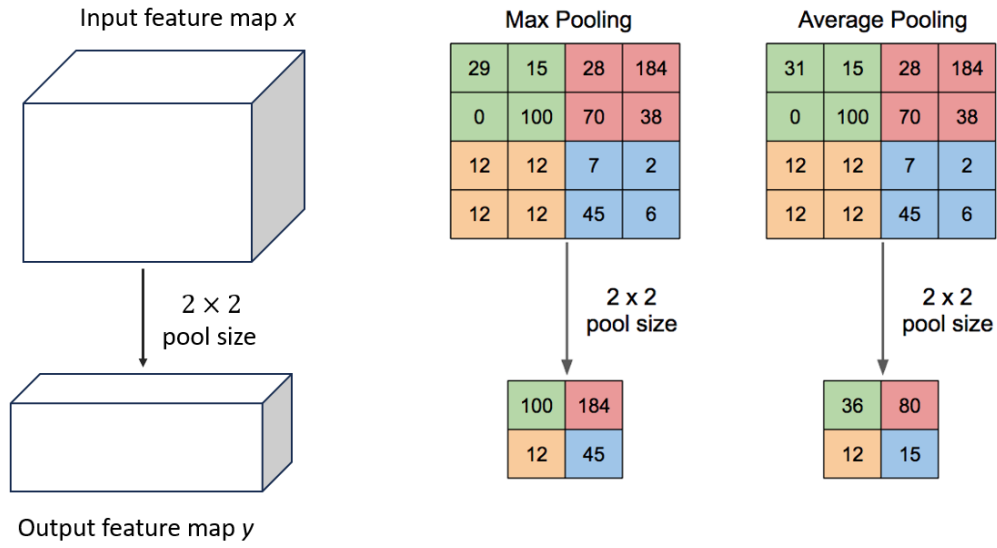
For a region  $R_{m,n}^d$ , maximum pooling selects the maximum value of all neurons in this region as the representation of this region.

$$y_{m,n}^d = \max(x_i), i \in R_{m,n}^d \quad (3.11)$$

Average pooling calculates the average value in a neighbourhood area, the average value in each block represents this part of information or feature.

$$y_{m,n}^d = \frac{1}{|R_{m,n}^d|} \sum_{i \in R_{m,n}^d} x_i \quad (3.12)$$

The pooling layer can effectively reduce the number of neurons, thereby reducing the feature dimension and avoiding overfitting. It can also make the network remain invariant to some small local morphological changes and have a larger receptive field. Figure 3.10 shows a schematic diagram of the maximum pooling and average pooling operations. However, an excessively large sampling area will drastically reduce the number of neurons and cause excessive information loss. How to set reasonable pooling parameters is an important process in adjusting deep networks.



**Figure 3.10 Example of Max Pooling and Average Pooling[164].**

CNNs have also been widely adopted beyond traditional image classification tasks, including applications in medical image analysis, autonomous driving, and natural language processing with the integration of attention mechanisms. As deep learning models continue to evolve, the integration of CNNs with other architectures promises further advancements in these fields.

### 3.4.3 Fully Convolutional Network

In recent years, Deep Learning has become the main research direction of computer vision problems. Especially after the amazing achievements of Convolutional Neural Network in image classification, a growing number of researchers choose to use deep neural networks to solve a series of image processing tasks, such as image recognition, object detection, and segmentation. With the rapid development of deep learning, deep neural networks have made amazing achievements in the field of computer vision. The advancements in deep learning have significantly impacted the field of computer vision and have also introduced novel approaches to medical imaging. The Fully Convolutional Neural Network (FCN) developed by Long et al [165]. has achieved excellent results in natural image segmentation tasks and provided ideas for future end-

to-end segmentation networks later. Avi et al. [166] further validated the effectiveness of FCN by showing that they can achieve high segmentation accuracy in clinical applications.

In the early days, neural networks were primarily employed for classification tasks, rather than segmentation. However, Ciresan et al. [167] developed a neural network and this network became the category champion for ISBI2012. This network uses a sliding window to predict the class label for each pixel, while providing the area around the pixel as input. Despite its success, this approach exhibited two major limitations. First, it was computationally inefficient due to the extensive overlap in the input, leading to redundant calculations. The same features were processed multiple times, which not only wasted computational resources but also increased the risk of overfitting. Second, there was a trade-off between localization accuracy and contextual information. Larger patches required more pooling layers to capture high-level features, which in turn reduced the resolution of the feature map and caused loss of detail. Conversely, smaller patches focused on local information but lacked sufficient context to accurately classify the pixel.

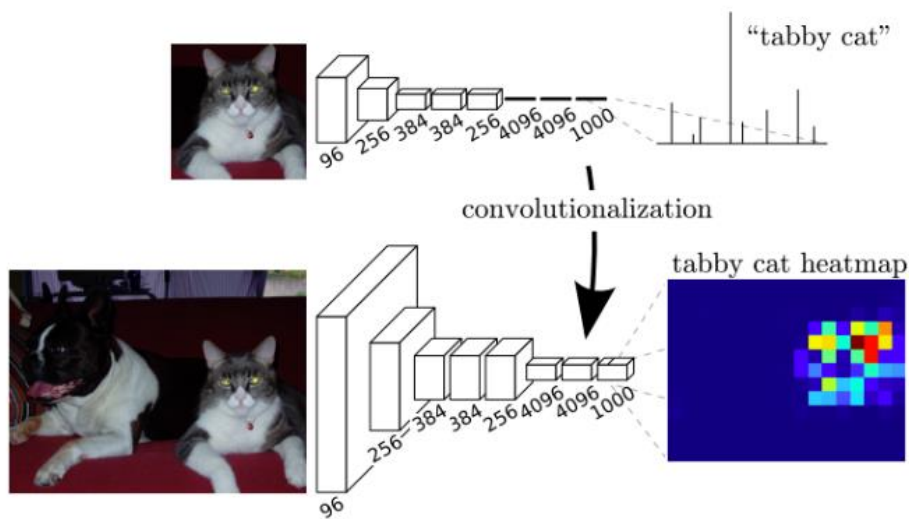
The power of CNN is that its multi-layer structure can automatically learn features and output specific results according to these features in image classification. Shallower convolutional layers capture local features with small receptive fields, while deeper layers have larger receptive fields, allowing them to extract more abstract features. And these features are strongly robust on the size, position and direction of the object target which helps to provide reliable predictions in all situations. These learned features are robust to variations in object size, position, and orientation, thereby enhancing the reliability of predictions. Although these features are highly effective for classification tasks, they are often insufficient for precisely delineating the fine contours of objects, thereby making pixel-level segmentation a challenging task.

Traditional CNN-based segmentation algorithms typically classify individual pixels based on their surrounding pixels by treating each pixel block as an input to the

network. This approach leads to several significant challenges. First, the memory requirements increase dramatically. For example, if a  $5 \times 5$  pixel block is used to represent a single pixel, the required memory space is 25 times than the initial picture. Second, computational efficiency is reduced because adjacent pixel blocks often contain overlapping elements, leading to redundant convolution operations without improving performance. Finally, small pixel blocks struggle to capture abstract features and fail to integrate information across regions, which negatively impacts the accuracy of segmentation results.

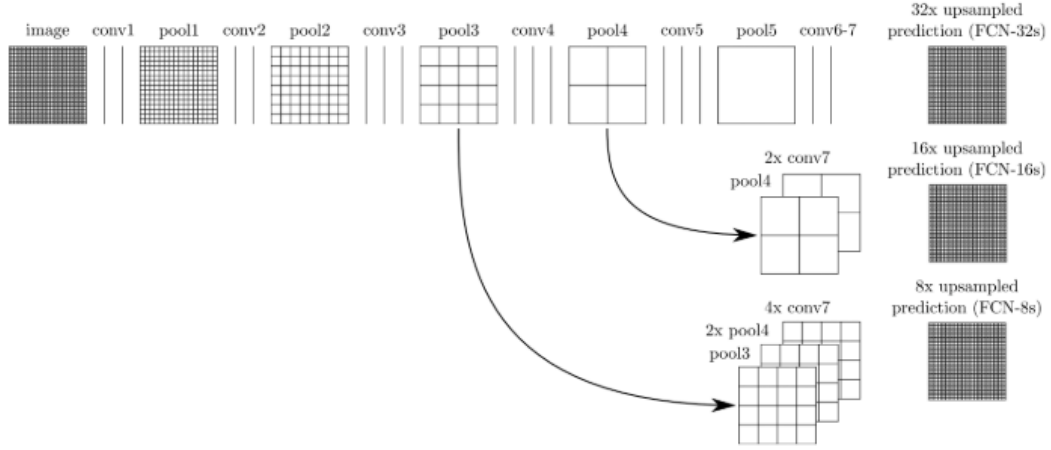
To address this series of challenges, Jonathan Long et al. [165] in Berkeley proposed Fully Convolutional Networks (FCN) for image segmentation. This network is used to recover the classification according to the abstract features for each pixel. It also means to achieve the classification at the pixel level instead of the previous classification at the image level. FCNs achieve this by replacing the fully connected layers in traditional CNNs with convolutional layers, thereby allowing the network to perform dense predictions for each pixel.

As shown in the following Figure 3.11, in the traditional CNN architecture, the first five layers consist of convolutional operations with activation functions, while the 6th and 7th layers are fully connected, each with 4096 units, followed by the 8th layer which outputs a 1000-dimensional vector representing the probabilities for 1000 categories. In contrast, the FCN replaces the fully connected layers with convolutional layers, generating dense feature maps of sizes  $(7 \times 7 \times 4096)$ ,  $(7 \times 7 \times 4096)$ , and  $(7 \times 7 \times 1000)$ , respectively. These feature maps are obtained by applying  $1 \times 1$  convolutions to the outputs of the preceding layers, followed by a SoftMax function to assign a label to each pixel. Due to the fully convolutional nature of the hidden layers, this architecture is referred to as an FCN.



**Figure 3.11 The simple structure of a Fully Convolutional Network [159].**

The process of convolution and pooling progressively reduces the spatial resolution of the image, resulting in down sampled feature maps. FCN gradually restores the spatial resolution of the feature map by introducing up-sampling operations, thereby achieving full-size output for the input image. This up-sampling strategy is able to restore the feature map to the same resolution as the input image to generate a pixel-level segmentation map. After five stages of convolution and pooling, the feature map is down sampled to  $1/32$  of its original size, such that the final layer's output has dimensions  $(7 \times 7 \times 1000)$ . To recover the original resolution, the final layer is up sampled by a factor of 32. However, the output of this 32x up-sampling process lacks sufficient detail, prompting the authors to refine the segmentation by performing additional deconvolutions on the outputs of the 4<sup>th</sup> and 3<sup>rd</sup> layers, with 16x and 8x up-sampling, respectively. This multi-level up-sampling approach results in finer segmentation, although challenges related to blurring and insufficient separation of object boundaries remain. The following Figure 3.12 shows the process of various up sampling:



**Figure 3.12 Up-sampling process in FCN, including 8 times and 16 times up-sampling prediction [159].**

Compared with traditional CNN-based image segmentation methods, FCN has significant advantages in training efficiency. FCN reduces memory consumption and computational redundancy by removing the process of repeatedly processing pixel blocks containing overlapping pixels. However, FCN also exhibits certain limitations. First, the segmentation results are usually not precise enough. Although 8x up-sampling produces better results than 32x up-sampling, the output is still blurry, and the network has difficulty capturing fine details. Second, FCN classifies each pixel independently and does not fully consider the spatial dependencies between neighbouring pixels. The lack of spatial regularization commonly adopted in pixel-based segmentation methods leads to a lack of spatial consistency in the final segmentation output.

To address these limitations in practical medical applications such as liver segmentation from CT images, various improvements to the standard FCN framework have been proposed. Christ et al. [168] proposed a cascaded fully convolutional neural network (CFCN) method for automatic segmentation of the liver and its lesions. The method first uses FCN to perform preliminary segmentation of the liver in CT images and then uses a second FCN to further segment the lesions in the liver region segmented by the first FCN. The advantage of this cascade structure is that by gradually segmenting the liver and lesions, the false positive problem in lesion segmentation is



effectively reduced.

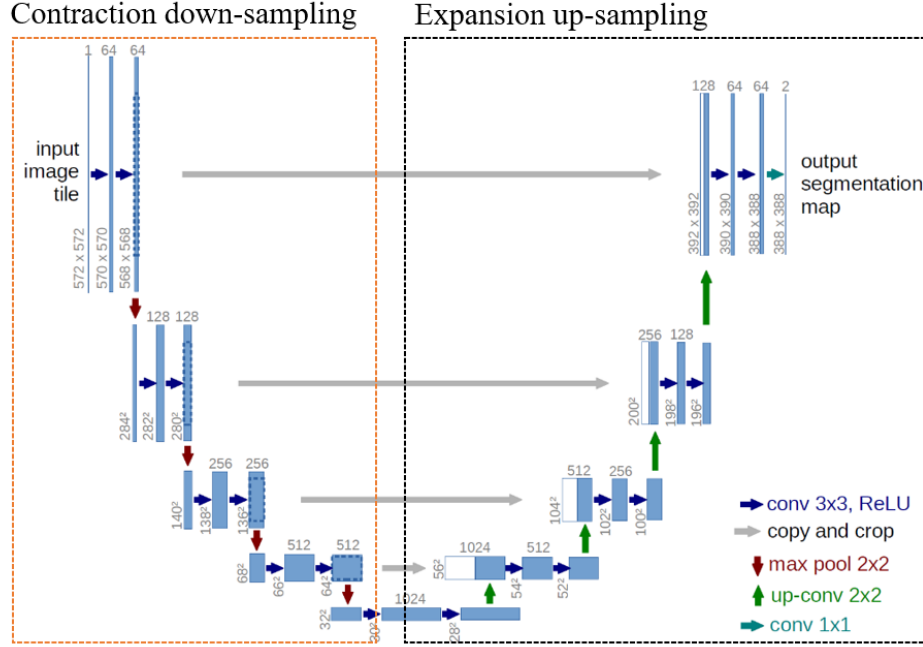
In addition, Sun et al. [169] proposed an automatic liver tumour segmentation method based on FCN, which demonstrated the effectiveness in this task. The study used multi-phase enhanced CT images, allowing FCN to perform well in capturing image features of different phases. The incorporation of multiphase contrast information improves the detection and segmentation performance of liver tumours, making it more widely applicable and effective in the field of medical image segmentation.

#### **3.4.4 U-Net**

U-Net is a convolutional neural network architecture for image segmentation. It was first proposed in the context of medical cell image segmentation by Olaf et al [105] but has been widely used in various image segmentation tasks. The main feature of U-Net is its U-shaped structure, which includes a contraction down-sampling path and an expansion up-sampling path, and there are many skip connections between these two paths.

U-Net consists of two parts including a downward encoder and an upward decoder, which is shown as Figure 3.13. The left half implements the function of the encoder. The encoder uses many convolution operations and pooling operations to gradually reduce the spatial dimension of the image and increase the number of feature channels to obtain feature maps of different levels. The encoder consists of 4 blocks, each block consists of repeated two  $3 \times 3$  valid convolutional layers, followed by a ReLU activation function and a  $2 \times 2$  maximum pooling operation to reduce the feature map the size. After each down-sampling, the number of feature maps is doubled to obtain more deep features, so there is a change in the size of the feature map shown in the Figure 3.13. Finally, a feature map with a size of  $32 \times 32$  is obtained. The right half implements the function of the decoder, which performs the opposite operation to encoder, gradually recovering the spatial dimension of the image through up-sampling. The decoder first performs a  $2 \times 2$  up-sampling operation and then merges the result with a skip connection

in the encoder part. At the end of the decoder, there is a  $1 \times 1$  convolution operation to transform the multi-channel feature maps into the desired output categories.



**Figure 3.13** The architecture of a classical U-Net.

In image segmentation tasks, it is usually necessary to combine the local details of the image (low-level features) and a larger range of contextual information (high-level features) to obtain more accurate results. The skip connections in U-Net connect the feature maps of each level in the encoder to the corresponding level in the decoder, which enables the network to obtain fine-grained features of the encoding stage in the decoding stage. Another effect of skip connections is to restore spatial resolution. In the encoding stage, the spatial resolution of the image (feature maps) gradually decreases due to successive pooling operations. In the decoding stage, deep features alone might not be sufficient to recover all the details of the original resolution despite the up-sampling operation. Skip connections provide these lost details and help restore the original spatial structure of the image.

To ensure that the predicted segmentation maps are closely aligned with the ground truth, it is crucial to use an appropriate loss function during training. The loss function

helps guide the network to learn the correct pixel-by-pixel classification by measuring the difference between the predicted output and the true segmentation label.

In binary classification tasks such as foreground-background segmentation, the cross-entropy (CE) loss is often used. It effectively guides the model optimisation process by comparing the predicted probability of each pixel with the true label (0 or 1) and applying a larger penalty for incorrect predictions. This helps the network gradually improve its accuracy in distinguishing different regions in the image, especially when the difference between regions is small.

$$CE\ Loss = -[g_i \log(p_i) + (1 - g_i) \log(1 - p_i)] \quad (3.13)$$

Where  $p_i$  is the predicted probability for pixel  $i$  and  $g_i$  is the true label (0 or 1) for pixel  $i$ .

While the cross-entropy loss performs well in binary pixel classification tasks, it can be challenging to handle data imbalance problems, especially when the background occupies a large area in the image. For example, in the liver tumour segmentation task, the tumour region usually occupies only a small part of the image, making it difficult for the model to accurately segment the tumour tissue. To address this issue, the Dice loss is often combined with the cross-entropy loss. The Dice loss focuses on the similarity of pixel sets by directly maximizing the overlap between the predicted segmentation and the true segmentation, so that it focuses on the pixel sets of the target object, thereby improving segmentation accuracy.

$$Dice\ Loss = 1 - \frac{2 \times |P \cap G|}{|P| + |G|} = 1 - \frac{2 \times \sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \quad (3.14)$$

Where  $p_i$  is the predicted probability for pixel  $i$ ,  $g_i$  is the true label (0 or 1) for pixel  $i$ , and  $N$  is the number of all pixels.

The application of U-Net-based neural networks in medical image segmentation has made significant progress. Han [170] proposed a 2.5D deep convolutional neural network that integrates U-Net and ResNet architectures for liver and lesion segmentation. This model takes multiple adjacent CT slices as input and generates a segmentation map for the central slice, effectively addressing the significant variability

in liver anatomy, including variations in shape, size, and intensity. The innovation of this network lies in combining the long-range feature connections of U-Net with the short-range residual connections of ResNet, enabling it to handle issues like low contrast and noise artefacts in CT images with high efficiency

In another study, Li et al. [171] introduced a modified U-Net variant called Bottleneck Supervised (BS) U-Net, designed for pixel-wise segmentation of the liver and liver tumours. The base U-Net model incorporates dense modules, inception modules, and dilated convolution in the encoding path, which significantly enhances segmentation accuracy. The BS U-Net first uses an encoding U-Net to learn anatomical features from the label maps and then utilises this information to guide the segmentation network during training. This approach ensures that the segmentation results maintain the anatomical structure of the liver while effectively reducing false positives and false negatives. Experiments demonstrated that this model not only achieved excellent overall segmentation performance but also addressed issues such as shape distortion and boundary segmentation, which are common challenges in liver segmentation.

### **3.4.5 Application of deep learning in liver and tumour segmentation**

In terms of automatic segmentation of liver and liver tumours, in addition to the above-mentioned fully convolutional network and U-net, derivative models of U-Net are also widely used. In recent years, many studies have improved U-Net to improve its segmentation accuracy. For example, Residual U-Net [172, 173] alleviates the vanishing gradient problem in deep networks by adding residual connections, thereby effectively improving the stability and accuracy of the network. Attention U-Net [174] introduces an attention mechanism to significantly improve the segmentation effect by focusing on the liver and tumour areas and ignoring irrelevant background noise. Jiang et al. [175] also proposed the AHCNet which combines soft and hard mechanism and long-short skip connection to achieve liver and tumour CT image segmentation. They also proposed a cascade network structure, including a liver localization network, a liver segmentation network and a tumour segmentation network. The collaborative

work of the three networks significantly improved the accuracy of tumour segmentation and effectively dealt with the class imbalance problem.

In addition to simply adding some mechanisms, there are also network models such as U-Net++ [176] that enhance their performance by introducing more complex hierarchical structures. The basic principle of U-Net++ is to introduce multi-level dense skip connections between the encoder and decoder, so that the high-resolution feature maps are gradually enhanced before being fused into the semantically rich decoder features. This progressive feature fusion method can effectively reduce the semantic difference between the encoder and decoder, allowing the model to better capture the fine-grained details and boundary information of the foreground object, thereby improving the accuracy of segmentation. U-Net sends high-resolution feature maps directly from the encoder to the decoder network through skip connections, and semantically different feature maps are combined, resulting in a semantic gap problem between the encoder and decoder feature maps. U-Net++ uses stacked skip connections and thick skip connections to achieve progressive feature fusion, allowing the encoder to gradually increase the integration of low-level and high-resolution information during the decoding process, so that the decoder can better adapt to the deep semantic features extracted by the encoder while gradually restoring spatial information. Moreover, dense blocks and convolutional layers are added between the encoder and decoder to improve segmentation accuracy. Due to the stacked skip paths, U-Net++ builds high-resolution feature maps at multiple semantic levels, that is, at different depths, and uses four semantic levels to quantify the loss. These improvements enable U-Net++ to achieve higher accuracy in the segmentation tasks of liver and liver tumours. Li et al. [177] used the U-Net++ architecture with an attention-aware mechanism to segment the liver from CT images and achieved excellent results with 98.15% dice score.

Huang et al. [178] proposed U-Net 3+, which considers full-size skip connections and deep supervision. Deep supervision learns hierarchical representations from

aggregated feature maps of multiple sizes by adding support loss functions at multiple levels. This means that while the final output is supervised, losses are calculated at different stages of the decoder to guide feature learning at each stage, ensuring that the features of each layer contribute to the final segmentation task. They applied the U-Net 3+ architecture to liver segmentation and achieve superior performance with 96.75% dice score compared to other U-Net architectures.

In addition to U-Net based architectures and their variants, Mask R-CNN has demonstrated remarkable performance in various medical imaging tasks. Originally proposed by He et al. [179] for pixel-level instance segmentation, Mask R-CNN extends the Faster R-CNN framework, which is specifically designed for object detection [180]. The extension involves the addition of a mask prediction branch, enabling pixel-wise segmentation for individual objects. Moreover, a key innovation introduced by Mask R-CNN is the ROI Align operation, which replaces the ROI Pooling mechanism used in Faster R-CNN [180]. ROI Align decreases the quantization errors associated with ROI Pooling by employing bilinear interpolation, thereby achieving spatial alignment and enhancing segmentation accuracy. This architectural refinement makes Mask R-CNN particularly well-suited for high-precision tasks in medical imaging, such as tumour segmentation and organ delineation.

Haq et al. [181] proposed a liver segmentation method based on Mask R-CNN. First, the CT volume is pre-processed and normalized into standardized CT image slices to obtain the basic outline of the liver. These normalized slices were then input into the ResNet-101 network for feature extraction. Finally, Mask R-CNN is used to segment the tumour in the liver. This method has demonstrated the effectiveness of Mask R-CNN for liver CT segmentation on public datasets. Chen et al. [182] proposed an enhanced Mask R-CNN algorithm for liver segmentation in CT images, which contained the following innovations. First, the k-nearest neighbour (K-NN) algorithm is introduced in the training stage to cluster the target liver pixels and obtain a suitable aspect ratio for adjusting the size and shape of the anchor points, thereby improving the

accuracy of liver positioning. Second, a fully convolutional network (FCN) is used to perform pixel-level segmentation of foreground objects and refine the detection of liver areas.

H-DenseUNet, proposed by Li et al. [183], is a hybrid architecture that integrates 2D DenseUNet for intra-slice feature extraction and 3D DenseUNet for volumetric context aggregation. This design leverages hybrid feature fusion to combine 2D and 3D features, effectively avoids the limitation of 2D networks ignoring volume context information and the high computational cost of 3D networks. The network incorporates dense connectivity for efficient feature reuse and employs a cascaded segmentation approach to sequentially localise the liver and refine tumour segmentation within the ROI.

Similar designs come from the X-Net proposed by Chi et al. [184], which takes full advantage of the ability of intra-slice and inter-slice feature extraction. Global features are extracted through the backbone Dense U-Net, and they are combined with the liver segmentation branch and tumour segmentation branch to optimise the feature extraction and segmentation performance of local areas respectively. The network significantly reduces the computational cost of 3D feature extraction by stacking 2D features instead of traditional 3D convolution. On this basis, the contextual information between slices is efficiently integrated with the detailed information in the two-dimensional slices, thereby improving the segmentation performance.

Zhang et al. [185] proposed a liver and tumour segmentation framework with a coarse-to-fine approach. Initially, a 3D U-Net is employed to segment the liver, followed by a Multi-Scale Candidate Generation (MCG) step to identify tumour candidates, leveraging superpixel-based techniques to extract all potential tumour regions. Subsequently, a 3D Fractal Residual Network is introduced to refine the tumour region delineation, serving as the coarse segmentation output. Finally, an Active Contour Model is applied to refine the tumour segmentation further, achieving precise boundary delineation.

Generative Adversarial Networks (GANs) were proposed by Ian et al. in 2014 [151], have become a powerful tool in various computer vision tasks, particularly in image synthesis. In medical imaging, GANs have been successfully applied to tasks such as image enhancement, modality translation, and image segmentation [186, 187]. Their ability to learn complex data distributions makes them particularly useful in settings with limited labelled data or domain shifts. GANs consist of two neural networks, one for the generator and the other for the discriminator. The generator is used to create new data, while the discriminator evaluates the generated data based on its similarity to the training data. The GAN architecture has been shown to achieve good performance in liver and liver tumour segmentation. Xie et al. [188] proposed a radiomics-guided GAN for liver tumour segmentation without contrast agents. This method achieves segmentation through the mapping relationship between contrast images and non-contrast images and uses the radiomics features in the discriminator as guidance information to help extract features in non-contrast images. Experiments show that this method achieves a segmentation accuracy of 95.85% and a Dice coefficient of 92.17% on the data of 200 subjects. Enokiya et al. [189] combined U-Net and Wasserstein GAN for automatic liver segmentation. Wasserstein GAN is mainly used to improve the training of U-Net under small data set conditions. Experimental results show that this method can obtain a dice score of 90% with only 10 training images.

Zhang et al. [190] proposed a tumour 3D conditional generation confrontation segmentation network (T3scGAN) based on conditional generation confrontation network. T3scGAN only fuses feature of two scales, while ensuring the accuracy of 3D segmentation, it minimises network parameters and computational complexity to avoid losing details of small tumour areas due to excessive down-sampling. At the same time, they adopted a coarse-to-fine 3D segmentation framework to achieve fine segmentation from liver to tumour.

To overcome challenges in medical image segmentation, such as limited annotated



data and differences between different patients and imaging protocols, researchers have adopted a variety of techniques, including data augmentation, multi-scale feature extraction, and transfer learning. Among these methods, transfer learning has proven to be particularly effective when insufficient training data is available. Transfer learning involves using models that have been pre-trained on large-scale natural image datasets. These pre-trained models can capture general features that are applicable to different visual tasks. In the field of medical imaging, transfer learning enables models to benefit from previously learned representations and reduces the need for large-annotated datasets. This is particularly important because collecting and labelling high-quality medical data requires expert efforts and is often resource intensive. By applying transfer learning, researchers can improve model performance and training efficiency in medical image segmentation tasks.

Heker et al. [191] used SE-RESNet to achieve the goal of liver segmentation. Before training, they used pre-trained ImageNet for feature extraction. The authors found that using transfer learning significantly improved weight initialization and faster convergence speed. In order to solve with the challenge of insufficient 3D medical image data, Chen et al. [192] have aggregated data sets from multiple medical challenges, constructed the 3DSeg-8 data set, and designed the Med3D pre-training model to accelerate the training of 3D medical tasks and improve accuracy. Experiments showed that the Med3D model significantly improved the training convergence speed and accuracy in multiple 3D medical tasks. A dice score of 94.6% was achieved for liver segmentation based on the LiTS challenge dataset.

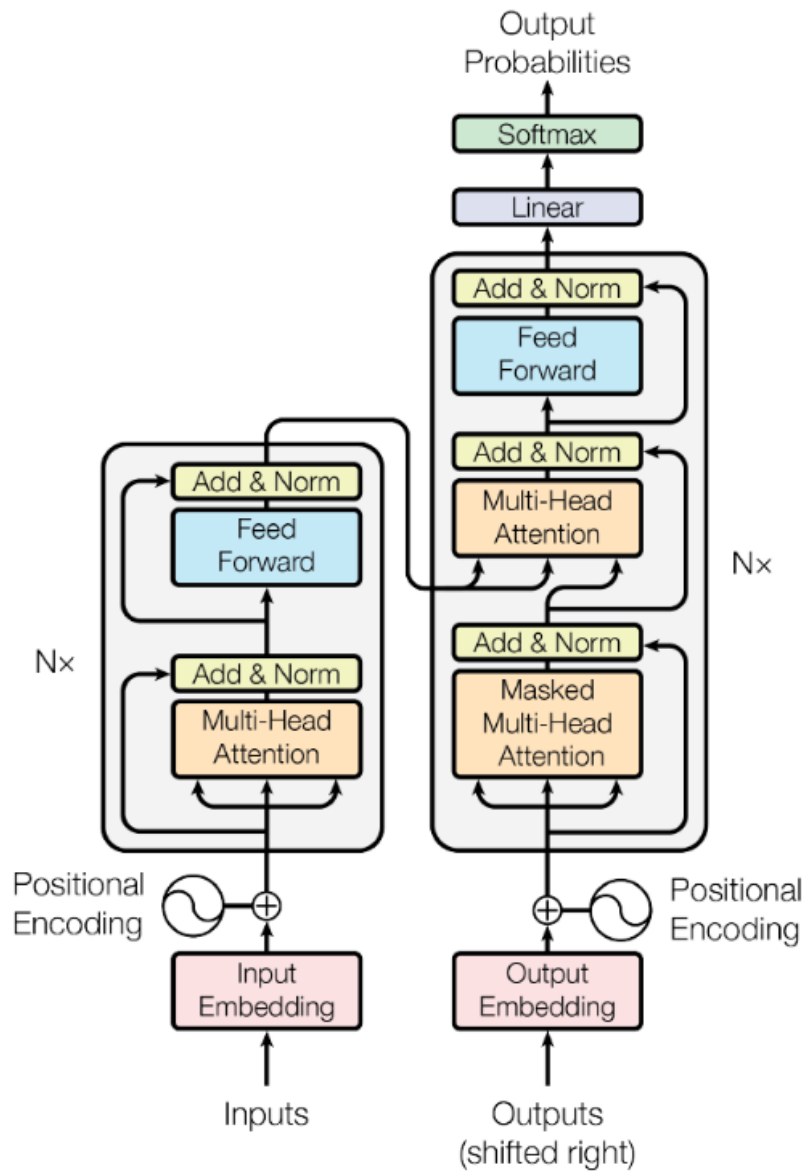
In recent years, the Transformer architecture, originally developed for natural language processing, has been successfully applied to various computer vision tasks, including image segmentation. Various models based on convolutional neural networks have dominated this field, especially architectures such as U-Net, DeepLab, and Mask-RCNN. However, the emergence of transformer models has broken this status quo. The development of transformer-based models shows that the self-attention mechanism

can capture long-range dependencies and global contextual connections more effectively than CNN, which helps to improve the performance of segmentation models. The basic principle of Transformer relies on the self-attention mechanism, which can effectively process the global information of sequence data through multi-head attention, position encoding, and encoder-decoder structure. Figure 3.14 shows the encoder-decoder architecture of the Transformer model.

Transformer-based image segmentation architectures include Vision Transformer (ViT), Swin Transformer, Segmentation Transformer (SETR), Detection Transformer (DETR), Mask Former, etc. ViT is one of the early models that applied Transformer to visual tasks, mainly for image classification and segmentation. ViT divides the input image into fixed-size, non-overlapping patches, and subsequently transforms each patch into a feature representation known as a patch embedding. After undergoing a linear transformation, these patch embeddings serve as input tokens for the Transformer encoder. This patch-based segmentation approach enables the model to consider the entire image and capture global contextual information. To retain the spatial position of each patch, ViT incorporates positional embeddings, allowing the model to remain aware of each patch's original arrangement within the image.

Swin Transformer is an improved version of ViT, especially optimised for the needs of vision tasks. It introduces a hierarchical structure with shifted windows, where self-attention is calculated within a local window instead of the entire image. This design reduces the computational cost and enables the model to scale to high-resolution inputs. The hierarchical design also allows Swin Transformer to capture multi-scale features by reducing the resolution of feature maps layer by layer. These properties make Swin Transformer well suited for dense prediction tasks, including semantic and instance segmentation, as it balances local and global feature extraction. Coa et al. [193] introduced Swin-U-Net for medical image segmentation. Their architecture is based on Swin Transformer, where the encoder adopts a shifted window method and has patch embedding layers in the decoder. Both encoder and decoder architectures are designed

hierarchically to enhance segmentation accuracy and robustness.



**Figure 3.14** The encoder-decoder architecture of the Transformer model [147]. The core structure of the Transformer model includes the encoder module on the left and the decoder module on the right. Each module consists of multiple layers of stacked substructures: multi-head attention mechanism, feedforward network, residual connection and layer normalization.

Transformer U-Net (TransUNet) is a hybrid architecture that combines the strengths of Transformer models and CNNs for medical image segmentation tasks. It extends the traditional U-Net structure by incorporating Transformer blocks into the encoder to capture long-range dependencies and global contextual information more effectively. The encoder of TransUNet use a ViT to process flattened image patches,

allowing the network to maintain a global perception of the input data. This is particularly effective in medical imaging, where lesions and organs often exhibit varying scales, shapes, and locations. Meanwhile, the decoder follows the U-Net design, which employs convolutional layers to reconstruct high-resolution segmentation maps, effectively combining the global features extracted by the Transformer with local details learned through convolution.

Li et al. [194] proposed a U-Net-based hybrid variable structure, RDCTrans U-Net for liver tumour segmentation in computed tomography (CT) examinations. They design a backbone network dominated by ResNeXt50 and supplement by dilated convolution to increase the network depth, expand the perceptual field, and improve the efficiency of feature extraction without increasing the parameters. At the same time, a Transformer is introduced in down-sampling to increase the network's overall perception and global understanding of the image and to improve the accuracy of liver tumour segmentation.

Table 3.4 summarises the various liver and tumour segmentation methods proposed over the years. It includes the datasets used, the liver dice score and tumour dice score obtained by each method.

**Table 3.4 Comparative Analysis of Liver and Tumour Segmentation Methods.**

Ref	Year	Proposed Method	Dataset	Liver Dice Score	Tumour Dice Score
[164]	2020	Attention U-Net	LiTS 3DIRCADb	0.961 0.977	0.59 0.830
[165]	2019	AHC-Net	LiTS	0.953	0.734
[167]	2023	Eres-UNet++	LiTS	0.956	0.913
[168]	2020	Unet3+	LiTS	0.960	/
[171]	2021	Mask-R-CNN	LiTS	0.95	/
[172]	2021	Mask-R-CNN	Codalab competition	0.943	/
[173]	2018	H-DenseUNet	LiTS 3DIRCADb	0.965 0.982	0.824 0.937
[174]	2021	X-net	LiTS	0.971	0.843
[175]	2019	3D MCG-FRN	3DIRCADb	/	0.764
[177]	2019	Radiomics-guided GAN	Clinical CT data sets	/	0.92
[178]	2018	U-Net with GANs	Clinical CT data sets	0.94	/
[179]	2021	T3scGAN	LiTS	0.961	0.796
[180]	2020	Transfer learning	LiTS	/	0.71
[181]	2019	Med3d	LiTS	0.946	/
[182]	2022	Swin-Unet	ACDC	0.943	/
[183]	2022	RDCtrans U-Net	LiTS	0.934	0.898

### 3.5 Conclusion

In Chapter 3, various image segmentation technologies based on CT data are introduced, including semi-automatic segmentation technology and fully automatic segmentation methods. This chapter shows the evolution from simple thresholding and graph cut techniques to more complex statistical model algorithms and learning segmentation methods, especially the application of deep learning in medical image

segmentation, marking a major advance in this field. With the improvement of computing power and the continuous optimisation of algorithms, image segmentation technology based on deep learning is expected to play a greater role in improving the accuracy and efficiency of medical image processing.

In the section on semi-automatic segmentation techniques, the principles of threshold setting, and graph cut techniques and their applications in medical image processing are discussed. Although these techniques require a certain degree of manual intervention to select parameters or markers, they are still effective when processing specific types of images. Subsequently, among the fully automatic segmentation methods, methods based on statistical models are first introduced. This type of method automatically performs segmentation by establishing a statistical model of image data without manual intervention. Learning-based segmentation methods include machine learning-based algorithms and deep learning-based algorithms, greatly improving the accuracy and efficiency of segmentation by manually selecting or automatically learning image features from large amounts of data.

This chapter focuses on the application of deep learning in medical image segmentation; the history and basic theory of deep learning are reviewed. The development and iteration of a large number of deep learning algorithms reveals how this technology is driving the development of medical image processing. In particular, the introduction of fully convolutional neural networks (FCN) and U-net architectures demonstrates how deep learning can be specifically optimised to adapt to the unique needs of medical image segmentation. U-net architecture has received special attention due to its high sensitivity to small objects in medical images and effective utilization of image context information and has become an important development in the field of medical image segmentation.

## **Chapter 4. CT dataset and evaluation parameters**

As one of the most widely used benchmark datasets for liver tumour segmentation, the Liver Tumour Segmentation (LiTS) dataset [17] is selected in this study due to its clinical diversity, expert-validated annotations, and widespread adoption in academic research. Advances in medical imaging technology have resulted in an explosion of high-quality imaging data, necessitating the development of efficient and accurate image analysis methods. Among them, liver and tumour segmentation from abdominal CT images is a critical task for diagnosis, treatment planning, and disease progression monitoring. The LiTS dataset [17] is derived from seven clinical sites around the world and represents a comprehensive collection of such images. The dataset is both large and diverse, covering a variety of liver tumour types of varying stages, sizes and stages of treatment. Segmentation of liver and tumour structures was validated by experienced radiologists, providing a solid basis for developing and testing automated segmentation algorithms. This chapter describes the complexity of the dataset, including the distribution of liver volumes, tumour incidence, and the complexity of tumour characteristics. Furthermore, it outlines the basic evaluation parameters that are crucial for evaluating the performance of segmentation models, thereby guiding researchers and developers to improve their methods to achieve higher accuracy and reliability in medical image processing.

### **4.1 The Liver Tumour Segmentation (LiTS) Dataset**

The image data for the LiTS [17] challenge is collected from seven clinical sites all over the world, including medical centres in Germany, the Netherlands, Canada, Israel, and France. The LiTS dataset includes abdominal CT images of 201 patients, of which 194 have liver lesions. This dataset covers a variety of liver tumour types, including primary and secondary liver tumours. The images are of varying quality and resolution, showing different tumour-to-background ratios, as these images contain a mixture of pre- and post-treatment abdominal CT scans and are acquired using different CT scanners and acquisition protocols.

Data collection encompasses a wide range of imaging technologies from different vendors, employing a variety of CT scanners and protocols, resulting in significant variation in image resolution and quality. The axial resolution of the data ranges from  $0.56\text{ mm}$  to  $1.0\text{ mm}$ , while the z-axis resolution ranges from  $0.45\text{ mm}$  to  $6.0\text{ mm}$ . Furthermore, the number of z-axis slices varies from 42 to 1026, and some scans show imaging artefacts such as metal artefacts, reflecting challenges encountered in real clinical settings.

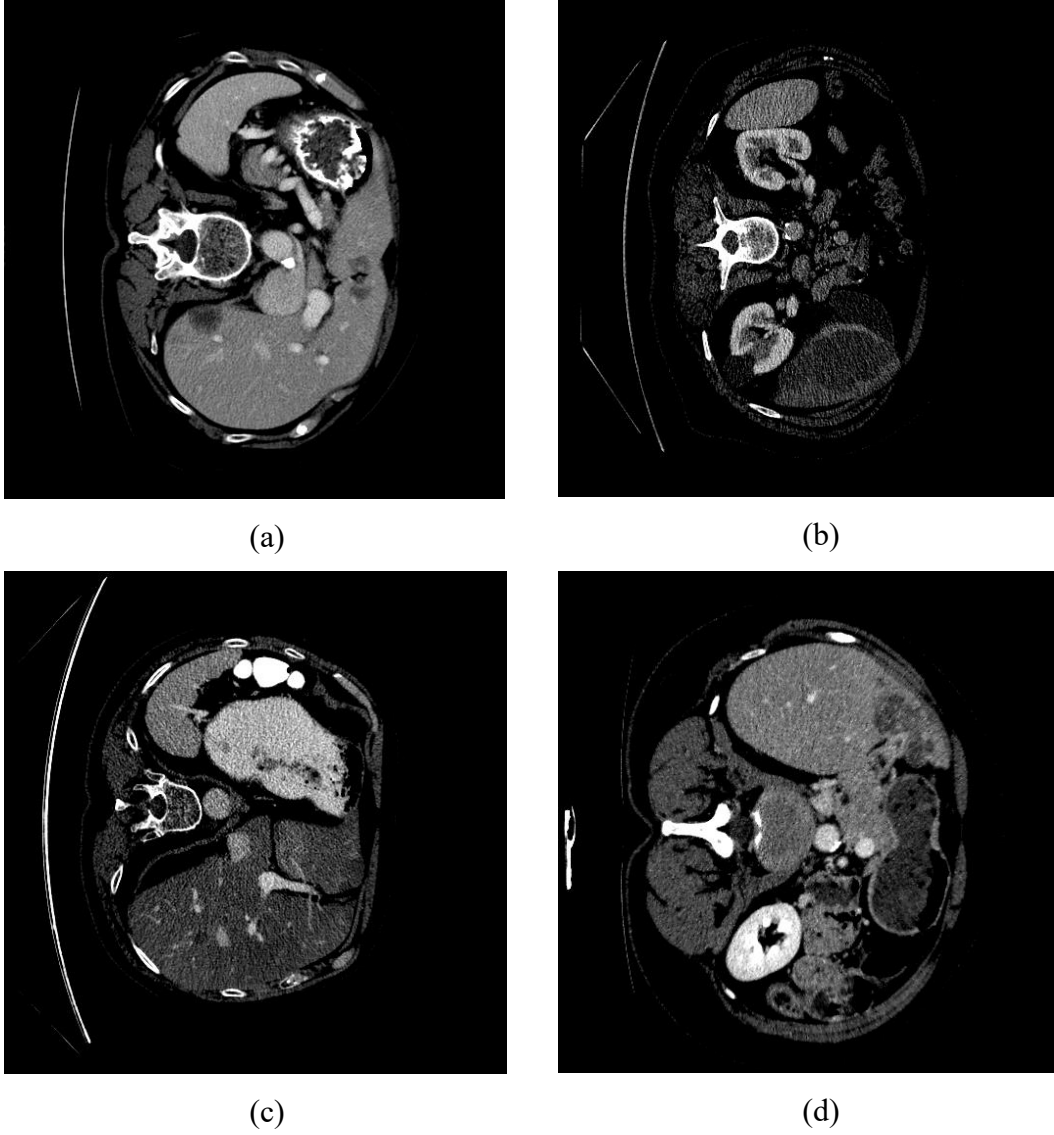
The initial ground truth segmentation of the liver and tumour structures are generated by trained radiologists at each participating clinical site. After the initial annotation, the segmentation masks undergo a verification process by three experienced radiologists in a blinded review. This additional step is crucial for maintaining the quality and reliability of the dataset. By involving multiple radiologists in both the annotation and verification phases, the dataset aims to minimise subjective bias and maximise the accuracy and reliability of the segmentation masks provided for training and testing automated segmentation algorithms.

The distribution of liver volumes in the training and test data sets is approximately normal. The incidence of tumours in these datasets varies widely, from none to as many as 12, with the test set showing a higher incidence of tumours compared to the training set. Tumour sizes range from  $38\text{ mm}^3$  to  $1231\text{ mm}^3$ . Mean difference in intensity between liver and tumour tissue, calculated as the mean absolute difference in Hounsfield unit (HU) values between liver and tumour voxels, ranging from 0 to 98 with a mean of 31.94 (SD = 20) and the median is 29.61. Figure 4.1 shows four CT scans of the abdominal cavity with different HU values. This variation in tumour characteristics and imaging conditions highlights the complexity and diversity of LiTS challenge datasets.

From a total of 201 CT volumes, 131 are chosen randomly for training and 70 CT volumes for the test set. The split is performed in such a way to have the same clinical site distribution in training and test set. The participants could download training data



from the LiTS Challenge website including training reference data, The test data without ground truth segmentation is also available for download.



**Figure 4.1 Four CT scans with different HU values, which can be seen from the depth of the liver.**

To enable model training, the original NIFTI-format 3D CT volumes are converted to 2D axial slices and stored as PNG images. A standard liver window is used to normalize the intensity values, which has been described in Section 3.2. In addition, histogram equalization is applied to enhance image contrast, and all slices are resized to  $512 \times 512$  pixels to retain sufficient anatomical details. All images are normalized to a fixed intensity range to ensure consistency across the dataset.

The LiTS dataset also suffers from a severe class imbalance problem, as tumor voxels account for a very small proportion of the entire image volume. Moreover, the features of tumours vary greatly in size, number, and intensity. To alleviate this problem, we carefully construct the test set to contain a balanced representation of large, medium, and small tumours (with a ratio of 1:1:1). In addition, the training loss function is designed to combine Dice loss and binary cross entropy loss to enhance the robustness of the model under imbalanced conditions.

## 4.2 Evaluation Parameters

In the field of medical image processing, especially in tasks like semantic segmentation, model evaluation plays an indispensable role in the entire development process. Quantifying the performance of the model through different evaluation metrics can help developers or researchers intuitively evaluate the performance of the model and make subsequent adjustments and optimisations.

The Confusion Matrix is an important tool for evaluating model performance and the basis for evaluating classification task performance. It provides an intuitive way to identify the model's performance on different categories by comparing the model's predictions with the true values, revealing the model's accuracy and error in predicting positive and negative categories. The confusion matrix involves four basic terms: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). Table 4.1 Confusion Matrix provides the discriminant conditions of the confusion matrix.

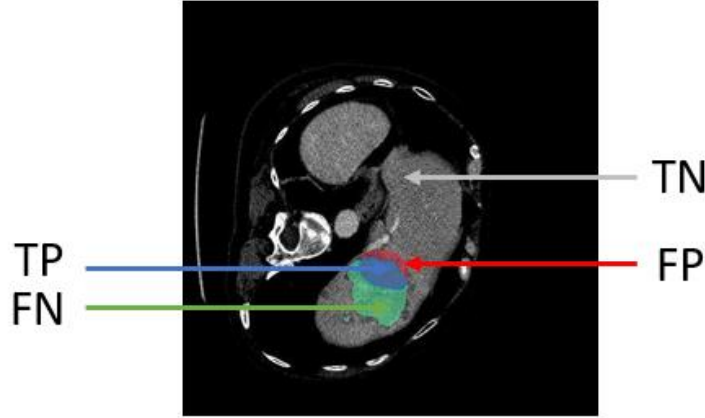
**Table 4.1 Confusion Matrix.**

		Predict Class	
Reality/Prediction		Predict Positive	Predict Negative
Actual Class	Actual Positive	True Positive	False Negative
	Actual Negative	False Negative	True Negative

The four elements of the confusion matrix are defined as follows:

1. True Positives (TP): When the model predicts a positive class, and the prediction is correct, this situation is called a true positive. In the liver/tumour segmentation task, TP represents the number of pixels or voxels correctly predicted by the model to be liver/tumour.
2. False Positives (FP): When the model incorrectly predicts a negative class as a positive class, this situation is called a false positive. In the liver/tumour segmentation task, FP represents the model incorrectly predicted non-liver regions as pixels or voxels of the liver/tumour.
3. True Negatives (TN): When the model predicts a negative class, and the prediction is correct, this situation is called a true negative. In the liver/tumour segmentation task, TN represents the number of pixels or voxels correctly predicted by the model as non-liver/non-tumour.
4. False Negatives (FN): When the model incorrectly predicts a positive class as a negative class, this situation is called a false negative. In the liver/tumour segmentation task, FN represents the model failed to predict the liver/tumour region as the number of pixels or voxels of the liver/tumour.

Figure 4.2 indicates four elements in liver tumour segmentation. The difference between segmentation tasks and classification tasks is that segmentation tasks require attention to the classification performance of each pixel or voxel.



**Figure 4.2 Confusion matrix in the liver tumour segmentation.**

From the confusion matrix, more advanced classification indicators can be obtained, including Accuracy, Precision, Recall, Specificity and F1score.

Accuracy measures the overall ability of a classifier to predict correctly. It is the ratio of correctly predicted samples (positive and negative classes) to the total number of samples. Its mathematical expression is:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.1)$$

Although accuracy is one of the most intuitive evaluation indicators, it does not always provide sufficient information to judge the quality of an algorithm. Especially when the data set is imbalanced, accuracy can be misleading.

Specificity measures the proportion of true negative samples that are correctly identified by the model. A high specificity indicates the model's effectiveness in correctly recognizing negative cases, resulting in fewer false positives. Its mathematical expression is:

$$Specificity = \frac{TN}{TN + FP} \quad (4.2)$$

Precision measures the proportion of samples correctly identified as positive by the model that are actually positive. A high precision means that among the elements predicted by the model to be positive, the proportion of elements that are actually

positive is higher. Its mathematical expression is:

$$Precision = \frac{TP}{TP + FP} \quad (4.3)$$

Recall also called Sensitivity measures the proportion of all actual positive classes that are correctly recognised by the model. A high recall means that the model can capture more positive examples, as well as lower false negative. Its mathematical expression is:

$$Recall = \frac{TP}{TP + FN} \quad (4.4)$$

It is not scientific and comprehensive enough to evaluate an algorithm model solely by one standard. For example, sensitivity and precision are often a pair of contradictory indicators in classification problems. When improving precision, some true positives may be missed, resulting in a decrease in sensitivity. On the contrary, when sensitivity is improved, but false positives are increased at the same time, resulting in decreased precision. The F1 Score can be used to balance precision and sensitivity. It is the harmonic mean of accuracy and sensitivity and can be used as a comprehensive indicator of both.

$$F1 \text{ score} = 2 \times \left( \frac{Precision \times Recall}{Precision + Recall} \right) = \frac{2TP}{2TP + TN + FP} \quad (4.5)$$

The F1 score ranges from 0 to 1, where 1 indicates perfect precision and recall, and 0 indicates that at least one metric is zero.

In the LiTS challenge set, six official evaluation criteria are provided to measure the performance of liver or tumour segmentation. A brief introduction on six statistical performance measures commonly utilised in liver related segmentation methods is given below [17].

The first one to introduce is the dice score. On the two-classification problem, the dice score can be considered to be equivalent to the F1 score introduced before. The dice coefficient is a set similarity measure function, used to calculate the similarity between samples, and the value range is between [0, 1]. Its mathematical expression is

$$Dice(X, Y) = \frac{2TP}{2TP + FN + FP} \quad (4.6)$$

For semantic segmentation problems, as shown in Figure 4.3, the dice coefficient can be understood as

$$Dice(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|} \quad (4.7)$$

Where  $X \cap Y$  represents the intersection of X and Y, and X and Y represent the number of elements in the sets X and Y. The coefficient of 2 for the numerator part is due to double counting of common elements between X and Y. When X represents prediction result, Y represents the ground truth.

$$Dice = \frac{2 \times \text{Area of overlap}}{\text{Total area}} = \frac{2 \times \text{Prediction} \cap \text{Ground truth}}{\text{Prediction} + \text{Ground truth}}$$

**Figure 4.3 Dice score in segmentation problem.**

In LiTS challenge, Global Dice score and Dice per Case are commonly employed.

Global Dice Score is calculated by aggregating all true positive (TP), false positive (FP), and false negative (FN) voxels across the entire dataset before computing the Dice coefficient. It measures the overall segmentation accuracy across all slices and patients, and it is especially sensitive to cases with large volumes or a high number of slices.

$$Global\ Dice = \frac{2 \cdot \sum TP}{2 \cdot \sum TP + \sum FN + \sum FP} \quad (4.8)$$

Dice per Case, in contrast, is calculated individually for each patient by evaluating the Dice score for that patient's entire scan (or tumour region), and then averaging the results over all patients. This metric gives equal weight to each patient regardless of

tumour size or the number of tumour-containing slices.

$$Dice\ per\ Case = \frac{1}{N} \sum_{i=1}^N \frac{2 \cdot TP_i}{2 \cdot TP_i + FN_i + FP_i} \quad (4.9)$$

These two metrics complement each other, while Global Dice reflects overall volumetric performance, Dice per Case captures case-wise consistency and fairness. Significant discrepancies between the two may indicate dataset imbalance or inconsistent model performance across cases with different tumour sizes.

Volume Overlap Error, as its name suggests, is used to quantify the overlap error between two volumes or areas. Its mathematical formula is as follows:

$$VOE(X, Y) = 1 - \frac{|X \cap Y|}{|X \cup Y|} \quad (4.10)$$

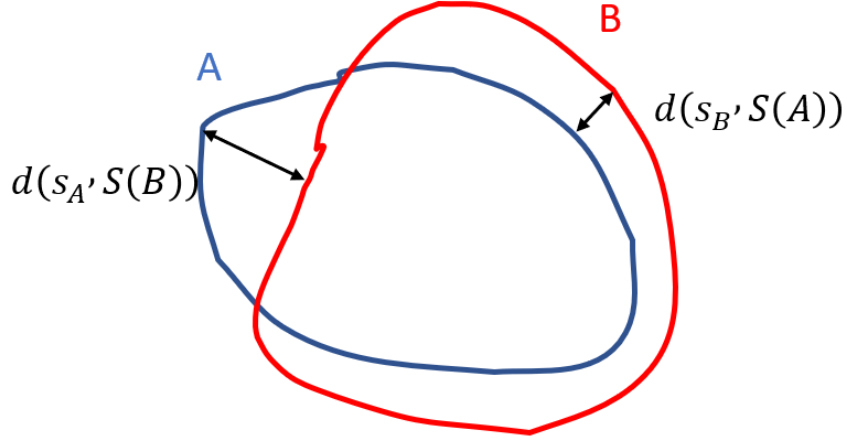
Where  $\frac{|X \cap Y|}{|X \cup Y|}$  represents the Jaccard index, which is the intersection and union ratio.

When the Jaccard index is equal to 1, VOE equals 0 means that the two parts completely overlap, which means that the image is perfectly segmented.

Relative Volume Distance is another metric used to evaluate the difference between two volumes. Its mathematical formula is as follows:

$$RVD(X, Y) = \frac{|X| - |Y|}{|Y|} \quad (4.11)$$

Where X is the predicted volume and Y is the true volume. RVD can be positive or negative depending on whether the predicted volume is larger or smaller than the true volume. If the predicted volume is greater than the true volume, RVD is positive; otherwise, RVD is negative. It is clear that smaller absolute value of RVD means better segmentation result.



**Figure 4.4 Example of calculating the ASSD and MSD of two graphs.**

Average symmetric surface distance (ASSD) and Maximum symmetric distance (MSD) are measures of similarity between two sets of points in a space. ASSD is calculated by taking the average distances from each point in one set and its closest point in the other set. In the biomedical image segmentation field, it is the average of all distances from points on the segmentation boundary to the boundary of ground truth. Its calculation formula is as follows,

$$ASD(A, B) = \frac{1}{|S(A) + S(B)|} \times \left( \sum_{s_A \in S(A)} d(s_A, S(B)) + \sum_{s_B \in S(B)} d(s_B, S(A)) \right) \quad (4.12)$$

Where  $S(A)$  and  $S(B)$  denote the set of surface voxels of A and B,  $d(s_A, S(B))$  represents minimum Euclidean distance from each point on A to edge B,  $d(s_B, S(A))$  represents minimum Euclidean distance from each point on B to edge A, as shown in Figure 4.4. A smaller ASSD value indicates better segmentation results. Similar to ASSD, MSD still finds the distance between each point in one set and its closest point in the other set. But it selects the maximum distance from points between two image voxels as the measurement results, which can be calculated as:

$$MSD(A, B) = \max \left\{ \max_{s_A \in S(A)} d(s_A, S(B)), \max_{s_B \in S(B)} d(s_B, S(A)) \right\} \quad (4.13)$$



### **4.3 Conclusion**

The LiTS dataset presents unique challenges and opportunities to the field of medical image analysis. Through a detailed exploration of their composition, characteristics, and the rigorous validation process they undergo; this chapter highlights the critical role of high-quality, diverse datasets in advancing automated segmentation technology. Evaluation metrics, including confusion matrix, accuracy, precision, recall, and F1 score, are essential tools for evaluating the effectiveness of segmentation models. These metrics are complemented by more specialized metrics such as Dice score, volumetric overlap error, and average symmetry distance, allowing for a nuanced assessment of model performance on different aspects of segmentation tasks. The analysis presented here not only highlights the inherent complexity of liver tumour segmentation but also points to the ongoing need for innovative methods that can expertly handle the variability of real medical imaging data.

# **Chapter 5. Liver segmentation method using an improved Multi Scale Feature U-Net (MSF U-Net)**

## **5.1 Introduction**

Successful liver image segmentation can yield accurate liver volume, which has an important impact on liver surgery and patient's postoperative recovery evaluation. However, due to highly variable shape, close proximity to other organs and diverse pathologies, the liver tissue in the abdominal CT image will have the same intensity value as the adjacent organs like stomach, heart, pancreas, kidney [195]; and furthermore its shape will be deformed. In clinical applications, liver segmentation is usually done by experienced radiologists. This remains an extremely time-consuming task that can reach 90 minutes or more per patient, with a degree of inter and intra segmentation errors.

This chapter introduces an improved U-Net model, which aims to automatically carry out image segmentation of the liver from CT data. Building on the foundational architecture U-Net covered in Section 3.4.4, the improved model proposed in this chapter makes several innovations based on the original architecture to improve segmentation performance and efficiency. These include the use of techniques such as pyramid pooling and dilated convolution, that aim to enhance the model's ability to obtain contextual information and the size of the receptive field when processing complex liver CT images.

The remainder of the chapter is organised as follows. Section 5.2 reviews the basic structure and function of the classic U-Net and then elaborates on the concepts of pyramid pooling modules and dilated convolutions and their specific applications in the improved Multi Scale Feature U-Net. Specific implementation details of using this model for liver segmentation, including the processing of the dataset, the network training process, and the visualisation analysis of the results are described in Section 5.3. Section 5.4 provides the conclusion of this Chapter.

## 5.2 Liver CT scans segmentation using Modified MSF U-Net

### 5.2.1 Overview of Modified MSF U-Net

A modified U-Net is proposed to achieve automatic liver segmentation from CT scans. As introduced in Section 3.4, a classic U-Net is mainly composed of a down-sampling path, an up-sampling path, and skip connections. The down-sampling path extracts features at different levels through convolution, pooling, and dropout, and the up-sampling path restores the features to the input image size through deconvolution or interpolation. The skip connections introduce the down-sampled feature information into the up-sampling process, providing multi-scale information to support image segmentation. Finally, the feature map is mapped to between 0 and 1 through the sigmoid function to generate the prediction result.

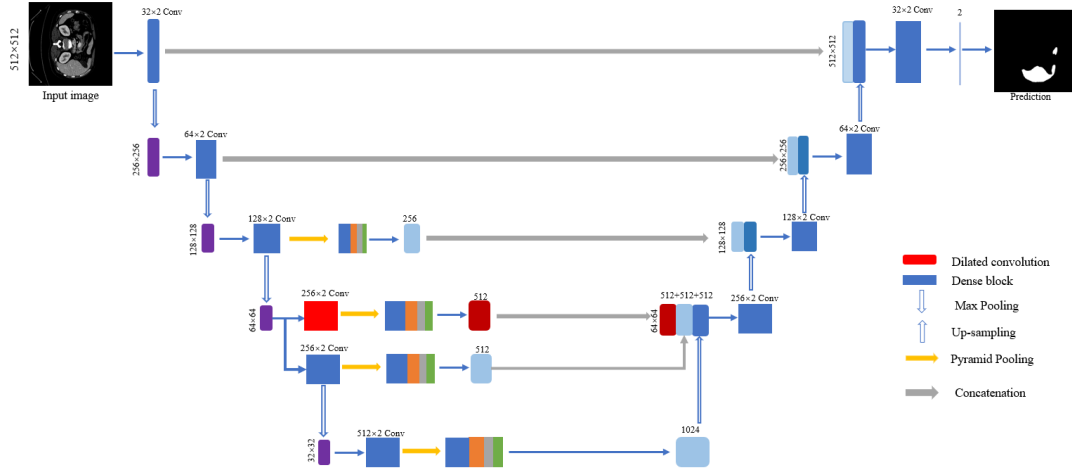
However, the original U-Net architecture has limitations when applied to liver segmentation in CT scans. Specifically, it may struggle to capture complex boundaries or small lesions accurately, and the decoder may not fully utilise deep semantic information. To address these issues, modifications are introduced to the U-Net structure. These changes aim to enhance feature representation, improve boundary delineation, and increase the model's robustness in handling liver shape and intensity variation. The rationale behind these changes is based on the need for finer localization and stronger contextual understanding in challenging medical imaging processing. Based on the original U-Net, several adjustments are made to achieve better segmentation performance, including using pyramid pooling to obtain contextual information at different scales and using dilated convolution to obtain a larger receptive field in high-dimensional channels. These two parts will be introduced in detail in the following part.

The main path of U-Net contains four down-sampling and up-sampling. On the encoder path, the number of convolution kernels is increased from 64 to 1024 in a double increment and the size of the convolution kernel is set to  $3 \times 3$ . As illustrated in Figure 5.1, each dense block contains two convolution layers, each followed by batch

normalization and ReLU activation, with ‘same padding’ (described in Section 3.4.2.1) to ensure that the final prediction result matches the input image. The  $2 \times 2$  max pooling with stride of 2 is used to halve the feature map size. On the decoder path, deep features are restored to feature map size through bilinear interpolation [105]. Subsequently, the feature map obtained by the corresponding encoder and the feature map obtained by up-sampling are concatenated together through skip connection; these feature blocks containing deep feature and shallow feature are passed through the feature extraction module (dense block) again to obtain more detailed features. The up-sampling operation is repeated four times to restore the abstract high-dimensional features to the same size as the original input. The number of convolution kernels is consistent with the number of convolution kernels used for feature extraction in the encoder path, which is reduced from 1024 to 64 by a factor of 2 each time. At the end of the network, the sigmoid function is used to classify the feature map between 0-1 to obtain the prediction probability map. The model generates two values for each pixel, corresponding to the probability that the pixel belongs to each category. Therefore, the last layer has two channels, and each channel outputs a probability map of the corresponding category.

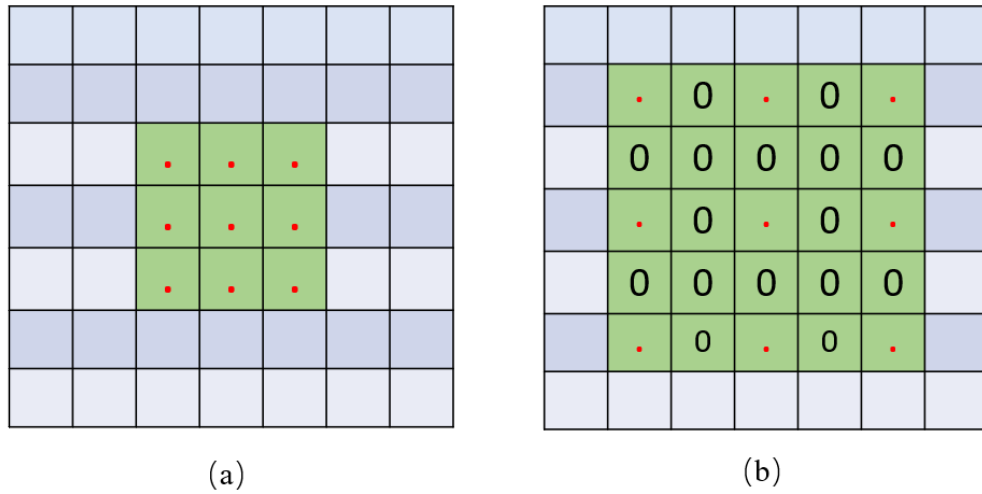
### **5.2.2 Dilated Convolution**

In deep neural convolutional networks, as the depth of the network increases and the pooling operation is involved, the number of the feature maps will gradually increase (As shown in Figure 5.1, the number of features is 1024). In order to obtain a larger receptive field without excessively increasing parameters, dilated convolutions[196] are used in the deep layers of the network to obtain broader contextual information and global features.



**Figure 5.1** The architecture of proposed modified U-Net for automatic liver segmentation.

Section 3.4.2 described the definition of reception and the calculation method of convolution. Compared with traditional convention operation, a dilated convolution introduces a dilated rate parameter based on the conventional convolution. The hole ratio indicates the "gap" between adjacent elements in the convolution kernel.



**Figure 5.2** Normal Convolution calculation vs Dilated Convolution with a dilated ratio of 2.

As shown in Figure 5.2, when the dilated ratio is 2, there will be a "hole" between each element in the convolution kernel, so that the convolution kernel covers a larger space. In this way, dilated convolution can extract feature information in a larger range without increasing the size of the convolution kernel, thereby expanding the network's

perceptual field while maintaining computational efficiency. It seems confusing because the size of the convolution kernel increases from  $3 \times 3$  to  $5 \times 5$ , but in fact only a  $3 \times 3$  convolution kernels (9 kernel values) are used with the corresponding feature values for convolution, the extra 16 positions are replaced by holes, which are zeros. Therefore, although the physical size of the convolution kernel remains  $3 \times 3$ , the receptive field is significantly increased. Dilated convolution allows the network to enlarge the receptive field without reducing the resolution of the feature maps, enabling it to capture broader contextual information. This is particularly beneficial in liver segmentation, where livers may vary in size and shape, and spatial context is critical for distinguishing organ boundaries from surrounding tissues. By incorporating dilated convolutions, the model can better aggregate global and local features, improving segmentation accuracy, especially in cases involving small or diffuse lesions.

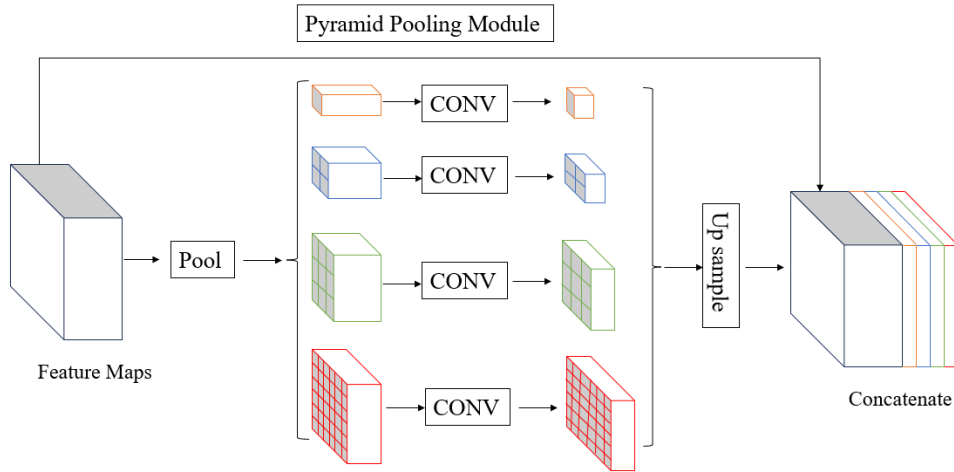
In the proposed modified U-Net, as illustrated in Figure 5.1 in the red block, the dilated convolution is used in the fourth resolution dimension, where the feature map size is  $64 \times 64$  with dilated ratio 2. The reason for merely using dilated convolution in the deep layer rather than at all resolution layer is to avoid excessive information loss. Since a large number of holes are filled in the convolution kernel, numerous local information will be ignored during the calculation process. Shallow features extraction requires continuous convolution kernels to process in shallow layers. Currently, using dilated convolution is not conducive to feature extraction in the early layers. In order to ensure the richness of information, the normal convolution kernel is used in another branch, and down-sampling is performed again to obtain more abstract deep segmentation information. The three feature maps generated by the two branches are fused together for the subsequent up-sampling process.

### 5.2.3 Pyramid Pooling Module

In image semantic segmentation, information at different scales is crucial for understanding objects and context in images, from tiny details to overall objects, which may have an important impact on the task. In a cross-section of a CT scan of the

abdomen, the liver will appear in the image in different sizes and proportions. To better capture this multi-scale information, a pyramid pooling module [189] is introduced to ensure that the network can effectively handle features of different scales.

In order to make full use of the features in each scale, a pyramid pooling structure is employed in the information extraction stage to capture more contextual details, which is to extract image features at different regional scales. First, several feature maps are extracted from the input image using one CNN. Then, these feature maps undergo pooling at different scales, resulting in multiple down-sampled versions of the original feature map. As shown in Figure 5.3, the input feature map is divided into multiple grids of different scales ( $1 \times 1$ ,  $2 \times 2$ ,  $3 \times 3$ ,  $6 \times 6$ ), and then perform a global average pooling operation which is described in Section 3.4.2 on the features in each grid to obtain feature maps of various scales. Next, a  $1 \times 1$  convolutional kernel is used to reduce the number of channels, thus reducing the network parameters. The feature maps after pooling need to be resized as the original feature map and then are concatenated along channels with the original feature map. The merged multi-scale feature map is passed to the decoder part via short skip connections.



**Figure 5.3 Architecture of Pyramid Pooling Module.**

In the proposed modified U-Net, the pyramid pooling module is applied on the third to fifth resolution layers (feature map size equals  $128 \times 128$ ,  $64 \times 64$ ,  $32 \times 32$ ). When the feature map is in this interval, the network can extract abstract semantic

information, applying pyramid pooling at these layers can ensure that multi-scale information is integrated in features that contain richer semantic information. On the other hand, using pyramid pooling operations in the first few layers, the feature layer will generate too many sub-feature maps after being divided into grids of different sizes due to high resolution of feature map, which will inevitably require increased memory and computational consumption. A joint loss function of dice loss and binary entropy loss, which is described in Section 3.4.4, is used to optimise the parameters of model.

## **5.3 Visualisation and statistical analysis of liver segmentation results**

### **5.3.1 Network Implementation**

The data set used for training and testing is the Liver Tumour Segmentation Challenge (LiTS), which is described in Chapter 4. The data set contains CT scans of 130 patients, which are divided into training sets and test sets. The training set and test set contained 100 patients and 30 patients, respectively. All scans have been provided in *nii* format with an axial size of 512×512.

To improve the generalization ability and robustness of the network, data augmentation techniques are applied during training. This is particularly important in medical image segmentation, where annotated datasets are often limited in size, and organs like the liver can vary significantly in shape, size, and location across patients. Following the strategy used in the original U-Net, the applied augmentations include small random rotation (0.2 range), shift on horizontal (0.05) and vertical (0.05) direction, shear (0.05), zoom (0.05), and flip (horizontal).

The new algorithms are implemented in Anaconda with Python, running on PC with 32GB RAM, 3.8GHz AMD Ryzen7 3800X 8-core CPU, and a NVIDIA RTX2080 GPU with a total 8GB memory. This work is implemented using Keras based on the TensorFlow backend. The proposed modified cascaded U-Net is trained for 80 epochs using the Adam optimiser, and the learning rate is gradually reduced from 1e-4 to 1e-5

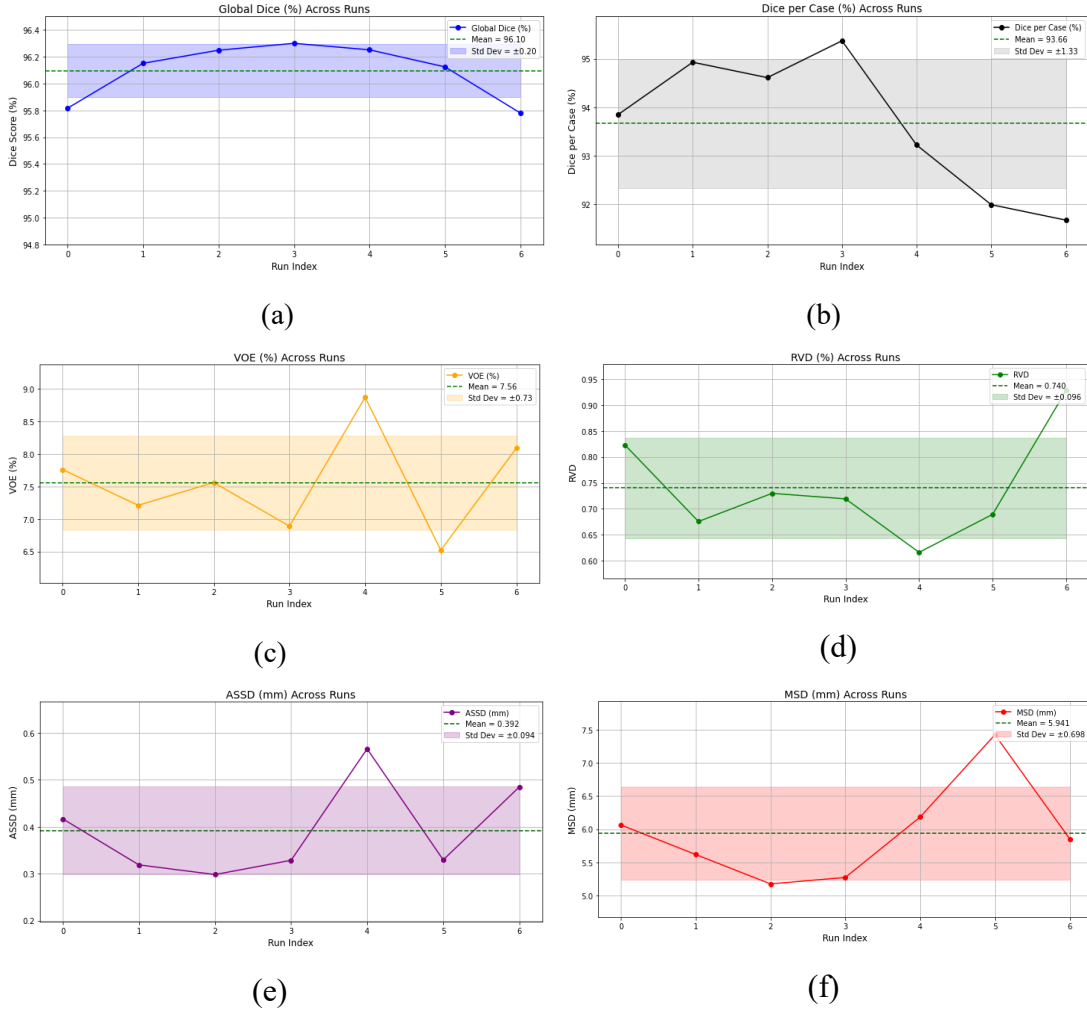


when the accuracy did not change within 5 epochs. In this approach, a batch size of 2 is set during training. This choice is primarily due to the high spatial resolution of abdominal CT images and the limited GPU memory available, which constrained the number of images that could be processed simultaneously. While large batch sizes may promote more stable convergence by narrowing the range of learning rates and producing smoother gradient estimates, they typically require much higher computational resources. In contrast, small batch sizes have been shown to improve generalization in some scenarios [197, 198], as they introduce stochasticity into the gradient estimation process, potentially helping the model avoid sharp local minimum. This is particularly important in medical image segmentation, where overfitting is a common challenge.

### 5.3.2 Statistical Analysis

To evaluate the effectiveness of the proposed segmentation method, a statistical analysis is conducted based on multiple evaluation metrics, including global Dice score, Dice per case, volume overlap error (VOE), relative volume distance (RVD), average symmetric surface distance (ASSD) and maximum symmetric distance (MSD), which have been discussed in Section 4.2. These metrics together reflect the volume accuracy and boundary precision, thus providing a comprehensive evaluation of the segmentation performance.

Table 5.1 presents six relative metrics of ten different liver segmentation methods. The method proposed in this study, as the improved U-Net using multi-scale feature fusion, achieves a global Dice coefficient of 0.961, which is comparable to the manual annotation results of professional radiologists [199]. To ensure the robustness and statistical reliability of the proposed method, experiments are conducted across seven random training–testing splits, and the results are visualised in Figure 5.4.



**Figure 5.4 Statistical analysis of segmentation performance across seven random training-testing splits using six evaluation metrics.**

Figure 5.4 presents the statistical evaluation of the proposed segmentation model across seven independent random training–testing splits, using six commonly adopted metrics. The global Dice scores and Dice per case across seven runs show a narrow standard deviation of  $\pm 0.20$  and  $\pm 1.33$ , respectively, indicating high consistency and robustness of the model across different data splits. Similarly, the variability in VOE ( $\pm 0.20$ ), RVD ( $\pm 0.096$ ), ASSD ( $\pm 0.094$  mm), and MSD ( $\pm 0.698$  mm) remains within acceptable bounds reinforcing the generalization capability of the model.

VOE and RVD assess volumetric agreement between predicted and ground truth regions, where lower values denote more accurate and balanced volume estimates. The proposed method achieves a mean VOE of 7.56% and a mean RVD of 0.74%, reflecting reliable volumetric performance. ASSD and MSD evaluate boundary precision. A lower

ASSD indicates finer alignment of the predicted segmentation with the true anatomical boundary. The model obtains a mean ASSD of 0.392 mm and a mean MSD of 5.941 mm, demonstrating high accuracy even in delineating complex or irregular boundaries.

**Table 5.1 Comparison of the proposed model with other nine liver segmentation methods based on six measured metrics. The symbol “N/A” represents unreported result. Bold font represents the highest score on each measurement.**

Methods	Global Dice(%)	Dice per case (%)	VOE (%)	RVD (%)	ASSD (mm)	MSD (mm)
Manual [196]	98.6	N/A	N/A	N/A	N/A	N/A
Level sets[135]	N/A	N/A	10.4	2.7	1.9	28.9
Region growing[200]	N/A	N/A	26.4	-11.5	10.2	74.0
U-Net+TL loss[201]	94	<b>95</b>	10	5	1.89	32.71
U-Net+Level set[202]	95.6	N/A	N/A	N/A	N/A	N/A
Semi-supervised[203]	84.85	N/A	N/A	N/A	N/A	N/A
DAR-Net[204]	96.13	N/A	5.03	<b>0.38</b>	N/A	5.32
RDCTrans-Unet[194]	93.38	N/A	N/A	N/A	N/A	N/A
MCI-Net[205]	<b>96.22</b>	N/A	<b>2.802</b>	-0.637	1.711	N/A
Cascade ResNet[206]	95.51	N/A	8.55	N/A	N/A	N/A
MSF U-Net	96.10	93.66	7.56	0.74	<b>0.392</b>	<b>5.941</b>

The following is a comparative analysis: Comparing traditional methods such as region growing and level set with the proposed model, it is evident that the MSF U-Net significantly outperforms these methods, particularly in terms of average surface ASSD and MSD. For instance, MSF U-Net achieves an ASSD of 0.282 mm, compared to 10.2 mm for region growing, demonstrating its superior accuracy in liver contour segmentation. Moreover, traditional methods necessitate multiple iterative steps for parameter optimisation, requiring over 20 minutes to process a set of ten images. In contrast, the proposed deep neural network significantly accelerates this process, achieving the same task with an obviously reduced inference time. MSF U-Net processes each image in approximately 0.2 seconds, making it more efficient for

clinical applications where time-sensitive decision-making is crucial. This marked improvement in both accuracy and speed highlights the clear advantages of using deep learning models over traditional approaches.

Compared to the standard U-Net architecture with special Tversky loss functions or post-processing the U-Net output using the level set method, the proposed method still slightly improves the Dice score of the model, thereby intuitively achieving a better image segmentation effect. Compared to cascaded network architectures, the proposed model achieves a comparable Dice score and superior Volume Overlap Error (VOE) with a more concise and lighter structure, utilising fewer network layers.

In comparison with the RDCTrans U-Net model, although Li et al. integrated a Transformer module at the bottom layer of the network to capture the dependencies between different positions in the image and establish a feature representation of the global context, the proposed method still shows a higher Dice score. In addition, the pyramid pooling strategy adopted in the deeper network layers also effectively incorporates contextual information, further enhancing the performance of the model.

The result is not as good as MCI-Net and DAR-Net. In the MCI network, Xie et al. designed a structure called Multi-scale Context Extraction Module, which is embedded in each feature extraction layer. The structure of this module is similar to the pyramid structure, but unlike the traditional pyramid, it uses more sparse convolutional layers to enhance the network's ability to capture contextual information. This design enables the MCI network to show higher accuracy in feature extraction and context information integration. From the experimental results, the MCI network shows better accuracy than the proposed network on the target task. However, when evaluated from the perspective of network capacity and computational cost, the proposed network has higher efficiency. Although the MCI network improves performance through its unique structure, it also leads to a larger amount of network parameters and computing resource consumption.

In another article published by Xie [204], dynamic adaptive pooling are used to replace the commonly used max pooling. This method also applies mathematical

modelling to analyse the relationship between the maximum value and other values of the pooling area to achieve a more refined pooling effect. In addition, they post-process the network's predicted values using CRF to further improve the accuracy.

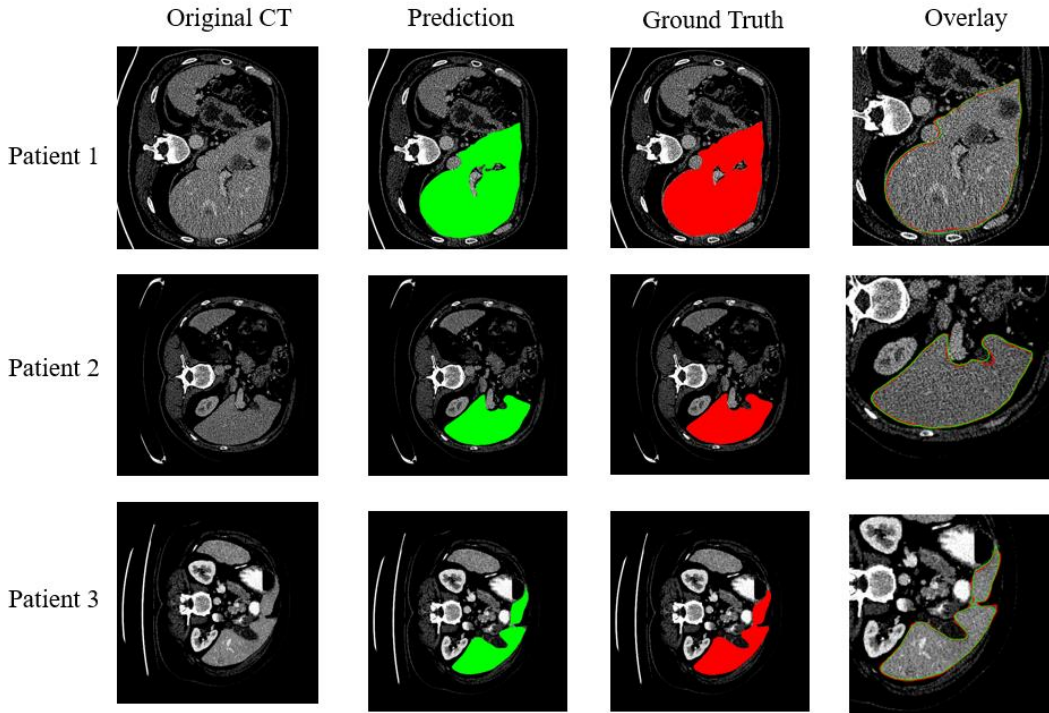
In summary, although the proposed method does not surpass all the best methods of the same kind in terms of Dice coefficient, it demonstrates outstanding performance in key metrics such as average surface distance and maximum surface distance, where it significantly outperformed traditional methods. These metrics are critical for accurately capturing the liver contour. In addition, the proposed method shows potential in terms of ease of use, computational efficiency, and generalization ability. Future work will explore further architecture optimisation and training strategies to comprehensively improve the performance of the model.

Figure 5.5 shows the visualisation of liver segmentation using the improved U-Net method on a patient's abdominal CT scan. The images are arranged in rows, each row corresponds to a different patient, and the columns represent the original scan, the predicted result, the true label, and the boundary overlay result. The first column presents the original CT scans for three different patients. These scans serve as the baseline images against which the segmentation results are compared. The second column displays the segmentation predictions generated by proposed modified U-Net method. The areas predicted to be liver are highlighted in green. This visualisation allows for a direct comparison of the model's output with the ground truth. The third column provides ground truth annotations for the livers, marked in red.

These annotations represent the manually labelled tumour regions by experts, serving as a reference standard for evaluating the accuracy of the model's predictions. The final column shows an overlay of the original CT scan, the prediction, and the ground truth. In this overlay, correctly predicted tumour regions appear as a blend of green and red, while discrepancies between the prediction and ground truth are visible in their respective colours. This visualisation helps in assessing the degree of overlap and identifying areas where the model either succeeded or failed to match the ground

truth.

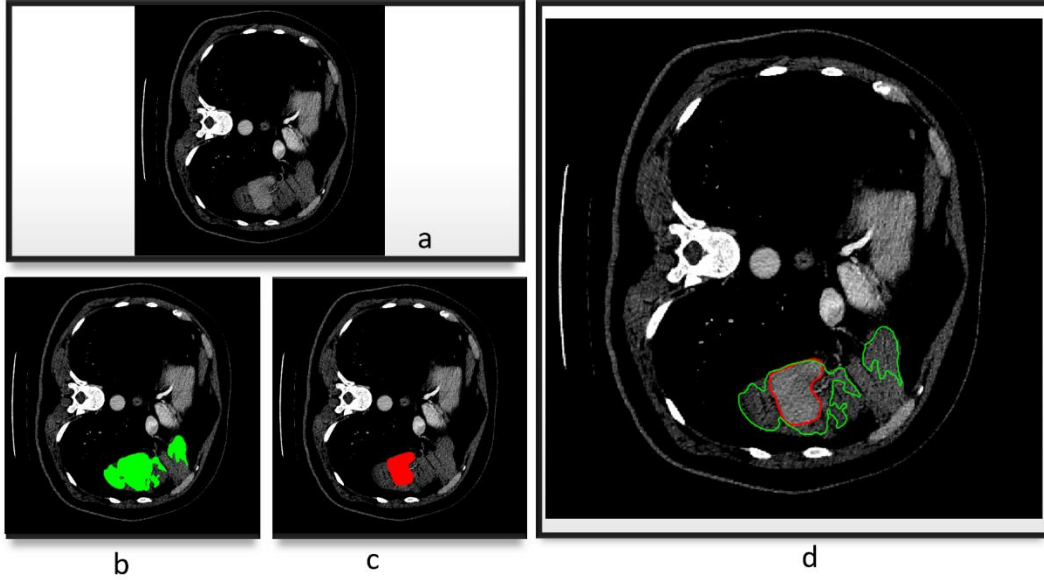
Since the liver has a relatively regular morphological structure and is highly recognizable in CT images, most prediction results can usually accurately approximate the true label. This morphological stability and predictability enable the model to achieve high accuracy when performing liver segmentation. It is also easy to see from the visualisation results given that the prediction results are almost the same as the true label. Of course, under-segmentation or over-segmentation may still occur on some fuzzy boundaries.



**Figure 5.5** Visualisation results of liver segmentation from three patients, each row corresponding to a different patient, and columns representing different stages of the analysis process.

Although the proposed method performs well in most cases, there are still some limitations in some challenging scenarios. As shown in Figure 5.5, the model incorrectly classifies the area around the liver as part of the liver due to the similar contrast. In addition, anatomical abnormalities, compressed liver morphology, or large lesions with unclear boundaries make the segmentation task more complicated. These

issues show that the model still has difficulty distinguishing complex boundary regions, especially in cases of poor image quality or abnormal anatomical structures. Future improvements may include the use of multimodal imaging (such as enhanced CT or MRI), combined with anatomical prior knowledge, to alert clinicians to potential unreliable predictions.



**Figure 5.6** A challenging case where the liver segmentation model produces significant errors. (a), (b), (c), and (d) represent the original image, liver prediction result, liver label, and overlapped image, respectively. The green contour shows the predicted liver region, while the red represents the ground truth.

## 5.4 Conclusion

In this chapter, an MSF U-Net model tailored for automatic liver segmentation from CT scans is present. The modifications introduced to the traditional U-Net architecture, such as the incorporation of pyramid pooling and dilated convolutions, are designed to improve the model's ability to capture contextual information and manage complex variations in liver morphology. The enhancements in the model's structure allowed for more effective processing of the liver's diverse shapes and proximity to other abdominal organs, which often present similar intensity values in CT images.

The chapter began by revisiting the foundational structure of the classic U-Net, followed by a detailed discussion of the new components integrated into the modified U-Net. The contribution of the pyramid pooling module and dilated convolutions to the

model's improved performance is explored, focusing on their roles in expanding the receptive field and capturing multi-scale features. Additionally, the implementation details, including data preprocessing, network training, and the visualization of segmentation results are elaborated to provide a comprehensive understanding of the proposed approach.

The experimental results demonstrated that the modified U-Net achieves competitive performance across multiple metrics. Through a comparative analysis with other state-of-the-art methods, the proposed model illustrated a balanced performance with notable strengths in computational efficiency and generalization ability. Despite not surpassing all benchmark methods, particularly in terms of Dice coefficient, the proposed approach proved to be a robust and efficient solution for liver segmentation tasks.



# **Chapter 6. Novel CT liver tumour segmentation method based on Attention Module**

## **6.1 Introduction**

This chapter will focus on the segmentation method of liver tumours. Compared to liver CT image segmentation, tumour segmentation is more challenging due to the diversity and complexity of tumours in terms of shape, size, borders, density and serious class imbalance problem.

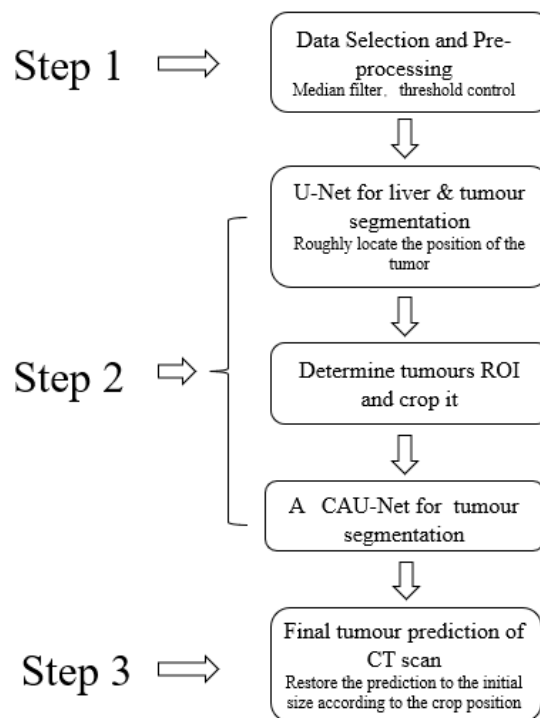
Tumour segmentation from CT images is crucial for early cancer diagnosis, treatment planning, efficacy monitoring, and medical research. It not only helps doctors detect tiny tumour lesions, but also provides detailed information about tumours such as volume, shape, location, and density. This aids in quantitative analysis, surgical planning, radiotherapy and chemotherapy planning, and monitoring of treatment effects. In addition, tumour segmentation plays a key role in cancer research and clinical trials, driving advancements in the medical field. Through accurate tumour segmentation, patients' survival rate and quality of life can be better improved.

In this Chapter, deep learning for liver tumour segmentation from CT data will be presented. The work will involve two modified versions of U-Net. The remainder of this chapter is organised as follows. Section 6.2 introduces the use of a cascaded U-Net with a custom attention mechanism (CAU-Net) to complete tumour segmentation. The custom attention module focuses on the approximation of feature regions in the up-sampling stage, guiding the network to focus on key features in the ROI region. Section 6.3 describes another cascaded hybrid attention network (CHAU-Net) which can be trained end-to-end. The proposed hard attention module allows the network to obtain the ROI region during training, and the soft attention mechanism is responsible for feature weighting during feature extraction, thereby selecting more critical semantic information. In the Section 6.4, the conclusion of this chapter will be summarised.

## 6.2 A novel Cascade U-Net with customised attention mechanism (CAU-Net) for liver tumour segmentation

### 6.2.1 Overview of the liver tumour segmentation process

The proposed segmentation workflow is shown in Figure 6.1. The workflow consists of three major steps. The first step includes data selection and data preprocessing for the network. In the second step, a cascade of a U-Net followed by a CAU-Net are used to jointly segment the liver and tumour, where the tumour segmentation is based on the segmentation result of the liver. This two-stage design reduces false positives by limiting the tumor search space to anatomically plausible regions, which is particularly important in abdominal CT where tumours may have low contrast and appear similar to surrounding tissues. In the final step, the predicted tumour lesions are restored to the original image size according to the boundary of the cropping within the network.



**Figure 6.1** The flow chart for proposed liver tumour segmentation process.

The liver is segmented using the method proposed in Chapter 5 to determine the approximate location of the liver and the tumour on the liver. The purpose of segmenting the liver is to exclude those tumour predictions out of the liver organ. Based on the prediction results of the first liver segmentation network, the input image is cropped to obtain a smaller input image to increase the ratio of foreground to background, which can improve the class imbalance problem. The centre position  $(x_i, y_i)$  of the tumour distribution can be calculated using the four boundary points of the tumour, and a  $512 \times 512$  image is then cropped around this centre point. The specific cutting position is:

$$w_l = x_i - 128, \quad w_r = x_i + 127 \quad (6.1)$$

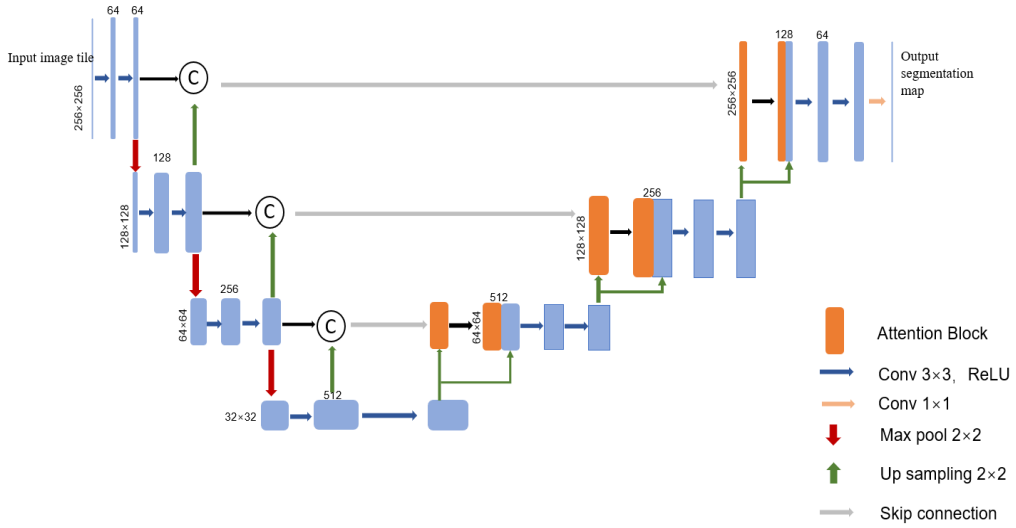
$$h_u = y_i - 128, \quad h_b = y_i + 127 \quad (6.2)$$

where  $w_l$  and  $w_r$  represent the starting and ending points of horizontal cutting respectively, and  $h_u$  and  $h_b$  respectively represent the starting and ending points of longitudinal cutting. All images in the dataset were individually reviewed and automatically cropped to obtain slices containing the liver region. The cropping positions were recorded for each case to ensure that the tumor predictions could be accurately mapped back to the original image space. However, the image resolution is not altered during this process, and no contrast enhancement or other intensity adjustments were applied after cropping. A CAU-Net with a custom attention module is used to perform high-precision tumour segmentation on the cropped ROI. The predicted segmentation result is restored to the size of the original CT scan, and the final predicted area of the tumour is restored according to the cropping position.

### 6.2.2 Details of proposed Customised Attention U-Net (CAU-Net)

As illustrated in Figure 6.2, the fine tumour segmentation process comprises a U-Net structure. The initial U-net designed 4 or 5 pooling layers for extracting deeper features. However, the excessive pooling process will inevitably lead to the loss of information, which is not negligible in small target segmentation. The size of earlier liver tumours is usually less than 3 cm, and it only accounts for approximately 5%-8% of the entire CT

image. After several times of down-sampling, this important tumour information will be reduced or even disappear completely. Based on experiments, using three pooling layers provides a good balance between feature abstraction and spatial detail preservation. Increasing the depth to four pooling layers does not lead to noticeable improvements in segmentation performance, while adding additional computational complexity. Both three and four pooling layer configurations are tested, with the four-layer version resulting in only marginal changes in Dice score ( $<0.3\%$ ) while increasing model size and training time. Therefore, the three-pooling-layer configuration is selected as a more efficient design.

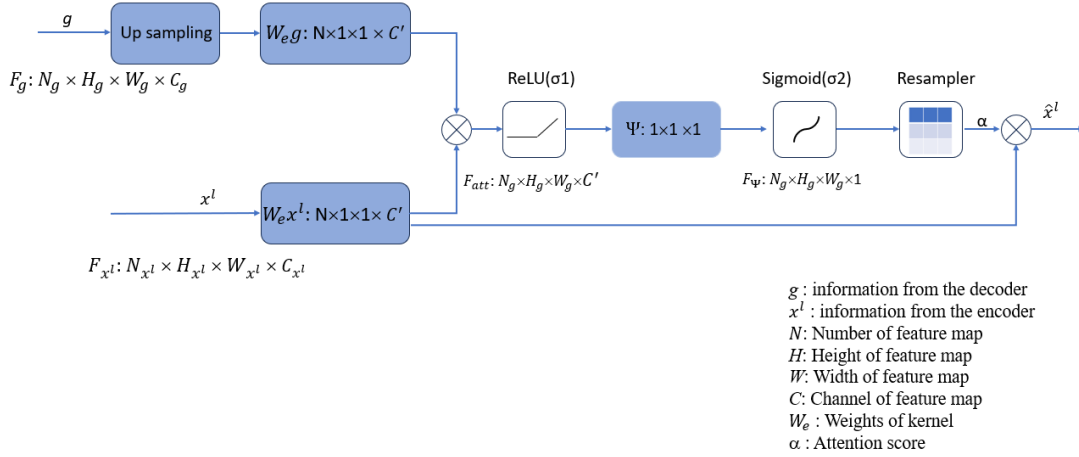


**Figure 6.2 The architecture of CAU-Net with attention block.**

A novel attention module is incorporated into the traditional U-Net to selectively emphasize specific features. This attention module is applied to the original skip connection path. The decoder in U-Net gradually restores the highly abstract semantic information to the original spatial dimension of the image through multiple up-sampling operations. However, this will inevitably lead to information loss, so skip connections are used to attach fine details to the up-sampled feature blocks at a deep level. The attention block's function is to attach a strong guidance to the up-sampled feature block to reweight the features in the skip connection, so that the network can

focus on those regions that are more relevant to the target structure.

As illustrated in Figure 6.3,  $x^l$  is the feature maps from the encoder, while  $g$  is the gate signal that carries information from the decoder. The extracted feature  $x^l$  consists of the current layer's feature map, which is concatenated with the upsampled feature map from the lower layer. This ensures that  $x^l$  retains both high-level semantic information and spatial details. The gate signal is  $g$  obtained by up sampling the feature maps from the current decoder layer, ensuring that it retains high-level semantic information. This gate signal is then used to guide the attention mechanism by selecting the most relevant spatial regions in  $x^l$ , allowing the model to focus on meaningful features while suppressing irrelevant ones.



**Figure 6.3 The structure of Customised Attention Block.**

Then  $g$  and  $x^l$  are converted to the same number of channels using two different  $1 \times 1$  convolutional layers for subsequent operations. The inner product between  $g$  and  $x^l$  is calculated, which can be simply understood as a measure of the similarity of vectors. If the directions of  $g$  and  $x^l$  are the same (or tend to be similar), the value of the dot product will be large, which means that the corresponding  $x^l$  should be assigned a higher weight. The ReLU function is followed by adding non-linearity to the attention module, which allows the attention mechanism to capture more complex relationships between features, not just simple linear mappings. Another  $1 \times 1 \times 1$  convolution is used to reduce the number of channels, which can reduce the number of parameters and

calculations of the model. On the other hand,  $1 \times 1 \times 1$  convolution allows information interaction between different channels, which means the multi-channel information is compressed into a more compact form to calculate the attention weight. The attention weight is normalised by the sigmoid function to have a value between 0 and 1. Then it will pass through a resampling module to align them with the dimensions of the input features  $g$  to ensure that the shape of the weights and input data matches so that they can be multiplied element by element. Finally, the resulting attention weights  $\alpha$  are element-wise multiplied with the original decoder feature map  $x^l$  to obtain a new feature map  $\hat{x}^l$ , thereby adjusting the importance of each location in the feature map. In this architecture, the attention gate is applied to the encoder feature map before it is fused with the decoder feature. This attention mechanism filters out irrelevant or low-response regions in the encoder output, allowing only the most informative features to be retained. The refined encoder feature map is then concatenated with the up-sampled decoder feature map, enabling the network to focus selectively on contextually relevant regions. This targeted feature fusion enhances the decoder's ability to reconstruct fine structures and contributes to improved segmentation accuracy.

As mentioned above, the proposed tumour segmentation method is a two-stage segmentation process, which requires first locating and extracting the liver part in the CT scan using the method proposed in Chapter 5. In addition, the coordinates of the liver extraction need to be recorded to facilitate the subsequent restoration step. In the second step, the cropped CT images are processed by the modified U-net to obtain the corresponding tumour prediction results, and the predicted images are restored to the size of the original image according to the cropping coordinates.

### **6.3 A novel end-to-end Cascade Hybrid Attention U-Net (CHAU-Net) for liver tumour segmentation**

#### **6.3.1 Overview of proposed end-to-end Cascade Hybrid Attention U-Net**

Another cascaded hybrid attention U-Net with different mechanisms is now described. In Section 6.2, the liver organs in the CT scans are classified and segmented with an original U-Net, images need further cropping to reduce the image size. The system is based on the fact that network can automatically determine the location of the liver and reduce the area of interest during training. Therefore, the following cascade network structure is designed to achieve end-to-end liver tumour segmentation. A hard attention module is designed into the cascade network to reduce the size of the image so that the network can pay more attention to liver part feature extraction. The hard attention block significantly reduces the size of the input image by locating the region of interest (ROI) and cropping out irrelevant background areas, thereby optimizing the model's computational efficiency and feature expression capabilities. In addition, a soft attention mechanism is employed in the second tumour segmentation block to enhance the detection of small lesions. This mechanism consists of spatial attention and channel attention and run through the entire feature extraction process. Channel attention emphasizes the modelling ability of global important features by assigning weights to feature channels, while spatial attention strengthens the localization ability of key areas by weighting spatial positions. Introducing soft attention in the feature selection stage effectively improves the network's ability to selectively focus on relevant features, thereby optimizing segmentation performance.

#### **6.3.2 Details of end-to-end cascade attention U-Net**

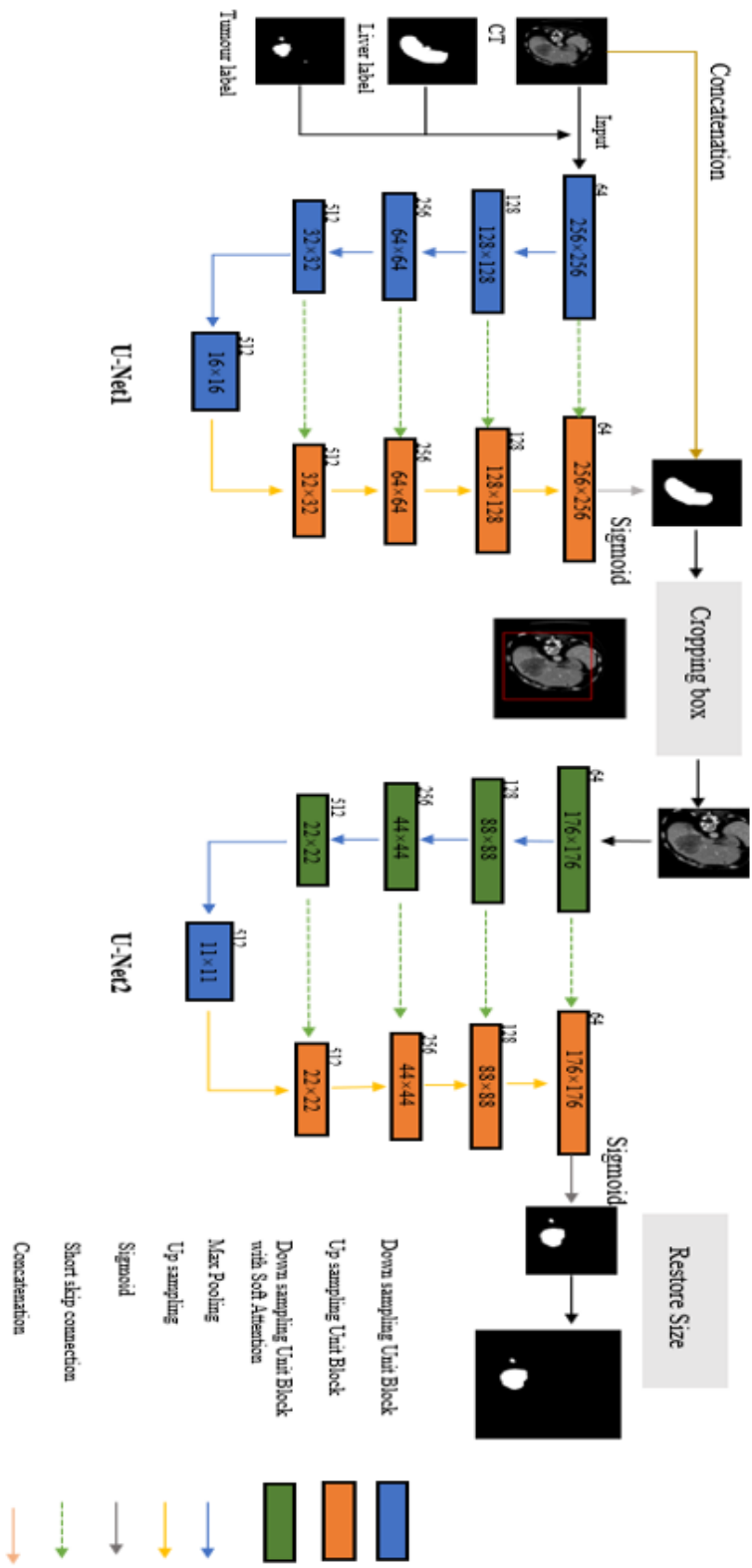
As illustrated in Figure 6.4, two basic U-Nets called U-Net1 and U-Net2 are used as the basic segmentation tools. A grayscale image with size  $256 \times 256 \times 1$  is used as the input to the whole network, corresponding to liver and tumour label as output. The image resolution provided by the public datasets used for research is usually  $512 \times 512$ . Due to

the limitation of computing power, all images are reduced to 1/4 of their original size to reduce the hardware requirements. Each U-Net has 4 down-sampling blocks, one bottle neck block and 4 up-sampling blocks with pre-processed CT scans. In the feature extraction path, 64,128,256 and 512  $3\times 3$  convolution kernels followed by ReLU functions are used, corresponding to 4 down-sampling paths. Then max pooling is used to halve the size of feature map and obtain the higher-level semantic features after each feature extraction path. In the bottleneck block, two  $3\times 3$  convolution layers are used to compute the most abstract information. In the up-sampling path, bilinear interpolation is used to double the size of features from the previous layer. The subsequent feature extraction steps are the same as in down-sampling path, except that the number of convolution kernels are 512,256,128 and 64, decreasing from deep layer to shallow layer. Short skip-connection connects the features in the corresponding down-sampling and up-sampling path, to obtain more contextual information in the process of up-sampling. In the last step, one  $1\times 1$  convolution kernel and sigmoid function are employed to generate the segmentation prediction of liver or tumour.

Excessive use of pooling inevitably leads to information loss. Moreover, tumour segmentation pays more attention to detailed information than liver segmentation due to relatively small size. Therefore, in the U-Net2, in the last two steps of the down sampling path, a convolution operation with stride 2 is used instead of the pooling operation to ensure the integrity of the information as much as possible.

As seen in Figure 6.4, the two U-Nets are concatenated sequentially with a hard attention block. The output of U-Net1 is concatenated with initial input passed through the hard attention block, to generate a smaller size CT image based on the liver segmentation result, including location and size, which is used as the input of U-Net2. Since the input image size of U-Net2 is reduced, the calculation cost of U-Net2 is relative lower than U-Net1. The details of hard attention block are described in the following part.





**Figure 6.4** The architecture of proposed modified cascaded attention U-Net.

The main contribution of this chapter is in the use of a hard attention block between two U-Nets, to extract the main area of liver through simple mathematical calculations without training. A liver prediction mask with values  $[0,1]$  is obtained from the last layer of U-Net1 using the sigmoid function, enabling direct determination of the liver location. The prediction mask passes through a thresholding function, so it contains only 0 or 1. Then this mask is multiplied by the initial input as in the following operation.

Define a row vector  $A \in \mathbb{R}^{1 \times N}$  with continuous real values ranging from 1 to  $N$ , where  $N$  equals the size of image. Define  $M \in \mathbb{R}^{N \times N}$  is the processed prediction mask from U-Net1, where  $M_n$  is the  $n^{th}$  row of  $M$ ,  $n \in [1, N]$ . The matrix  $O$  is the element-wise product of these, where the liver part gets assigned while the rest remains 0. For each row in  $O$ , the following holds:

$$O_n = A \cdot M_n \quad (6.3)$$

Then the left ( $w_x$ ) and right ( $w_y$ ) boundary of ROI (region of interest) in  $O$  can be obtained by determining the indexes of columns, where the global minimum and maximum values occur in  $O$ , respectively. Likewise, the same method is used to achieve the value of up and bottom boundary. The difference is to use the transpose matrix of  $A$  and multiply it with  $M$ , to get  $O'_n$ . The top ( $t_x$ ) and bottom ( $t_y$ ) boundaries can be obtained by determining the indexes of row where the global minimum and maximum values occur in  $O$ , respectively.

$$O'_n = A^T \cdot M_n \quad (6.4)$$

Then the centre point of ROI can be calculated with

$$c_{x,y} = \frac{1}{2}(w_y - w_x), \frac{1}{2}(t_y - t_x) \quad (6.5)$$

Let the size of the image to be cropped be  $a \times a$ , four boundary points are

$$w_{off} = c_x - \frac{1}{2}a, w_{end} = c_x + \frac{1}{2}a \quad (6.6)$$

$$h_{off} = c_y - \frac{1}{2}a, h_{end} = c_y + \frac{1}{2}a \quad (6.7)$$

To ensure the cutting point does not exceed the boundaries of the image, three points are set. First of all, the image size is defined as  $n \times n$ ;

- (1) The horizontal coordinate or vertical coordinate is less than 0.5a.
- (2) The horizontal coordinate or vertical coordinate is between 0.5a and n-0.5a.
- (3) The horizontal coordinate or vertical coordinate is bigger than n-0.5a.

The corresponding boundary points can be calculated as

$$w'_{off} = \begin{cases} 10, & \text{if } c_x < \frac{1}{2}a \\ c_x - \frac{1}{2}a \\ n - a, & \text{if } c_x > n - \frac{1}{2}a \end{cases} \quad (6.8)$$

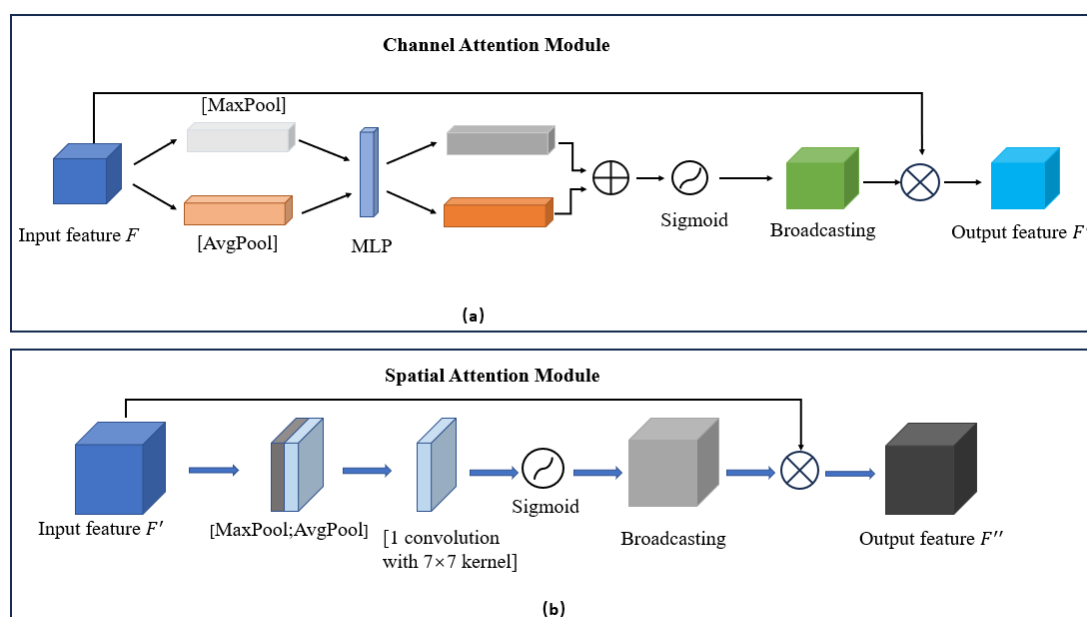
$$h'_{off} = \begin{cases} 10, & \text{if } c_y < \frac{1}{2}a \\ c_y - \frac{1}{2}a \\ n - a, & \text{if } c_y > n - \frac{1}{2}a \end{cases} \quad (6.9)$$

$$w'_{end} = w'_{off} + a \quad (6.10)$$

$$h'_{end} = h'_{off} + a \quad (6.11)$$

In this research,  $n$  is set to 256 and  $a$  is set to 176, so the cropped image as the input of the U-Net2 has size 176×176. According to the observation, the maximum size of the liver in the abdominal CT cross-section does not exceed 2/3 of the entire image, which

is why the size of the boundary box is set to  $176 \times 176$ . As the input size is reduced by  $2/3$ , the computation cost of U-Net2 is also reduced accordingly. However, the label of tumour still has size  $256 \times 256$ . As indicated on Fig 6.4(a), restore operation is needed after the last sigmoid layer. According to the boundary conditions of the cropping, padding the corresponding number of 0 around it can restore the prediction map to the initial input size. In this work,  $w_{off}$  zeros are padded to the left,  $80 - w_{off}$  zeros to the right,  $h_{off}$  zeros to the top and  $80 - h_{off}$  zeros to the bottom.



**Figure 6.5** The structure of channel attention module and spatial attention module respectively.

In addition, a soft attention mechanism is added to enhance the ability to detect small tumours. A combination of spatial and channel attention is applied to each feature block [207]. As shown in Figure 6.5(a), the input feature map  $F$  is processed through global max pooling (GMP) and global average pooling (GAP) along the spatial dimensions, generating two one-dimensional vectors with the same number of channels as the input features. These two vectors are then passed through a shared multi-layer perceptron (MLP), which consists of a dimension reduction layer, a ReLU activation function, and a dimension expansion layer. The outputs of the MLP are then combined using an

element-wise summation, followed by a sigmoid activation function to produce the final channel attention map. The channel attention mechanism assigns different importance weights to each channel, highlighting those that contain more meaningful features. The computed channel attention map is broadcasted along the spatial dimensions to match the shape of the input feature map and then multiplied element-wise with the input feature map, generating an enhanced feature representation, which serves as the input for the subsequent spatial attention module.

The spatial attention mechanism ensures that the network focuses on the most important regions of the image. As shown in Figure 6.5(b), the feature map  $F'$  after channel attention block is processed through the global max pooling and global average pooling along the channel axis to generate two weight maps with size  $\text{width} \times \text{height} \times 1$ . These two maps are concatenated along the channel dimension and passed through a  $7 \times 7$  convolution, which helps capture a broader spatial context. Finally, a sigmoid activation function is applied to produce the spatial attention map. The computed spatial attention map is then broadcasted along the channel dimensions and element-wise multiplied with the input feature map to enhance relevant spatial locations.

The spatial attention features are multiplied by the spatial attention block's input to access the final features. The combined spatial and channel attention block is applied only on each feature extraction path in U-Net2. It is a compromise between increasing accuracy and control the parameters of the network because a very precise liver segmentation result is not necessary. In the current tumor segmentation task, there is no need to achieve high-precision liver boundary segmentation, only liver localization is needed to provide spatial constraints for tumor detection. Liver segmentation is only used as a means to limit the image size rather than the main goal.

## 6.4 Visualisation and Statistical analysis of tumour segmentation results

### 6.4.1 CAU-Net and CHAU-Net Implementation for tumour segmentation

The data set used for training and testing is still the Liver Tumour Segmentation Challenge (LiTS), which comes from clinical sites around the world [208]. The data set contains CT scans of 130 patients. All scans have been provided in *nii* format with an axial size of 512\*512.

Unlike liver segmentation, where the target organ typically exhibits a well-defined anatomical structure across patients, liver tumours show substantial variability in location, size, morphology, and intensity. Furthermore, tumour voxels usually constitute only a small portion of a CT slice relative to the surrounding healthy liver tissue, leading to a severe class imbalance problem. In the absence of a public LiTS test set for tumour segmentation, the allocation scheme of public datasets needs to be carefully designed. Without such planning, arbitrary split could lead to unbalanced or unrepresentative test sets, thus compromising the fairness and reliability of the evaluation.

A total of 7,184 CT slices from 118 patients were used for training and testing. While the official LiTS dataset labels all 118 patients as having liver tumours, further inspection revealed that only 73 of them had tumours appearing in more than 15 axial slices. Given that each patient typically has around 150 liver slices, these 73 cases represent tumours with sufficient spatial extent for reliable segmentation. To ensure fair and meaningful evaluation, especially for metrics like global Dice score and Dice per case, the 45 patients with limited tumour presence were allocated exclusively to the training set. Including them in the test set would introduce a bias, potentially inflating global Dice while lowering Dice per case due to the sparse tumour distribution.

Additionally, to enhance generalization and robustness, the test set is constructed to include a 1:1:1 ratio of large tumour, medium tumour and small tumour cases. This

ensured that the model could be evaluated across diverse tumour types. As a result, the final dataset is split into 88 patients for training and 30 patients for testing.

The cascaded CAU-Net proposed in Section 6.2 is a two-stage segmentation process, which requires first locating and extracting the liver part in the CT scan. Therefore, the liver segmentation results from Chapter 5 are used as the baseline to extract the area containing the liver as the training set of the neural network. This step is performed simultaneously on the label. In addition, the coordinates of the liver extraction need to be recorded to facilitate the subsequent restoration step. In the second step, the cropped CT images are processed by the CAU-net to obtain the corresponding tumour prediction results, and the predicted images are restored to the size of the original image according to the cropping coordinates.

For the method CHAU-Net described in Section 6.3, since this is an end-to-end network, only one network needs to be trained, eliminating the need for additional data preprocessing as required by the before mentioned methods. However, considering the network's capacity and GPU memory limitations, all data are downscaled by half to  $256 \times 256$  to satisfy the computational requirements.

The training protocol for this model follows that of the modified U-Net described in Chapter 5, including the same experimental environment. However, the hardware configuration is upgraded to include two RTX 2080 GPUs, providing a total of 16 GB of video memory. In addition, the number of training epochs is increased to 150 to support more extensive learning. All other hyperparameter settings (including learning rate strategy and batch size) remain unchanged.

#### **6.4.2 Discussion with segmentation results**

In order to evaluate the effectiveness of the proposed methods, the experimental results are compared with several mainstream tumour segmentation methods, the specific comparison is shown in Table 6.1. It is obvious that the accuracy of tumour segmentation is far inferior to that of liver segmentation due to the variability of number

and shape. Nevertheless, the performance of the proposed two different attention-based models on this complex task still reaches a satisfactory level.

The Table 6.1 presents a comparative analysis of various models based on their performance in tumour segmentation tasks, as measured by several key metrics: Global Dice, Dice per Case, Volume Overlap Error (VOE), Relative Volume Difference (RVD), Average Symmetric Surface Distance (ASSD), and Maximum Surface Distance (MSD).

Global Dice and Dice per Case evaluate the overall segmentation accuracy of the models. The H-Dense U-Net [164] achieves the highest Global Dice score of 0.824 and a strong Dice per Case score of 0.722, indicating its superior ability to accurately segment tumours compared to other models. However, the Hybrid Attention model also performs very well, with a Global Dice of 0.798 and the highest Dice per Case of 0.762, suggesting that incorporating attention mechanisms can significantly enhance segmentation accuracy.

**Table 6.1 Comparison of proposed model with other seven tumour segmentation methods based on six measured metrics. The symbol “N/A” represents unreported result. Bold font represents the highest score on each measurement.**

Model	Tumour					
	<i>Global dice</i>	<i>Dice per case</i>	<i>VOE (%)</i>	<i>RVD (%)</i>	<i>ASSD (mm)</i>	<i>MSD (mm)</i>
Shape-parameter[209]	N/A	0.754	0.709	<b>0.124</b>	1.6	N/A
U-Net+Level set[202]	0.700	N/A	N/A	N/A	N/A	N/A
AHCnet[175]	0.591	0.574	1.507	0.329	1.462	7.538
2D-dense[210]	N/A	0.725	0.589	N/A	N/A	N/A
H-dense U-net[183]	<b>0.824</b>	0.722	N/A	N/A	N/A	N/A
CU-Net[211]	N/A	0.595	0.460	N/A	N/A	N/A
MCG-FRN[185]	0.764	0.674	0.324	0.194	4.408	<b>7.113</b>
CAU-Net	0.782	0.736	0.422	0.272	<b>0.898</b>	9.075
CHAU-Net	0.798	<b>0.762</b>	<b>0.395</b>	0.327	0.912	7.302

VOE measures the volumetric overlap error, where a lower value indicates better performance. The MCG-FRN [166] model performs best in this regard, with a VOE of 0.324%, closely followed by the Hybrid Attention model at 0.395%. These models also



show competitive performance in RVD, with MCG-FRN achieving 0.194% and the Hybrid Attention model achieving 0.327%, indicating their effectiveness in minimising volume discrepancies between predicted and true segmentations.

ASSD and MSD evaluate the boundary accuracy of the segmentation. Lower values indicate better boundary delineation. The customised attention model achieves the best ASSD of 0.998 mm, suggesting it is particularly effective in capturing detailed boundary information. However, the Hybrid Attention model shows a balance between both ASSD (0.912 mm) and MSD (7.302 mm), indicating a well-rounded performance in both overall accuracy and boundary precision.

Overall, the H-Dense U-Net demonstrates the highest overall segmentation accuracy, while the Hybrid Attention model shows strong, balanced performance across multiple metrics, highlighting the importance of attention mechanisms in enhancing segmentation results.

Compared with the Shape-parameter method, although this method performs better in terms of Dice per case coefficient, reaching 0.754, it is not effective in dealing with tumour boundaries, resulting in a higher VOE value. This shows that this method has certain limitations when dealing with tumours with complex shapes. Compared with the basic U-Net, although the method combining U-Net and Level Set can effectively handle some simple tumour segmentation tasks, it performs poorly when dealing with complex tumours (such as small tumours or tumours with blurred boundaries). This is mainly limited by the insensitivity of U-Net itself to tiny tumours. Although the use of Level Set can further segment the detected tumours, using only U-Net will still miss a large number of small tumour targets.

The H-Dense U-Net achieved state-of-the-art results in the LiTS Challenge, ranking first in lesion segmentation and delivering highly competitive performance in liver segmentation. This is due to its ability to effectively combine 2D and 3D convolutional neural networks to capture both intra-slice (2D) and inter-slice (3D) features. This hybrid approach takes full advantage of the strengths of both 2D and 3D convolutions, enabling

the network to efficiently extract detailed spatial features. However, the introduction of 3D convolution significantly increases the parameter amount and computational complexity of the network, and also requires more stringent hardware resources, which limits the training speed of the network to a certain extent. In contrast, the proposed hybrid attention method just requires an additional 10496 non-trainable parameters to build hard-attention block which result more focus area of interest without manually cropping operation.

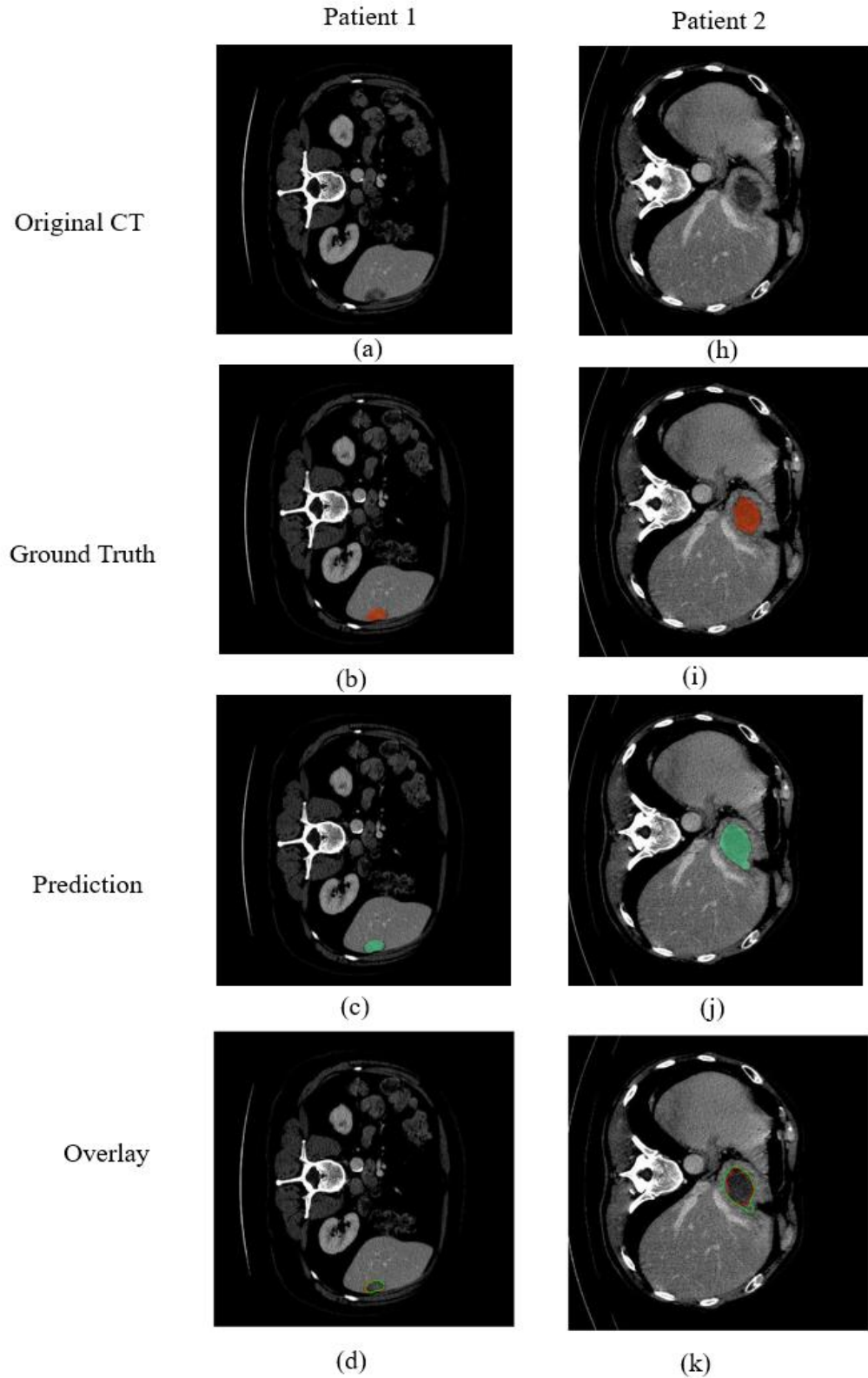
Most of the alternative mentioned networks discussed focus on the exchange of information and feature integration between layers within the network. In contrast, the proposed approach emphasizes feature selection. To enable the deep neural network to prioritize and identify key regions of the image before performing detailed segmentation, the attention mechanism is incorporated into the proposed model.

In the experiment, an in-depth comparative analysis was conducted between the proposed customised attention cascade U-Net and hybrid attention U-Net. Both have excellent overall performance, but they differ in specific scenarios: The cascade CAU-Net effectively locates the liver area through prior information and reduces unnecessary background areas by cropping CT images. This method is particularly suitable for processing large tumours because its segmented area is more concentrated and accurate. However, since the cropping operation depends on prior information, when the tumour location is complex or close to or even away from the liver boundary, part of the tumour area may be cropped, thus affecting the integrity of the segmentation. In addition, the CAU-Net has a high dependence on the quality of the input features, especially when extracting small tumour features, which may lead to incomplete input features due to information loss, thus limiting the module performance and even causing segmentation failure. The last disadvantage is that this is not an end-to-end network. From coarse segmentation to fine segmentation, manual intervention is required to achieve the goal of accurate tumour segmentation.

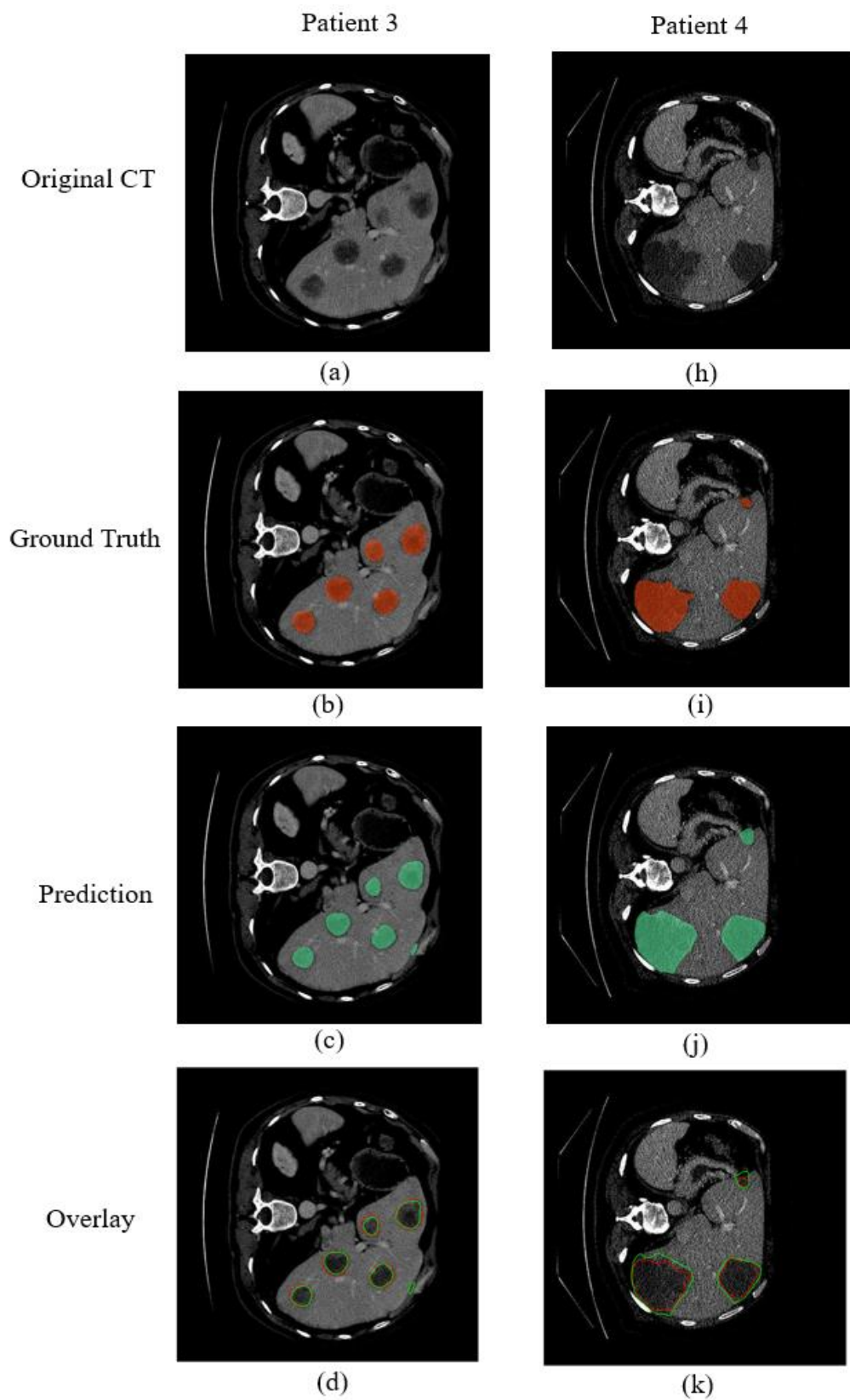
To overcome these limitations, a second CHAU-Net architecture is introduced, effectively combining hard attention and soft attention mechanisms to enhance performance on small and complex-shaped tumours. The soft attention module can dynamically adjust the model's attention focus, making the model more sensitive in capturing details and complex boundaries, thereby improving the detection rate and accuracy of small tumours. However, compared with the CAU-Net, the CHAU-Net has higher computational complexity and longer training time. In addition, when dealing with some simpler tumour segmentation tasks, the performance improvement is not significant, which may cause unnecessary waste of computing resources.

The following Figure 6.5 to Figure 6.8 show qualitative visualisation results of tumour segmentation from four patients with two proposed methods separately. Each figure displays the original CT slice, ground truth segmentation, model prediction, and the overlay of prediction and ground truth. The predicted tumor regions are shown in blue or green depending on the different algorithms, while the ground truth is consistently highlighted in red. Patients 1 and 2 both present cases with a single liver tumour. As shown in Figures 6.5 and 6.7, the model predictions of both methods closely match the annotated ground truth, with minor differences primarily along the tumor boundaries. These results demonstrate that the model performs robustly on solitary tumours with well-defined edges, achieving nearly perfect overlap in most slices.

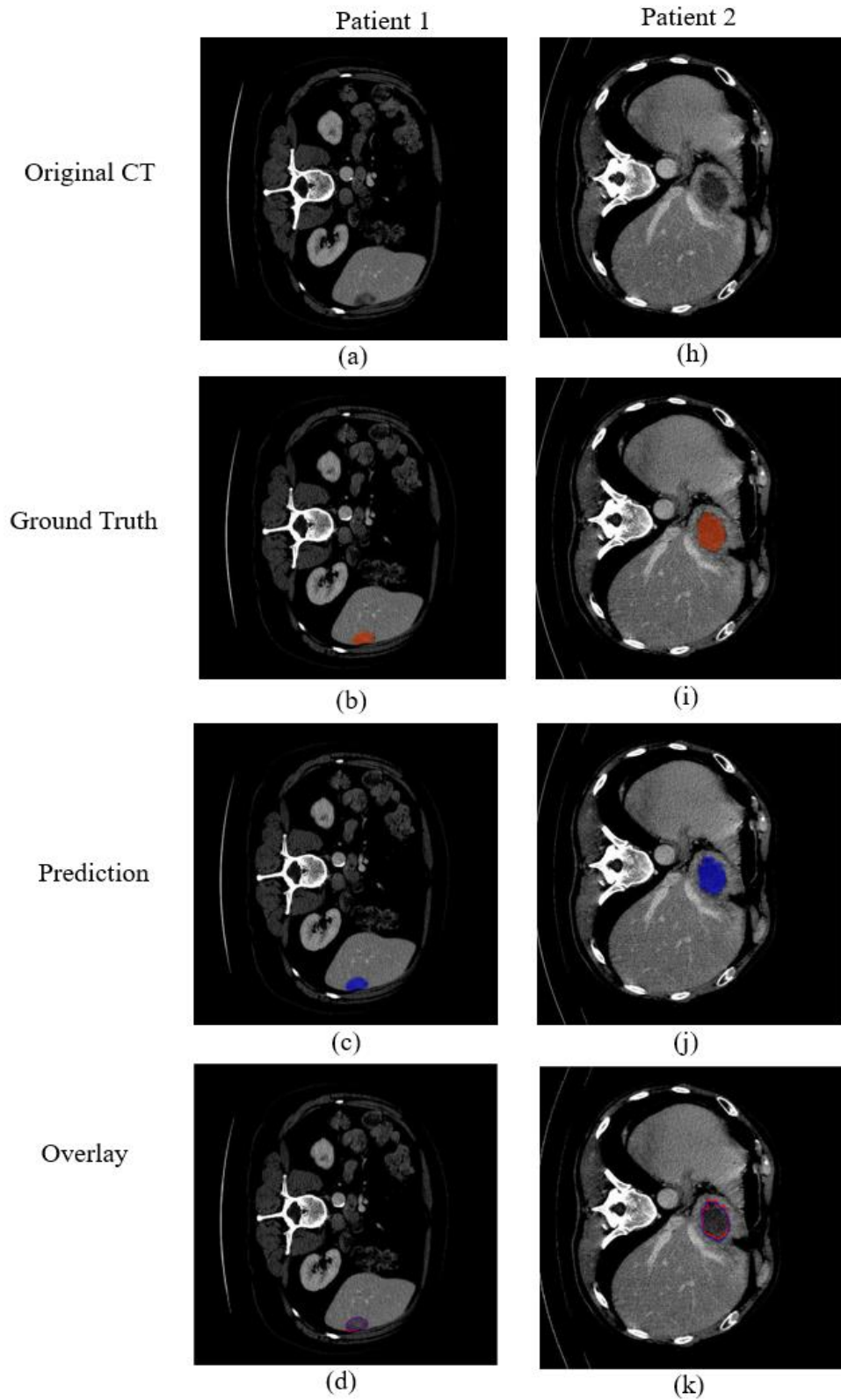
Patients 3 and 4 both present with multiple liver tumours of varying sizes and shapes, offering a more complex segmentation challenge. The visualisations in Figures 6.6 and 6.8 demonstrate that both models are capable of accurately segmenting most lesions, including small nodules and larger masses. Despite minor mismatches at the boundaries of some smaller lesions, the overall agreement between prediction and ground truth remains high. The overlay images confirm the models' effectiveness in managing tumor distributions, suggesting strong generalization to multi-lesion cases with diverse morphological characteristics.



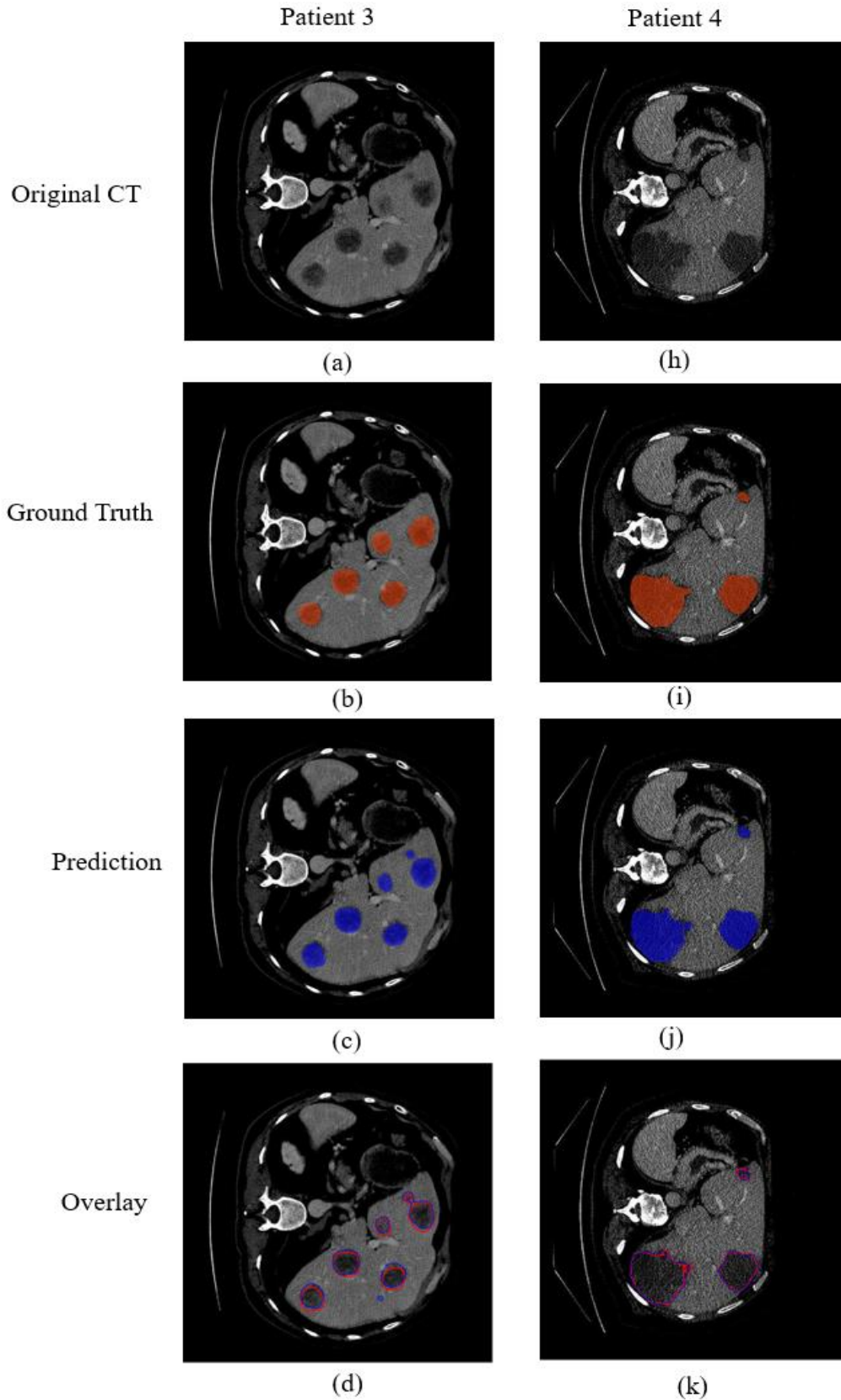
**Figure 6.6 Visualization of tumour segmentation results using the proposed CAU-Net method across patients with small and medium tumours. The first row shows the original CT slices, followed by the ground truth annotations (in red), model predictions (in green), and the overlay of prediction and ground truth.**



**Figure 6.7** Visualization of tumour segmentation results using the proposed CAU-Net method across patients with multiple medium tumours and large tumours. Overlay images confirm high spatial correspondence, with minor discrepancies on smaller lesions.

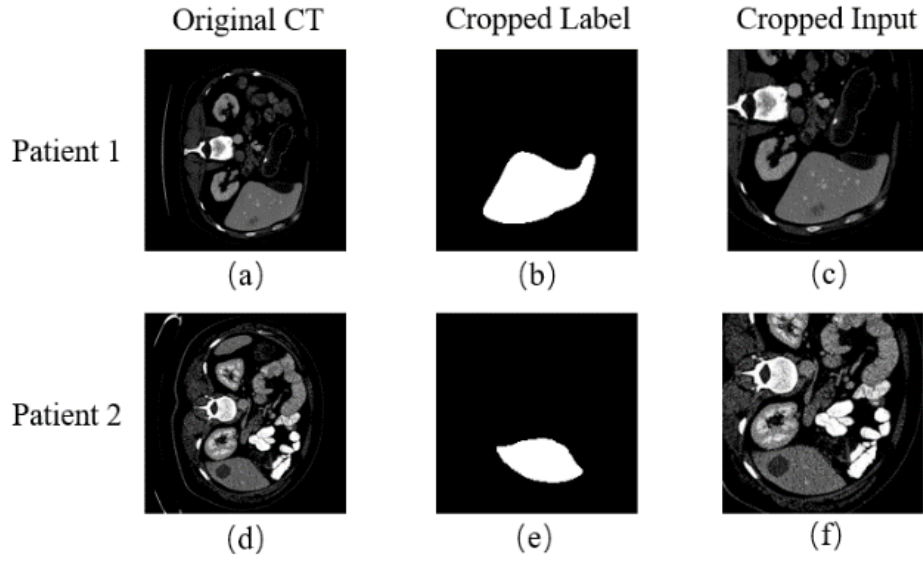


**Figure 6.8 Visualization of tumour segmentation results using the proposed CHAU-Net method across patients with small and medium tumours. The first row shows the original CT slices, followed by the ground truth annotations (in red), model predictions (in blue), and the overlay of prediction and ground truth.**



**Figure 6.9** Visualisation of tumour segmentation results using the proposed CAU-Net method across patients with multiple medium tumours and large tumours. Overlay visualizations show good alignment for multiple and medium-sized tumours, with slight bound.





**Figure 6.10 Visualisation results after hard attention stage. The second column is the cropped liver label, and the third column is the corresponding cropped input image.**

Figure 6.9 shows some visualisation results after cropping box block. It can be clearly seen that the embedded hard attention module can successfully extract the liver region and cut the corresponding region on the original input image as the input of the tumour segmentation network. Moreover, the tumour segmentation made on the cropped image can be correctly mapped to the original image without position shift.

## 6.5 Conclusion

This chapter explores advanced methods for liver tumour segmentation from CT images using modified U-Net architectures incorporating attention mechanisms. Tumour segmentation presents unique challenges due to the diversity in tumour shapes, sizes, and boundaries, as well as the inherent class imbalance in the dataset. The two approaches called customised attention cascade U-Net and hybrid attention U-Net are developed to address these challenges effectively.

The customised attention cascade U-Net leverages a two-stage segmentation process, where the liver is first segmented to narrow down the region of interest, followed by detailed tumour segmentation. This method excels in segmenting large



tumours by concentrating on relevant areas and reducing background noise through image cropping. However, its reliance on prior information for cropping can lead to incomplete segmentation in cases where the tumour is located near or outside the liver boundary.

To overcome the limitations of the cascade approach, the hybrid attention U-Net is designed, which combines hard and soft attention mechanisms. This model dynamically adjusts its focus, improving segmentation accuracy for small and complex-shaped tumours. While the hybrid attention U-Net demonstrates superior performance in handling small tumours, it also introduces higher computational complexity and longer training times.

Comparative analysis with other state-of-the-art methods revealed that both proposed models achieve competitive results, with the hybrid attention U-Net offering a balanced performance across multiple metrics, particularly excelling in boundary delineation. Visualisation of the segmentation results confirmed the models' ability to accurately segment tumours of various sizes, with good agreement between predictions and ground truth.

Although the proposed tumor segmentation method performs well in terms of accuracy, its results are not directly used for final clinical decision-making. After the automatic segmentation is completed, the radiologist still needs to review and optimize the tumor boundaries, if necessary, especially in complex cases where the lesions are small, low in contrast, or located near ambiguous areas of anatomical structures. This interactive process involving doctors not only improves the reliability of diagnosis, but also generates high-quality correction annotations for subsequent model fine-tuning or retraining, thus forming an effective user feedback mechanism. This closed-loop approach of human-computer collaboration helps to improve the efficiency of clinical workflows while enhancing the generalization and robustness of the model under different anatomical structures and imaging conditions.

# **Chapter 7. A Boundary Optimisation Scheme for Liver Tumours from CT Images**

## **7.1 Introduction**

In Chapter 6, two tumour segmentation networks were introduced based on different attention mechanisms and can achieve acceptable results. To further improve the segmentation performance, this chapter introduces an optimised tumour segmentation workflow that improves the accuracy and reliability of segmentation results.

In Section 7.2, a segmentation optimisation algorithm based on tumour boundaries is proposed and discussed. After the initial segmentation, the workflow advances by extracting image and mask patches based on the rough results. These patches are specifically targeted at the tumour boundaries due to the irregular shape of the tumour and the complex anatomical structures. Next, an optimisation network dedicated to refining the segmentation of these boundary patches is introduced. The network adopts a multi-level information fusion strategy to enhance the details and accuracy of the segmentation. By integrating features at different scales with residual structure and multi-resolution feature fusion, the network is designed to produce highly refined segmentation results for tumour boundaries. Finally, the optimised boundaries are mapped back to the original image, ensuring that the enhanced segmentation results are accurately aligned with the actual tumour location, which is critical to maintain the integrity of the spatial information.

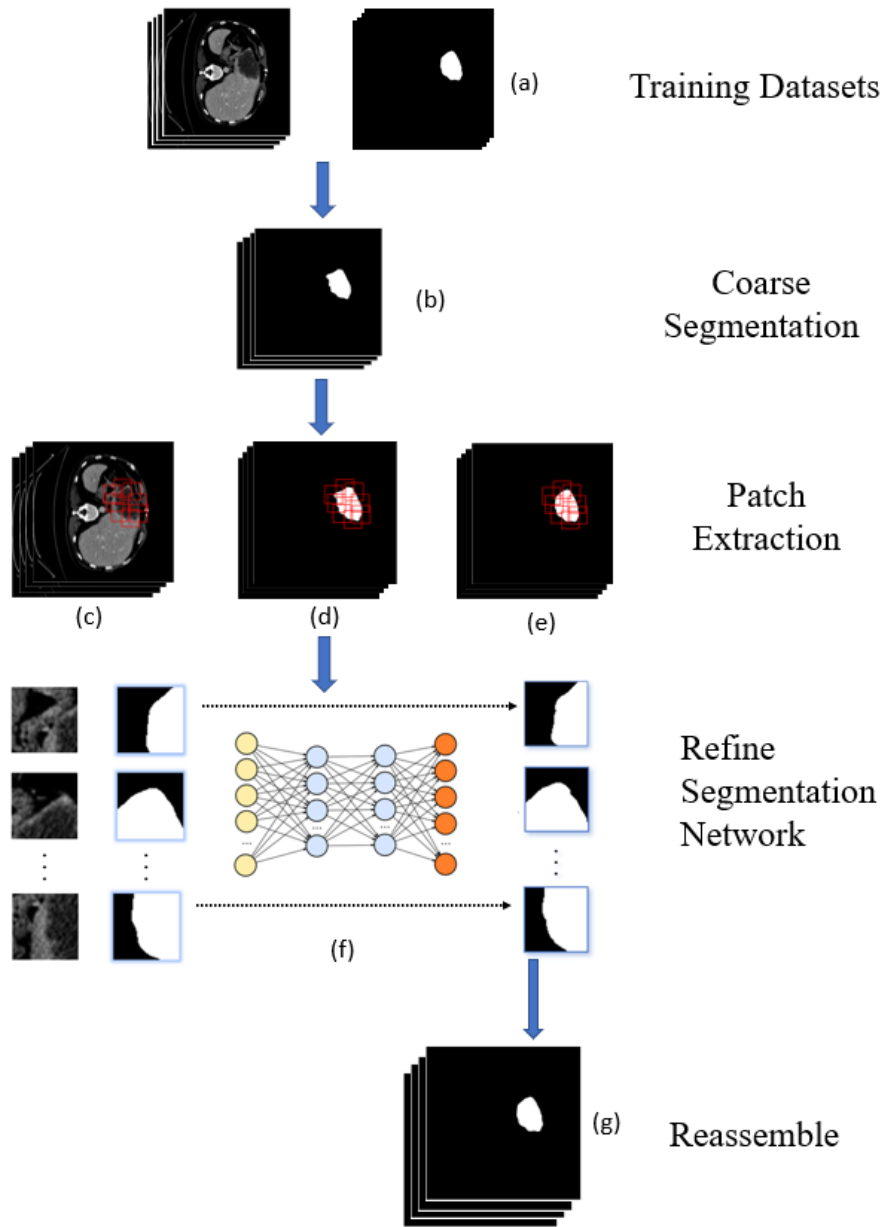
In Section 7.3, the statistical analysis and visualisation of segmentation performance is discussed to demonstrate the effectiveness of the proposed optimisation methodology. A brief summary is given in Section 7.4, illustrating the benefits provided by the adopted optimisation approach in the context of liver tumour segmentation. Through this optimised workflow, the goal is to achieve a level of segmentation accuracy that supports clinical decision making and ultimately improves patient

outcomes in oncology treatment.

## **7.2 Boundary Optimisation Process for liver tumour scans**

### **7.2.1 Overview of Proposed Optimisation Workflow**

The proposed segmentation optimisation workflow shown in Figure 7.1 consists of four major steps. The first step is to obtain all coarse segmentation results for all images in the training set. While the coarse segmentation provides a general location of the regions of interest, it may lack precision, especially at the boundaries of the segmented areas. This coarse result is crucial as it guides the subsequent refinement process. The second step is to extract all image patches and mask patches according to the coarse prediction result. The extraction of these focused patches allows the network to concentrate computational resources on refining the segmentation specifically where it is needed the most, improving the accuracy of the boundary delineation. The third step is an optimised refined segmentation network for boundary patch segmentation. By focusing on smaller and detailed patches of the image, the network can learn to correct the inaccuracies in the coarse segmentation, particularly around the edges of the segmented regions. The final step is to restore optimised boundaries to the corresponding position of the image. After the refined network has processed all the patches, the optimised boundary segments are stitched back together, aligning them with their original positions in the full-sized image. This reassembly step produces a final segmentation output that retains the benefits of both the coarse segmentation's global context and the refined segmentation's local accuracy.



**Figure 7.1** Proposed optimisation workflow, (a) liver CT image and corresponding ground truth, (b) extracted coarse segmentation (c) image and (d) mask boundary patches from the coarse segmentation (e) mask boundary patches from the ground truth (f) all patches pass the boundary refinement network (g) final refined segmentation result.

### 7.2.2 Boundary Patch Extraction

Expert radiologists' complete segmentation tasks, by first finding the target area and then carefully depicting the boundary. Segmentation results provide location information to determine the distribution of target regions. Boundary patches are then obtained on these target areas for subsequent optimisation. Initially, the original training

datasets consist of pairs of medical images and their corresponding ground truths. All images are used in the training data set through a deep segmentation network such as cascaded U-Net to obtain a coarse segmentation result. The coarse segmentation step generates a preliminary mask of the tumour regions. This step is crucial as it offers a baseline segmentation that captures the approximate location and shape of the tumour, which also simulates the coarse segmentation results in actual applications.

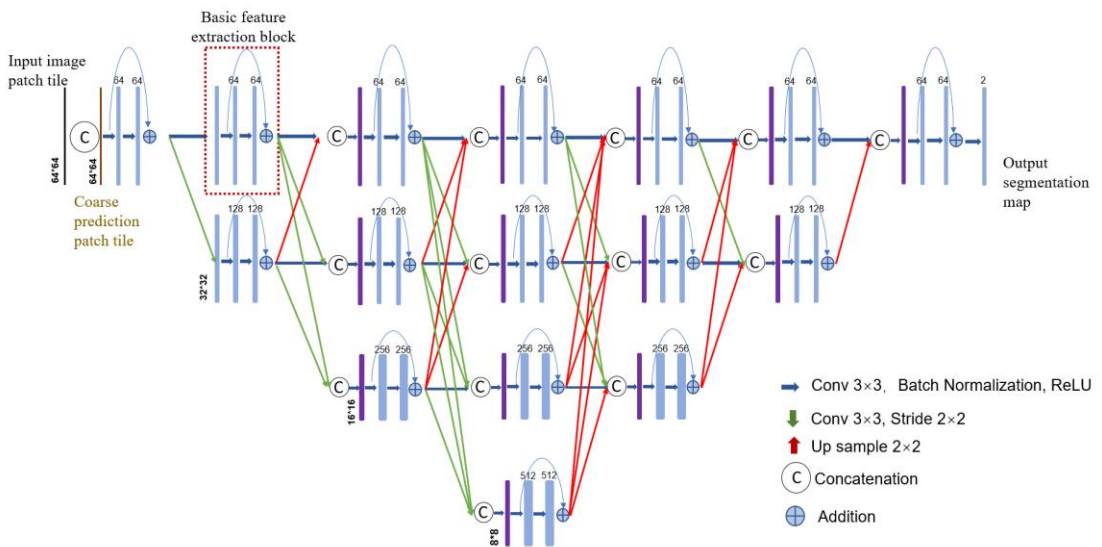
The Dice coefficient at the coarse segmentation stage is critical to supporting the effectiveness of subsequent fine segmentation. A sufficiently high Dice score ensures that most tumor regions are correctly located, providing a reliable basis for extracting boundary blocks and performing targeted refinement. If the coarse segmentation result deviates significantly from the true value (mis-segmentation or severe under-segmentation), the selected optimized region may not contain the true tumor region. In this case, even applying complex refinement strategies to non-tumor regions is unlikely to produce meaningful improvements. To reduce this risk, the improved U-Net architecture proposed in Chapter 6 is used, which combines an attention mechanism and a context enhancement module to improve the reliability of the initial coarse segmentation. However, given the limitations in detecting small tumours, especially those with sizes below 3cm in the early cancer stage, manual intervention is required. For these undetected regions, a small  $5 \times 5$  pixel block is manually marked in the image to assume that the tumour is present there, which is crucial to ensure the richness of the training set. Consequently, the dataset has been expanded from a pair of image and ground truth set, to an image set, a ground truth set, and a coarse segmentation set as in Figure 7.1(a) and (b).

Following the coarse segmentation, the predicted masks are processed as binary images, where tumour boundaries are explicitly defined. Boundary patches, as shown in Figure 7.1(d), are then extracted by sliding a window along the tumour's perimeter. The centre of each patch is positioned on a pixel belonging to the tumour boundary. This extraction is performed not only on the coarse segmentation output but also on the

original image and the corresponding ground truth (Figure 7.1 c and e), thereby creating a rich dataset focused on boundary information. The size of each boundary patch is set to 1/8 of the original image size to balance detail capture and computational efficiency. However, this patch size can introduce significant redundancy, which would increase the computational load without providing additional useful information. To mitigate this, the patches are extracted using a sliding stride of 2/3 the patch length. This approach reduces overlaps between adjacent patches, thereby minimising redundancy and computational cost.

### 7.2.3 Boundary Refinement Network (BRN)

For achieving precise segmentation, a multi-level information fusion boundary refinement network (BRN) is implemented, as visually detailed in Figure 7.2. This network is designed as a two-input, single-output architecture where the original image and the coarse segmentation map are concatenated and fed into the network. The purpose behind this concatenation is to impose a robust constraint on the input image, thereby directing the network's focus towards the boundary regions. By emphasizing these boundary areas, the network is able to accelerate the convergence process, ensuring it converges more rapidly and accurately towards the right direction.



**Figure 7.2 The architecture of proposed boundary refinement segmentation network.**

The BRN shown in Figure 7.1(f) and in more detail in Figure 7.2 comprises a feature extraction module, several down-sampling, up-sampling, and concatenation. As shown in the red dotted box in Figure 7.2, the basic feature extraction is processed with four  $3\times 3$  convolutions with batch normalization and ReLU activation function. The size of convolution kernels increases from 64 to 512 in a double increment with the number of down-sampling. This approach allows the network to capture increasingly abstract and complex features as the spatial dimensions of the input decrease. As the size of input images has significantly decreased compared to the original images, a convolution operation with a stride of 2 is used to accomplish the down-sampling rather than pooling. Although it will inevitably increase some computational costs, more information can be retained in the case of less input information, which is crucial given the reduced size of the input images compared to the original images.

Each feature extraction block is passed with residual structure to avoid gradient vanishing caused by the increase of network depth. We set 7,5,3,1 feature extraction block in four different resolutions respectively. As the number of down sampling increases, more abstract features and information are extracted with a smaller feature size as shown in Figure 7.2. Fully fusing information in different layers enables the model to learn more complex patterns and better capture the correlation of data. When the features received by each basic feature extractor come from different dimensions, the information is fused together through a cascade operation as shown in the purple bar in Figure 7.2 to extract new features. The previous feature map is adjusted to an appropriate size using a  $1\times 1$  convolution, down sampling (convolution with a stride size of 2) or up sampling (bilinear interpolation) to satisfy the concatenation requirements.

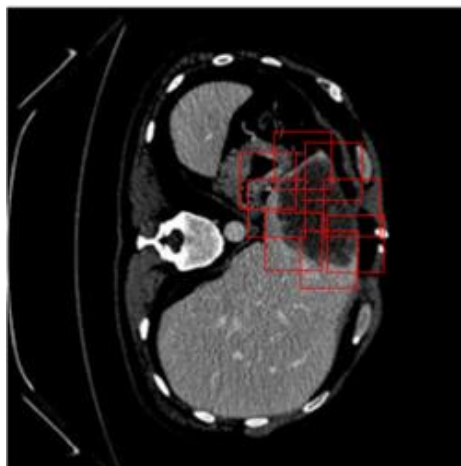
More fine-grained information is obtained in the low resolution by using multiple feature extractors which are used in multi-level information fusion. The fine-grained information in the shallow layer reduces the resolution by down sampling, while abstracted information in deep layer expands the feature size by interpolation. Features

in different dimensions are fused together by concatenation for information interaction to obtain richer semantic information and precise location information.

#### **7.2.4 Reassembling Process**

In the initial data preparing phase, boundary patches are extracted along the contours of the coarse segmentation results, and their spatial coordinates are recorded, as illustrated in Figure 7.3, which shows 8 overlapping patches. These patches, along with corresponding coarse segmentation patches, are fed into the trained refined segmentation network to produce accurate local predictions. After refinement, each patch is mapped back to its original location according to its recorded boundary coordinates, reconstructing the entire prediction map.

Due to the boundary patch extraction strategy, there is inevitably partial overlap between adjacent image patches. In the overlapping area, a majority voting scheme is used to determine the final label of each pixel. Specifically, for each overlapping pixel, the number of times it is classified as a tumor across all contributing patches is counted. If the majority of the predictions label the pixel as a tumor, it is classified as a tumor in the final segmentation map; otherwise, it is marked as background. This approach ensures consistency and robustness in handling overlapping predictions, especially in boundary areas.



**Figure 7.3 Patches are captured with a step size of 40 on the tumour boundaries.**



### 7.3 Visualisation and Statistical analysis of segmentation results

As described in Section 7.2.2, the boundary optimisation scheme requires the reconstruction of the training set and test set based on the existing coarse segmentation results. Each dataset should contain the initial image patch, the coarse segmentation patch, and its corresponding true label patch. Therefore, the liver segmentation results in Chapter 5 are used and the tumour segmentation results in Chapter 6 as coarse segmentation predictions to generate a new dataset containing hundreds of thousands of patches, where the patch size is  $64 \times 64$ . Considering the high similarity between adjacent CT images in the liver dataset, the data were screened to reduce redundancy and improve computational efficiency. For large liver areas that appear continuously, the strategy of selecting every other image is adopted. This method significantly reduces the data size while retaining key information. And the tumour dataset maintains the same number as Chapter 6, containing 7184 images. The training set consists of 90,064 liver patches and 30,183 tumour patches, which are used to train the refinement segmentation network.

The Boundary Refinement Net is implemented in Anaconda with Python, running on PC with 32G RAM, 3.8GHz AMD Ryzen7 3800X 8-core CPU, and NVIDIA RTX3090Ti GPU with total 24GB memory. The proposed Boundary Refinement Net is trained for 100 epochs using the Adam optimiser, and the learning rate is gradually reduced from  $1e-4$  to  $1e-5$  when the accuracy did not change within 5 epochs. The batch size is set to 16 based on the increase in available video memory and smaller image size to fully utilise hardware resources and improve training efficiency.

Table 7.1 shows the key metrics of ten models in tumour segmentation tasks on LiTS datasets. From the perspective of the global Dice coefficient, X-Net performed the best, reaching 0.843, which shows that its overall performance in the tumour segmentation task is excellent. The 2.5DnnU-Net also achieved a high score of 0.814, performing well in balancing computational efficiency and accuracy. The proposed optimisation algorithm also achieved an excellent score of 0.805. In terms of the three

indicators of VOE, RVD, and ASSD, our algorithm achieved the best results because the boundary of the predicted image is closer to the true value by growing or shrinking. While X-Net also performed well in these indicators.

**Table 7.1 Comparison of proposed model with other seven tumour segmentation methods based on six measured metrics. The symbol “N/A” represents unreported result. Bold font represents the highest score on each measurement.**

Model	Tumour					
	<i>Global dice</i>	<i>Dice per case</i>	<i>VOE (%)</i>	<i>RVD (%)</i>	<i>ASSD (mm)</i>	<i>MSD (mm)</i>
Mask-RCNN[212]	0.71	N/A	0.476	0.43	4.1	N/A
X-Net[184]	<b>0.843</b>	<b>0.764</b>	0.357	0.369	0.969	<b>5.407</b>
T3scGAN[190]	0.796	N/A	N/A	N/A	N/A	N/A
2.5DnnU-Net[213]	0.814	N/A	N/A	N/A	N/A	N/A
H-dense U-net[183]	0.824	0.722	N/A	N/A	N/A	N/A
FCN+PP[214]	0.796	0.676	N/A	N/A	N/A	N/A
2.5D P-U-Net[215]	0.735	N/A	N/A	N/A	N/A	N/A
Customised attention	0.782	0.736	0.422	0.272	0.898	9.075
Hybrid attention	0.798	0.762	0.395	0.327	0.912	7.302
Refinement result	0.805	N/A	<b>0.325</b>	<b>0.191</b>	<b>0.841</b>	7.359

The tumour segmentation algorithm based on Mask-RCNN [212] first uses the Mask R-CNN model to segment the liver part in the CT image. The secondary processing of the obtained liver image includes using power law transformation to enhance pixel differences and median filtering to remove noise. Then the MSER method is used to detect the tumour part from the processed liver image. MSER is a region detection algorithm based on image brightness changes, which can extract stable regions from CT images. Compared with deep neural networks, it can only detect areas with stable brightness and cannot capture complex high-level features such as edges, textures, contextual information. The morphological and texture characteristics of tumours are complex and diverse, and there may be brightness differences in CT images of different patients, while MSER is less sensitive to these details.

T3scGAN proposed by Liu et al. [190] is a 3D liver and tumour segmentation model built based on conditional generative adversarial network (cGAN). This method

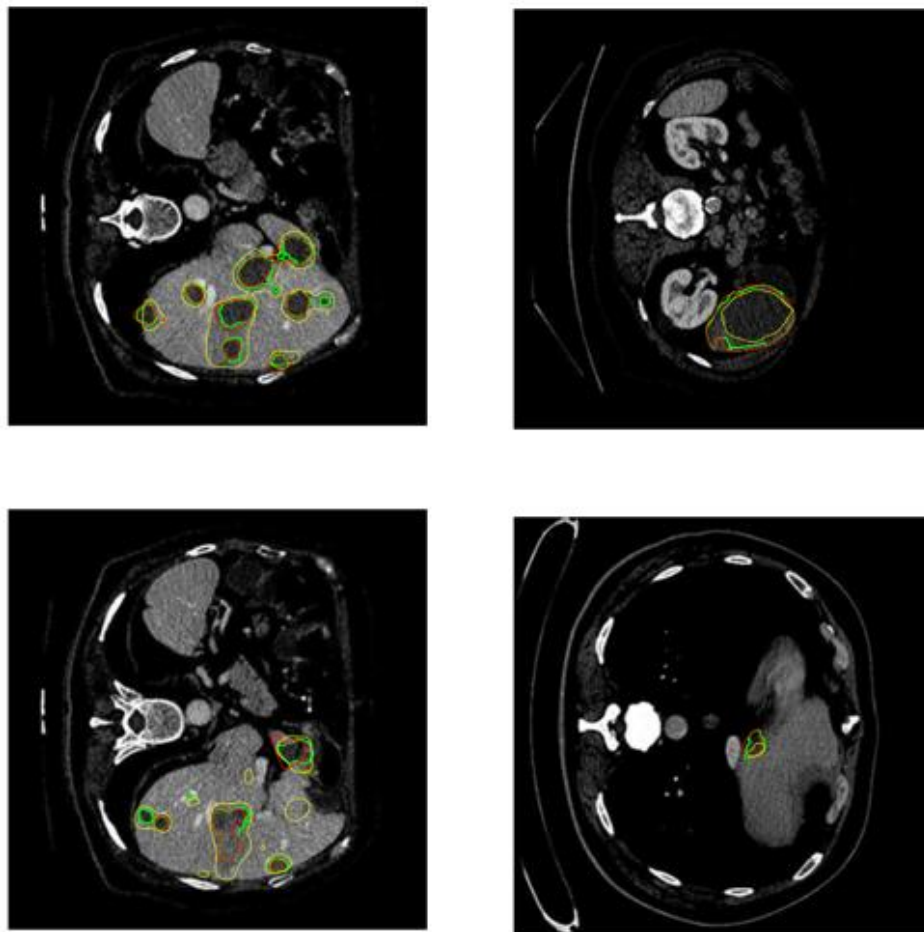
adopts a coarse-to-fine 3D automatic segmentation framework, from liver to tumour. The Dice coefficient of tumour segmentation of T3scGAN on the LiTS dataset was 0.796. The innovation of this method lies in the loss function mechanism of cGAN, which allows the network to perform adversarial training between the generator and the discriminator to optimise the segmentation accuracy.

The innovation of 2.5D nnU-Net [213] is that it uses a 2.5D convolutional network, which can capture 3D information at the same time by utilising the efficiency of 2D convolution, thereby improving the segmentation effect. In addition, it combines the Dense-Sparse sampling strategy and the depth-separable convolution to maintain high-precision tumour segmentation while reducing the number of model parameters. A Dice coefficient of 81.5% was achieved for tumour segmentation based on the LiTS dataset. A similar 2.5D Perpendicular U-Net [215] fuses three 2.5D Res-UNets in the coronal, sagittal, and axial planes, and then fuses their segmentation results. This multi-view combination can effectively improve the receptive field of the model, thereby enhancing the segmentation performance of the model. A Dice coefficient of 0.735 was achieved for tumour segmentation based on the LiTS dataset. Compared with the 3D model, the 2.5D P-U-Net has smaller parameters and memory usage, and is more efficient in training and inference, making it suitable for clinical application scenarios with limited hardware resources.

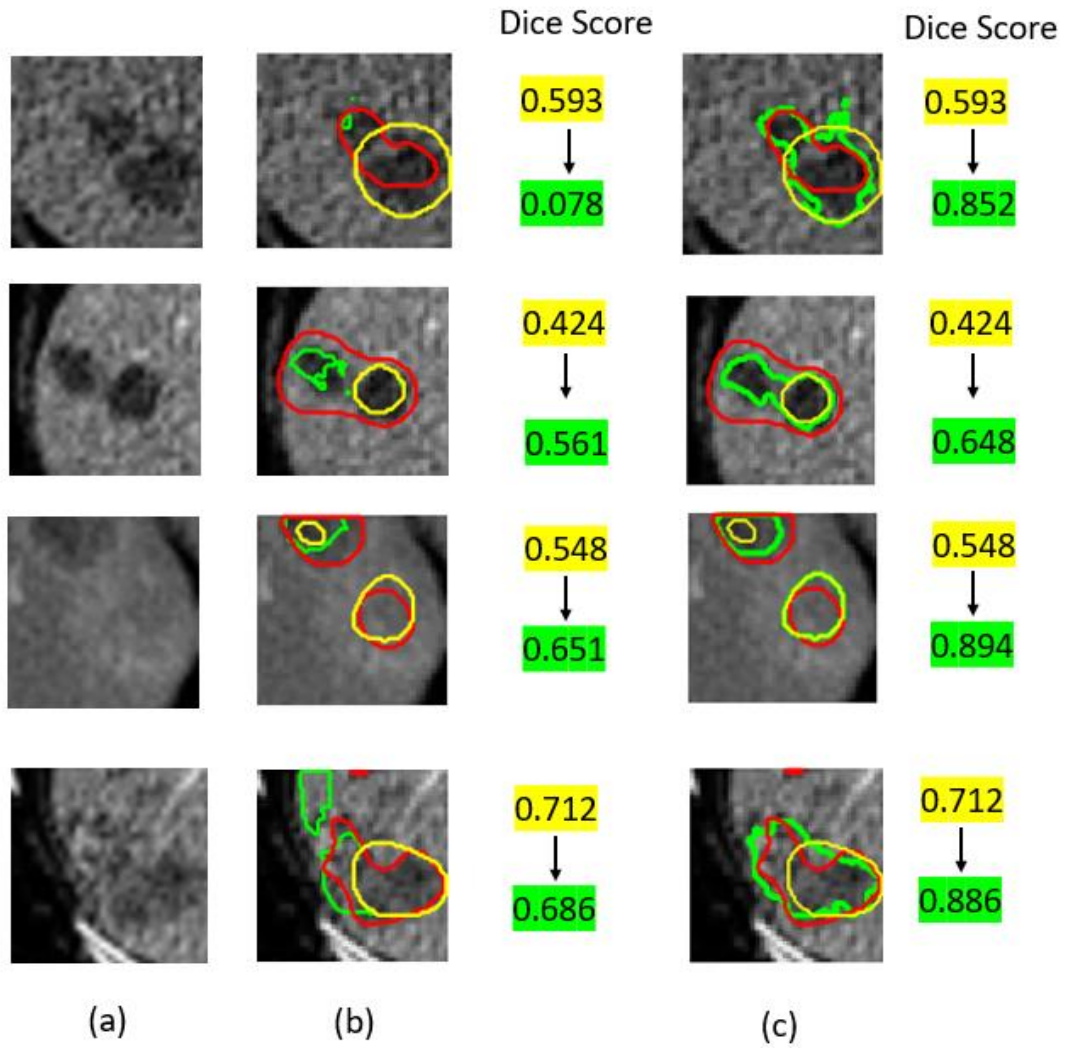
The fully convolutional network proposed by Grzegorz et al. [214] combined with an object-level post-processing strategy achieved a dice score of 0.796. Segmentation results generated solely by a fully convolutional network (FCN) may produce more false positives, so a post-processing step is introduced to further screen and classify these segmented tumour objects. Tumour objects are determined based on the 3D connected components output by the FCN and then classified as true positive or false positive using a random forest classifier containing 36 manually designed features, which cover information such as the shape, location, and distance from the liver boundary of the tumour. In contrast, the proposed boundary optimisation network

focuses on the precise correction of tumour boundaries through a multi-level information fusion network. The optimisation processing object in FCN is the classification of tumour candidate regions, while the boundary optimisation method deals with the refinement of tumour boundaries, which makes the final segmentation result closer to the real boundary.

Compared with the tumour segmentation algorithm proposed in Chapter 6, the proposed method shows significant improvement in multiple indicators, especially in the three evaluation indicators of VOE, RVD and ASSD. This shows that our proposed optimisation scheme is significantly effective in improving segmentation accuracy and boundary processing capabilities.



**Figure 7.4 Visualisation results of four different patients. Red line presents the ground truth of tumour boundary, yellow line present segmentation results need to be refined, green line presents refined segmentation boundary.**



**Figure 7.5 Comparison of boundary optimisation segmentation network with and without mask patch. Red: Ground Truth; Yellow: Coarse Segmentation; Green: Refined Segmentation. Yellow boxes show the Dice score between the coarse result and ground truth, while green boxes show the Dice score between the refined result and ground truth.**

shows some visualisations of tumour segmentation for four patients. The green line represents the optimised tumour boundary, the yellow line represents the coarsely segmented tumour boundary, and the red line represents the ground truth boundary. It can be seen that the proposed optimisation algorithm has a significant improvement in the segmentation results of small tumours. The optimisation algorithm can effectively shrink or expand the rough segmentation results to approach the real results. Figure 7.5 shows the effect of the mask patch on the optimisation algorithm. Column (a) displays the original CT images, column (b) shows segmentation results without using the mask

patch, and column (c) shows results with the mask patch. Red, yellow, and green contours represent the ground truth, coarse segmentation, and refined segmentation boundaries, respectively. The yellow boxes indicate Dice scores between the coarse segmentation and ground truth, while the green boxes represent Dice scores between the refined segmentation and ground truth.

When the model remains unchanged, the prediction results without the participation of the mask patch become very unreliable. For some small tumours, the model cannot obtain accurate prediction results or even completely wrong predictions, which are not as good as the results before optimisation. The use of the mask patch allows the network to focus more on the pixels near the area to be optimised, so that the optimisation result is closer to the real value. This improvement is clearly reflected in both the visual segmentation boundaries and the quantitative Dice Score gains shown in Figure 7.5.

## **7.4 Conclusion**

This chapter presents a new fully automatic method to optimise liver tumours segmentation results from CT scans, which can achieve end-to-end refinement for liver tumour segmentation. By leveraging image patches and mask patches derived from coarse segmentation, the proposed boundary refinement network enhances the model's attention to critical boundary areas, leading to improved segmentation accuracy. Unlike traditional U-shaped deep neural networks, the method employs additional feature extraction blocks in place of skip connections to maximise information extraction under limited input conditions. Experimental results on the liver tumour segmentation challenge demonstrate the effectiveness of the approach, achieving an average Dice score of 0.805 and a volume overlap error of 0.325. Furthermore, the reduced input image size contributes to alleviate the class imbalance problem, underscoring the method's potential for robust clinical application.

# Chapter 8. Conclusion

## 8.1 Conclusion

This thesis presents a comprehensive approach to the segmentation of liver and liver tumours in CT images, utilising advanced deep learning techniques to enhance the accuracy and efficiency of the segmentation process. The research aimed to address the critical need for precise identification and delineation of liver structures and tumours, which is essential for effective surgical planning and treatment in hepatocellular carcinoma (HCC) cases. Through the integration of advanced image processing techniques and deep learning frameworks, the research has successfully enhanced segmentation accuracy, robustness, and efficiency, providing valuable tools for clinical applications.

Chapter 2 reviews the literature related to the liver and liver diseases in detail, introduces the physiological structure and function of the liver, and common liver diseases, especially the epidemiology, causes, diagnosis and treatment of liver cancer. In addition, the basic principles of CT imaging and its application in the diagnosis of liver diseases are discussed, including common artefacts in CT imaging and their impact on image quality. Chapter 3 introduces image segmentation techniques based on CT data, covering classic image segmentation methods and modern methods based on deep learning. This chapter also discusses the role of preprocessing techniques and segmentation techniques in improving CT image quality and segmentation accuracy. Chapter 4 discusses CT datasets for liver and liver cancer segmentation, especially the composition and characteristics of the LiTS dataset. In addition, this chapter also introduces in detail the evaluation indicators of segmentation model performance, which provides a basis for the experiments and performance analysis in subsequent chapters.

The first key contribution, detailed in Chapter 5, proposed a modified Multi Scale Feature U-Net architecture designed for the automatic semantic segmentation of liver regions from CT scans. This modified MSF U-Net introduced dilated convolutions in

the deep layers of the U-Net, which effectively expand the receptive field, enabling the model to more sensitively capture subtle anatomical features and fuzzy boundaries without increasing the number of parameters. At the same time, the pyramid pooling module further aggregated multi-scale global features, significantly enhancing the model's ability to recognise complex liver anatomical structures. Experimental results demonstrated that the proposed MSF U-Net achieved an average dice score of 0.95 on the LiTS dataset, outperforming traditional U-Net models and exhibiting strong generalization even in challenging cases.

The second contribution in this thesis focuses on the tumour segmentation problem and proposed two modified cascaded U-Net models with the attention mechanism to address the challenges of tumour segmentation, especially when dealing with small and irregularly shaped tumours. The first tumour segmentation method reduces the interference of irrelevant areas by manually cropping the liver segmentation results. The introduction of a customised attention module in the CAU-Net can dynamically adjust the feature weight distribution during the decoding stage, thereby significantly improving the model's attention to the tumour area and improving the segmentation performance. The second proposed cascaded hybrid attention CHAU-Net implements an end-to-end training and inference process. The model introduces a hard attention mechanism for bounding box generation, which reduces the amount of unnecessary input data by dynamically adjusting the size of the image during the training phase. In the tumour segmentation phase, the encoder combines the channel attention module and the spatial attention module to enhance the feature extraction capability. The synergy of the two attention mechanisms enables the network to accurately focus on key tumour features while effectively reducing the interference of surrounding tissues, thereby improving the accuracy and robustness of segmentation. Both models achieved DSCs of 0.78 and 0.80, respectively, demonstrating better performance compared to baseline models.

In addition, this thesis also described a novel boundary refinement framework



aimed at optimizing liver and liver tumour segmentation results. This method extracts boundary patches from the coarse segmentation results and refines them using a boundary refinement network (BRN). Specifically, boundary patches are generated by sliding cropping along the contours of the coarse segmentation results, and then these boundary patches are jointly input into the boundary refinement network with the corresponding original image. By combining the semantic information of the original image and the structure of the coarse segmentation results, it aims to provide robust constraints for the input and guide the network to focus on key boundary areas. This can significantly enhance the network's perception of boundary features, accelerate the model's convergence process, and ensure that the segmentation results are more accurately optimised in the right direction. At the same time, the BRN network further enhances the ability to extract and fuse features by effectively utilising the information interaction between multiple layers. The boundary refinement framework improves boundary-related indicators by about 10%, providing clearer and more accurate segmentation contours for clinical use.

## **8.2 Future Work**

Although the liver and tumour segmentation methods proposed in this paper have made significant progress, there are still many possible research directions worth exploring to further improve the algorithm performance and promote its practical application.

First, the model can be trained with larger and more diverse datasets. The current method is mainly based on the LiTS dataset, which is limited in size and diversity. Future research can introduce multi-centre and multi-modal datasets, especially combining CT and MRI imaging data to capture a wider range of liver and tumour characteristics. This multi-modal integration helps to improve the segmentation model's ability to recognise complex pathological features.

Second, future research can explore the combination of 3D network architecture and Transformer technology to further improve segmentation performance. The current

work is mainly based on 2D segmentation strategies, which may have certain limitations when processing three-dimensional spatial information. The use of 3D networks can make full use of the volume information of CT images, thereby more accurately capturing the spatial structural characteristics of the liver and tumours. However, since 3D networks are usually accompanied by higher computational complexity, future research can consider introducing hierarchical or lightweight 3D network structures to balance performance and computing resource requirements. In addition, Transformer architecture provides a new perspective to handle medical image segmentation tasks. Compared with traditional convolutional neural networks, Transformers can capture long-range dependencies of global and local features through self-attention mechanisms and show higher flexibility and accuracy in complex scenarios. Combining the advantages of 3D networks and Transformers, embedding Transformer modules into 3D networks can achieve multi-scale feature fusion across levels and spaces, significantly enhancing the expressive power of the model.

Third, the proposed algorithm can be improved to learn from time-dependent data, such as CT or MRI scans of the same patient at different stages of treatment. This approach enables the model to capture the dynamic changes of the tumour, helping clinicians monitor the effectiveness of treatment and supporting personalised treatment planning. Segmentation algorithms combined with time series data are expected to play an important role in treatment evaluation and prediction.

Fourth, translating the proposed algorithms into practical applications requires the development of efficient and user-friendly clinical tools. Future research can integrate algorithms into software platforms, such as mobile applications or dedicated diagnostic systems, allowing clinicians to upload patients' CT scans, automatically run segmentation algorithms, and interactively visualise tumour areas. Such tools can accelerate the segmentation process, reduce the burden of manual operations, and improve diagnostic efficiency. Conducting usability studies with radiologists would help evaluate the reliability, interpretability, and practical value of such systems.

Integration with existing radiology platforms (such as PACS) and diagnostic workflows would be critical for adoption.

Finally, to ensure clinical deployment and patient safety, future efforts must consider regulatory compliance and interdisciplinary collaboration. Partnerships with radiologists, software engineers, and regulatory experts will be necessary to meet data protection standards, ensure system reliability, and navigate regulatory pathways. Early consideration of these aspects will help reduce translation gaps between algorithm design and clinical implementation.

## References

- [1] H. Sung *et al.*, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 71, no. 3, pp. 209-249, 2021.
- [2] P. H. Viale, "The American Cancer Society's facts & figures: 2020 edition," *Journal of the advanced practitioner in oncology*, vol. 11, no. 2, p. 135, 2020.
- [3] J. Albarrak and H. Al-Shamsi, "Current status of management of hepatocellular carcinoma in the Gulf region: challenges and recommendations," *Cancers*, vol. 15, no. 7, p. 2001, 2023.
- [4] R. S. Rates, "SEER cancer statistics review 1975-2004," *Bethesda, MD: NIH*, 2007.
- [5] A. Burton *et al.*, "Primary liver cancer in the UK: incidence, incidence-based mortality, and survival by subtype, sex, and nation," *JHEP Reports*, vol. 3, no. 2, p. 100232, 2021.
- [6] H. B. El-Serag and K. L. Rudolph, "Hepatocellular carcinoma: epidemiology and molecular carcinogenesis," *Gastroenterology*, vol. 132, no. 7, pp. 2557-2576, 2007.
- [7] J. M. Llovet *et al.*, "Hepatocellular carcinoma," *Nature reviews Disease primers*, vol. 7, no. 1, pp. 1-28, 2021.
- [8] C. P. Chen, "Role of radiotherapy in the treatment of hepatocellular carcinoma," *Journal of Clinical and Translational Hepatology*, vol. 7, no. 2, p. 183, 2019.
- [9] F. Pesapane, M. Codari, and F. Sardanelli, "Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine," *European radiology experimental*, vol. 2, pp. 1-10, 2018.
- [10] A. Gotra *et al.*, "Liver segmentation: indications, techniques and future directions," *Insights into imaging*, vol. 8, pp. 377-392, 2017.
- [11] E. K. Abdalla, R. Adam, A. J. Bilchik, D. Jaeck, J.-N. Vauthey, and D. Mahvi, "Improving resectability of hepatic colorectal metastases: expert consensus statement," *Annals of surgical oncology*, vol. 13, pp. 1271-1280, 2006.
- [12] M. Ahmad *et al.*, "[Retracted] Efficient Liver Segmentation from Computed Tomography Images Using Deep Learning," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, p. 2665283, 2022.
- [13] M. Y. Ansari *et al.*, "Practical utility of liver segmentation methods in clinical surgeries and interventions," *BMC medical imaging*, vol. 22, no. 1, p. 97, 2022.
- [14] M. Moghbel, S. Mashohor, R. Mahmud, and M. I. B. Saripan, "Review of liver segmentation and computer assisted detection/diagnosis methods in computed tomography," *Artificial Intelligence Review*, vol. 50, pp. 497-537, 2018.
- [15] L. Huang, M. Weng, H. Shuai, Y. Huang, J. Sun, and F. Gao, "Automatic Liver Segmentation from CT Images Using Single-Block Linear Detection," *BioMed research international*, vol. 2016, no. 1, p. 9420148, 2016.
- [16] T. Heimann *et al.*, "Comparison and evaluation of methods for liver

- segmentation from CT datasets," *IEEE transactions on medical imaging*, vol. 28, no. 8, pp. 1251-1265, 2009.
- [17] P. Bilic *et al.*, "The liver tumor segmentation benchmark (lits)," *Medical Image Analysis*, vol. 84, p. 102680, 2023.
  - [18] S. R. Abdel-Misih and M. Bloomston, "Liver anatomy," *The Surgical clinics of North America*, vol. 90, no. 4, p. 643, 2010.
  - [19] D. Maucourt-Boulch, C. de Martel, S. Franceschi, and M. Plummer, "Fraction and incidence of liver cancer attributable to hepatitis B and C viruses worldwide," *International journal of cancer*, vol. 142, no. 12, pp. 2471-2477, 2018.
  - [20] S.-C. Chuang, C. La Vecchia, and P. Boffetta, "Liver cancer: descriptive epidemiology and risk factors other than HBV and HCV infection," *Cancer letters*, vol. 286, no. 1, pp. 9-14, 2009.
  - [21] M. Reig *et al.*, "BCLC strategy for prognosis prediction and treatment recommendation: The 2022 update," *Journal of hepatology*, vol. 76, no. 3, pp. 681-693, 2022.
  - [22] P. Regmi *et al.*, "Efficacy and safety of sorafenib plus hepatic arterial infusion chemotherapy for advanced hepatocellular carcinoma," *Surgical Oncology*, vol. 39, p. 101663, 2021.
  - [23] S. Li *et al.*, "The role of hepatic arterial infusion chemotherapy in the treatment of hepatocellular carcinoma: a systematic review and meta-analysis," *Chemotherapy*, vol. 66, no. 4, pp. 124-133, 2021.
  - [24] P. Giraud and A. Houle, "Respiratory gating for radiotherapy: main technical aspects and clinical benefits," *International Scholarly Research Notices*, vol. 2013, no. 1, p. 519602, 2013.
  - [25] F. E. Boas and D. Fleischmann, "CT artifacts: causes and reduction techniques," *Imaging Med*, vol. 4, no. 2, pp. 229-240, 2012.
  - [26] R. M. Juza and E. M. Pauli, "Clinical and surgical anatomy of the liver: a review for clinicians," *Clinical Anatomy*, vol. 27, no. 5, pp. 764-769, 2014.
  - [27] J. H. Fasel and A. Schenk, "Concepts for liver segment classification: neither old ones nor new ones, but a comprehensive one," *Journal of clinical imaging science*, vol. 3, 2013.
  - [28] T. Germain, S. Favelier, J.-P. Cercueil, A. Denys, D. Krausé, and B. Guiu, "Liver segmentation: practical tips," *Diagnostic and interventional imaging*, vol. 95, no. 11, pp. 1003-1016, 2014.
  - [29] A. H. Elgazzar, *Synopsis of pathophysiology in nuclear medicine*. Springer, 2014.
  - [30] E. Britannica. "Anterior and posterior views of the liver." Encyclopædia Britannica. <https://www.britannica.com/science/liver#/media/1/344579/68633> (accessed 2024).
  - [31] H. Ishibashi, M. Nakamura, A. Komori, K. Migita, and S. Shimoda, "Liver architecture, cell function, and disease," in *Seminars in immunopathology*, 2009,

- vol. 31: Springer, pp. 399-409.
- [32] S. Lorente, M. Hautefeuille, and A. Sanchez-Cedillo, "The liver, a functionalized vascular structure," *Scientific Reports*, vol. 10, no. 1, p. 16194, 2020.
  - [33] C. Rouiller, *The liver: morphology, biochemistry, physiology*. Academic Press, 2013.
  - [34] J. C. Ozougwu, "Physiology of the liver," *International Journal of Research in Pharmacy and Biosciences*, vol. 4, no. 8, pp. 13-24, 2017.
  - [35] S. Euvrard and J. Kanitakis, "Skin cancers after liver transplantation: what to do?," *Journal of hepatology*, vol. 44, no. 1, pp. 27-32, 2006.
  - [36] K. Kagohashi, H. Satoh, H. Ishikawa, M. Ohtsuka, and K. Sekizawa, "Liver metastasis at the time of initial diagnosis of lung cancer," *Medical Oncology*, vol. 20, pp. 25-28, 2003.
  - [37] J. R. Diamond, C. A. Finlayson, and V. F. Borges, "Hepatic complications of breast cancer," *The lancet oncology*, vol. 10, no. 6, pp. 615-621, 2009.
  - [38] D. M. Parkin, F. Bray, J. Ferlay, and P. Pisani, "Estimating the world cancer burden: Globocan 2000," 2001.
  - [39] J. Ferlay *et al.*, "Cancer statistics for the year 2020: An overview," *International journal of cancer*, vol. 149, no. 4, pp. 778-789, 2021.
  - [40] M. Sherman, "Surveillance for hepatocellular carcinoma," *Best Practice & Research Clinical Gastroenterology*, vol. 28, no. 5, pp. 783-793, 2014.
  - [41] K. A. McGlynn, J. L. Petrick, and H. B. El-Serag, "Epidemiology of hepatocellular carcinoma," *Hepatology*, vol. 73, pp. 4-13, 2021.
  - [42] C. J. Murray *et al.*, "Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019," *The lancet*, vol. 396, no. 10258, pp. 1223-1249, 2020.
  - [43] D. Jain *et al.*, "Tobacco smoking and liver cancer risk: potential avenues for carcinogenesis," *Journal of Oncology*, vol. 2021, no. 1, p. 5905357, 2021.
  - [44] H. Rungay *et al.*, "Global burden of cancer in 2020 attributable to alcohol consumption: a population-based study," *The Lancet Oncology*, vol. 22, no. 8, pp. 1071-1080, 2021.
  - [45] S. Chen *et al.*, "Aristolochic acids exposure was not the main cause of liver tumorigenesis in adulthood," *Acta Pharmaceutica Sinica B*, vol. 12, no. 5, pp. 2252-2267, 2022.
  - [46] M. E. Rinella, "Nonalcoholic fatty liver disease: a systematic review," *Jama*, vol. 313, no. 22, pp. 2263-2273, 2015.
  - [47] E. Dhamija, S. B. Paul, and S. Kedia, "Non-alcoholic fatty liver disease associated with hepatocellular carcinoma: An increasing concern," *Indian Journal of Medical Research*, vol. 149, no. 1, pp. 9-17, 2019.
  - [48] D. Q. Huang, H. B. El-Serag, and R. Loomba, "Global epidemiology of NAFLD-related HCC: trends, predictions, risk factors and prevention," *Nature reviews Gastroenterology & hepatology*, vol. 18, no. 4, pp. 223-238, 2021.

- [49] M. E. Smela, M. L. Hamm, P. T. Henderson, C. M. Harris, T. M. Harris, and J. M. Essigmann, "The aflatoxin B1 formamidopyrimidine adduct plays a major role in causing the types of mutations observed in human hepatocellular carcinoma," *Proceedings of the National Academy of Sciences*, vol. 99, no. 10, pp. 6655-6660, 2002.
- [50] L. Ma, J. Wang, and Y. Zhang, "Probing the characterization of the interaction of aflatoxins B1 and G1 with calf thymus DNA in vitro," *Toxins*, vol. 9, no. 7, p. 209, 2017.
- [51] W. Zhang *et al.*, "Genetic features of aflatoxin-associated hepatocellular carcinoma," *Gastroenterology*, vol. 153, no. 1, pp. 249-262. e2, 2017.
- [52] H. B. El-Serag and J. A. Davila, "Surveillance for hepatocellular carcinoma: in whom and how?," vol. 4, ed: SAGE Publications UK, 2011, pp. 5-10.
- [53] O. O. Adigun, S. N. S. Yarrarapu, M. Zubair, and S. Khetarpal, "Alpha fetoprotein," 2017.
- [54] B. I. Carr, P. Pancoska, and R. A. Branch, "Low alpha-fetoprotein hepatocellular carcinoma," *Journal of gastroenterology and hepatology*, vol. 25, no. 9, pp. 1543-1549, 2010.
- [55] F. Farinati *et al.*, "Diagnostic and prognostic role of  $\alpha$ -fetoprotein in hepatocellular carcinoma: both or neither?," *Official journal of the American College of Gastroenterology| ACG*, vol. 101, no. 3, pp. 524-532, 2006.
- [56] C.-W. Lee *et al.*, "Normal alpha-fetoprotein hepatocellular carcinoma: are they really normal?," *Journal of clinical medicine*, vol. 8, no. 10, p. 1736, 2019.
- [57] L. C. Adams *et al.*, "Diagnostic performance of susceptibility-weighted magnetic resonance imaging for the detection of calcifications: a systematic review and meta-analysis," *Scientific Reports*, vol. 7, no. 1, p. 15506, 2017.
- [58] J. Neuberger *et al.*, "Guidelines on the use of liver biopsy in clinical practice from the British Society of Gastroenterology, the Royal College of Radiologists and the Royal College of Pathology," *Gut*, vol. 69, no. 8, pp. 1382-1403, 2020.
- [59] A. Boyd, O. Cain, A. Chauhan, and G. J. Webb, "Medical liver biopsy: background, indications, procedure and histopathology," *Frontline gastroenterology*, vol. 11, no. 1, pp. 40-47, 2020.
- [60] J. M. Llovet, C. Brú, and J. Bruix, "Prognosis of hepatocellular carcinoma: the BCLC staging classification," in *Seminars in liver disease*, 1999, vol. 19, no. 03: © 1999 by Thieme Medical Publishers, Inc., pp. 329-338.
- [61] F. Pons, M. Varela, and J. M. Llovet, "Staging systems in hepatocellular carcinoma," *Hpb*, vol. 7, no. 1, pp. 35-41, 2005.
- [62] J.-L. Raoul, A. Forner, L. Bolondi, T. T. Cheung, R. Kloeckner, and T. de Baere, "Updated use of TACE for hepatocellular carcinoma treatment: How and when to use it based on clinical evidence," *Cancer treatment reviews*, vol. 72, pp. 28-36, 2019.
- [63] R. J. Aragon and N. L. Solomon, "Techniques of hepatic resection," *Journal of gastrointestinal oncology*, vol. 3, no. 1, p. 28, 2012.

- [64] K. Hoffmann *et al.*, "Risk assessment for liver resection," *Surgery*, vol. 164, no. 5, pp. 998-1005, 2018.
- [65] R. Kakodkar and A. Soin, "Liver transplantation for HCC: a review," *Indian Journal of Surgery*, vol. 74, pp. 100-117, 2012.
- [66] J. Joseph and K. Rajappan, "Radiofrequency ablation of cardiac arrhythmias: past, present and future," *QJM: An International Journal of Medicine*, vol. 105, no. 4, pp. 303-314, 2012.
- [67] R. Lencioni and L. Crocetti, "Radiofrequency ablation of liver cancer," *Techniques in vascular and interventional radiology*, vol. 10, no. 1, pp. 38-46, 2007.
- [68] L. M. Sutherland, J. A. Williams, R. T. Padbury, D. C. Gotley, B. Stokes, and G. J. Maddern, "Radiofrequency ablation of liver tumors: a systematic review," *Archives of Surgery*, vol. 141, no. 2, pp. 181-190, 2006.
- [69] M. Koda *et al.*, "Complications of radiofrequency ablation for hepatocellular carcinoma in a multicenter study: an analysis of 16 346 treated nodules in 13 283 patients," *Hepatology Research*, vol. 42, no. 11, pp. 1058-1064, 2012.
- [70] D. S. Lu *et al.*, "Influence of large peritumoral vessels on outcome of radiofrequency ablation of liver tumors," *Journal of vascular and interventional radiology*, vol. 14, no. 10, pp. 1267-1274, 2003.
- [71] R. S. Finn *et al.*, "Outcomes of sequential treatment with sorafenib followed by regorafenib for HCC: additional analyses from the phase III RESORCE trial," *Journal of hepatology*, vol. 69, no. 2, pp. 353-358, 2018.
- [72] P. R. Galle *et al.*, "EASL clinical practice guidelines: management of hepatocellular carcinoma," *Journal of hepatology*, vol. 69, no. 1, pp. 182-236, 2018.
- [73] H. Ghanaati, M. Mohammadifard, and M. Mohammadifard, "A review of applying transarterial chemoembolization (TACE) method for management of hepatocellular carcinoma," *Journal of Family Medicine and Primary Care*, vol. 10, no. 10, pp. 3553-3560, 2021.
- [74] Y. Dai *et al.*, "Optimal timing of combining sorafenib with trans-arterial chemoembolization in patients with hepatocellular carcinoma: a meta-analysis," *Translational Oncology*, vol. 14, no. 12, p. 101238, 2021.
- [75] C. Jiang, G. Cheng, M. Liao, and J. Huang, "Individual or combined transcatheter arterial chemoembolization and radiofrequency ablation for hepatocellular carcinoma: a time-to-event meta-analysis," *World Journal of Surgical Oncology*, vol. 19, pp. 1-13, 2021.
- [76] Y. Yu, J. Fu, P. Xia, and C. Chu, "A systematic review and meta-analysis on the efficacy and safety of transcatheter arterial chemoembolization combined with radiofrequency ablation in the treatment of primary liver cancer," *Translational Cancer Research*, vol. 11, no. 5, p. 1297, 2022.
- [77] S. Kulkarni, J. A. Rumberger, and S. Jha, "Electron beam CT: a historical review," *American Journal of Roentgenology*, vol. 216, no. 5, pp. 1222-1228,



- 2021.
- [78] R. Zeman *et al.*, "Helical (spiral) CT of the abdomen," *AJR. American journal of roentgenology*, vol. 160, no. 4, pp. 719-725, 1993.
  - [79] J. Hsieh, "Computed tomography: principles, design, artifacts, and recent advances," 2003.
  - [80] T. Flohr, "CT systems," *Current Radiology Reports*, vol. 1, no. 1, pp. 52-63, 2013.
  - [81] L. W. Goldman, "Principles of CT and CT technology," *Journal of nuclear medicine technology*, vol. 35, no. 3, pp. 115-128, 2007.
  - [82] J. F. Barrett and N. Keat, "Artifacts in CT: recognition and avoidance," *Radiographics*, vol. 24, no. 6, pp. 1679-1691, 2004.
  - [83] M. Katsura, J. Sato, M. Akahane, A. Kunimatsu, and O. Abe, "Current and novel techniques for metal artifact reduction at CT: practical guide for radiologists," *Radiographics*, vol. 38, no. 2, pp. 450-461, 2018.
  - [84] M. Guziński, Ł. Waszczuk, and M. J. Sasiadek, "Head CT: image quality improvement of posterior fossa and radiation dose reduction with ASiR-comparative studies of CT head examinations," *European Radiology*, vol. 26, pp. 3691-3696, 2016.
  - [85] S. Xie, C. Li, H. Li, and Q. Ge, "A level set method for cupping artifact correction in cone-beam CT," *Medical physics*, vol. 42, no. 8, pp. 4888-4895, 2015.
  - [86] W. A. Kalender, *Computed tomography: fundamentals, system technology, image quality, applications*. John Wiley & Sons, 2011.
  - [87] A. G. Kane, K. C. Reilly, and T. F. Murphy, "Swimmer's CT: improved imaging of the lower neck and thoracic inlet," *American journal of neuroradiology*, vol. 25, no. 5, pp. 859-862, 2004.
  - [88] G. Glover and N. Pelc, "Nonlinear partial volume artifacts in x-ray computed tomography," *Medical physics*, vol. 7, no. 3, pp. 238-248, 1980.
  - [89] Y. Zou, E. Y. Sidky, and X. Pan, "Partial volume and aliasing artefacts in helical cone-beam CT," *Physics in Medicine & Biology*, vol. 49, no. 11, p. 2365, 2004.
  - [90] I. Mori, Y. Machida, M. Osanai, and K. Iinuma, "Photon starvation artifacts of X-ray CT: their true cause and a solution," *Radiological physics and technology*, vol. 6, pp. 130-141, 2013.
  - [91] P. J. La Riviere and C. R. Crawford, "From EMI to AI: a brief history of commercial CT reconstruction algorithms," *Journal of Medical Imaging*, vol. 8, no. 5, pp. 052111-052111, 2021.
  - [92] D. Fleischmann and F. E. Boas, "Computed tomography—old ideas and new technology," *European radiology*, vol. 21, pp. 510-517, 2011.
  - [93] M. D. Gilman, A. J. Fischman, V. Krishnasetty, E. F. Halpern, and S. L. Aquino, "Optimal CT breathing protocol for combined thoracic PET/CT," *American Journal of Roentgenology*, vol. 187, no. 5, pp. 1357-1360, 2006.
  - [94] S. Nehmeh *et al.*, "Effect of respiratory gating on reducing lung motion artifacts

- in PET imaging of lung cancer," *Medical physics*, vol. 29, no. 3, pp. 366-371, 2002.
- [95] L. A. Shepp and B. F. Logan, "The Fourier reconstruction of a head section," *IEEE Transactions on nuclear science*, vol. 21, no. 3, pp. 21-43, 1974.
  - [96] J. You, "Noise analysis and treatment for SPECT imaging via an FBP algorithm with classical filters," *Preprint*, 2005.
  - [97] F. E. Boas and D. Fleischmann, "Evaluation of two iterative techniques for reducing metal artifacts in computed tomography," *Radiology*, vol. 259, no. 3, pp. 894-902, 2011.
  - [98] R. Schulze *et al.*, "Artefacts in CBCT: a review," *Dentomaxillofacial Radiology*, vol. 40, no. 5, pp. 265-273, 2011.
  - [99] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, "Metal streak artifacts in X-ray computed tomography: a simulation study," in *1998 IEEE Nuclear Science Symposium Conference Record. 1998 IEEE Nuclear Science Symposium and Medical Imaging Conference (Cat. No. 98CH36255)*, 1998, vol. 3: IEEE, pp. 1860-1865.
  - [100] M. Drozdal *et al.*, "Learning normalized inputs for iterative estimation in medical image segmentation," *Medical image analysis*, vol. 44, pp. 1-13, 2018.
  - [101] K. Kaczor *et al.*, "Comparison of image pre-processing methods in liver segmentation task," in *2022 15th International Conference on Human System Interaction (HSI)*, 2022: IEEE, pp. 1-5.
  - [102] M. Islam, K. N. Khan, and M. S. Khan, "Evaluation of preprocessing techniques for U-Net based automated liver segmentation," in *2021 International Conference on Artificial Intelligence (ICAI)*, 2021: IEEE, pp. 187-192.
  - [103] G. Chartrand, T. Cresson, R. Chav, A. Gotra, A. Tang, and J. DeGuise, "Semi-automated liver CT segmentation using Laplacian meshes," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, 2014: IEEE, pp. 641-644.
  - [104] N. Sharma and L. M. Aggarwal, "Automated medical image segmentation techniques," *Journal of medical physics*, vol. 35, no. 1, pp. 3-14, 2010.
  - [105] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, 2015: Springer, pp. 234-241.
  - [106] C. Shi, Y. Cheng, F. Liu, Y. Wang, J. Bai, and S. Tamura, "A hierarchical local region-based sparse shape composition for liver segmentation in CT scans," *Pattern Recognition*, vol. 50, pp. 88-106, 2016.
  - [107] V. Czipczer and A. Manno-Kovacs, "Adaptable volumetric liver segmentation model for CT images using region-based features and convolutional neural network," *Neurocomputing*, vol. 505, pp. 388-401, 2022.
  - [108] S. Sun and R. Zhang, "Region of interest extraction of medical image based on

- improved region growing algorithm," in *2017 International Conference on Material Science, Energy and Environmental Engineering (MSEEE 2017)*, 2017: Atlantis Press, pp. 471-475.
- [109] R. Pohle and K. D. Toennies, "Segmentation of medical images using adaptive region growing," in *Medical Imaging 2001: Image Processing*, 2001, vol. 4322: SPIE, pp. 1337-1346.
  - [110] J. Yan, L. H. Schwartz, and B. Zhao, "Semiautomatic segmentation of liver metastases on volumetric CT images," *Medical physics*, vol. 42, no. 11, pp. 6283-6293, 2015.
  - [111] S. Jardim, J. António, and C. Mora, "Image thresholding approaches for medical image segmentation-short literature review," *Procedia Computer Science*, vol. 219, pp. 1485-1492, 2023.
  - [112] Z. Wang, C. Zhang, T. Jiao, M. Gao, and G. Zou, "Fully Automatic Segmentation and Three-Dimensional Reconstruction of the Liver in CT Images," *Journal of Healthcare Engineering*, vol. 2018, no. 1, p. 6797102, 2018.
  - [113] S. K. Siri, S. P. Kumar, and M. V. Latte, "Threshold-based new segmentation model to separate the liver from CT scan images," *IETE Journal of Research*, vol. 68, no. 6, pp. 4468-4475, 2022.
  - [114] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in ND images," in *Proceedings eighth IEEE international conference on computer vision. ICCV 2001*, 2001, vol. 1: IEEE, pp. 105-112.
  - [115] F. Yi and I. Moon, "Image segmentation: A survey of graph-cut methods," in *2012 international conference on systems and informatics (ICSAI2012)*, 2012: IEEE, pp. 1936-1941.
  - [116] X. Chen and L. Pan, "A survey of graph cuts/graph search based medical image segmentation," *IEEE reviews in biomedical engineering*, vol. 11, pp. 112-124, 2018.
  - [117] X. Chen, J. K. Udupa, U. Bagci, Y. Zhuge, and J. Yao, "Medical image segmentation by combining graph cuts and oriented active appearance models," *IEEE transactions on image processing*, vol. 21, no. 4, pp. 2035-2046, 2012.
  - [118] G. Li, X. Chen, F. Shi, W. Zhu, J. Tian, and D. Xiang, "Automatic liver segmentation based on shape constraints and deformable graph cut in CT images," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5315-5329, 2015.
  - [119] M. Liao *et al.*, "Efficient liver segmentation in CT images based on graph cuts and bottleneck detection," *Physica Medica*, vol. 32, no. 11, pp. 1383-1396, 2016.
  - [120] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433-459, 2010.
  - [121] A. Afifi and T. Nakaguchi, "Liver segmentation approach using graph cuts and iteratively estimated shape and intensity constrains," in *International*

- Conference on Medical Image Computing and Computer-Assisted Intervention*, 2012: Springer, pp. 395-403.
- [122] X. Lu, Q. Xie, Y. Zha, and D. Wang, "Fully automatic liver segmentation combining multi-dimensional graph cut with shape information in 3D CT images," *Scientific reports*, vol. 8, no. 1, p. 10700, 2018.
  - [123] T. McInerney and D. Terzopoulos, "Deformable models in medical image analysis: a survey," *Medical image analysis*, vol. 1, no. 2, pp. 91-108, 1996.
  - [124] H. Lamecker, T. Lange, and M. Seebass, "A statistical shape model for the liver," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2002: 5th International Conference Tokyo, Japan, September 25–28, 2002 Proceedings, Part II* 5, 2002: Springer, pp. 421-427.
  - [125] H. Lamecker, T. Lange, and M. Seebass, "Segmentation of the liver using a 3D statistical shape model," 2004.
  - [126] J. Lu, D. Wang, L. Shi, and P. A. Heng, "Automatic liver segmentation in CT images based on support vector machine," in *Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics*, 2012: IEEE, pp. 333-336.
  - [127] S. Marchesseau, T. Heimann, S. Chatelin, R. Willinger, and H. Delingette, "Fast porous visco-hyperelastic soft tissue model for surgery simulation: application to liver surgery," *Progress in biophysics and molecular biology*, vol. 103, no. 2-3, pp. 185-196, 2010.
  - [128] K. Nishii, G. Reese, E. C. Moran, and J. L. Sparks, "Multiscale computational model of fluid flow and matrix deformation in decellularized liver," *Journal of the mechanical behavior of biomedical materials*, vol. 57, pp. 201-214, 2016.
  - [129] R. M. Devi and V. Seenivasagam, "Automatic segmentation and classification of liver tumor from CT image using feature difference and SVM based classifier-soft computing technique," *Soft Computing*, vol. 24, no. 24, pp. 18591-18598, 2020.
  - [130] F. Ambellan, H. Lamecker, C. von Tycowicz, and S. Zachow, *Statistical shape models: understanding and mastering variation in anatomy*. Springer, 2019.
  - [131] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International journal of computer vision*, vol. 1, no. 4, pp. 321-331, 1988.
  - [132] B. Zhao, J. Soraghan, D. Grose, T. Doshi, and G. Di-Caterina, "Automatic 3D detection and segmentation of head and neck cancer from MRI data," in *2018 7th European Workshop on Visual Information Processing (EUVIP)*, 2018: IEEE, pp. 1-6.
  - [133] S. Osher and R. P. Fedkiw, "Level set methods: an overview and some recent results," *Journal of Computational physics*, vol. 169, no. 2, pp. 463-502, 2001.
  - [134] J. Lee *et al.*, "Efficient liver segmentation using a level-set method with optimal detection of the initial liver boundary from level-set speed images," *Computer methods and programs in biomedicine*, vol. 88, no. 1, pp. 26-38, 2007.
  - [135] Y. Li *et al.*, "Liver segmentation from abdominal CT volumes based on level set

- and sparse shape composition," *Computer methods and programs in biomedicine*, vol. 195, p. 105533, 2020.
- [136] R. Plantefeve, I. Peterlik, N. Haouchine, and S. Cotin, "Patient-specific biomechanical modeling for guidance during minimally-invasive hepatic surgery," *Annals of biomedical engineering*, vol. 44, pp. 139-153, 2016.
  - [137] S. Misra, K. Ramesh, and A. M. Okamura, "Modeling of tool-tissue interactions for computer-based surgical simulation: A literature review," *Presence*, vol. 17, no. 5, pp. 463-491, 2008.
  - [138] R. Ranjbarzadeh and S. B. Saadi, "Automated liver and tumor segmentation based on concave and convex points using fuzzy c-means and mean shift clustering," *Measurement*, vol. 150, p. 107086, 2020.
  - [139] S. Sangewar, A. A. Peshattiwar, V. Alagdeve, and R. Balpande, "Liver segmentation of CT scan images using K means algorithm," in *2013 International Conference on Advanced Electronic Systems (ICAES)*, 2013: IEEE, pp. 6-9.
  - [140] D. Krasnov, D. Davis, K. Malott, Y. Chen, X. Shi, and A. Wong, "Fuzzy c-means clustering: A review of applications in breast cancer detection," *Entropy*, vol. 25, no. 7, p. 1021, 2023.
  - [141] P. Yugander and G. R. Reddy, "Liver tumor segmentation in noisy CT images using distance regularized level set evolution based on fuzzy C-means clustering," in *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, 2017: IEEE, pp. 1530-1534.
  - [142] S. X. Ma *et al.*, "Bayesian networks in radiology," *Radiology: Artificial Intelligence*, vol. 5, no. 6, p. e210187, 2023.
  - [143] R. G. Mohamed, N. A. Seada, S. Hamdy, and M. G. Mostafa, "An adaptive method for fully automatic liver segmentation in medical MRI-images," *International Journal of Computer Applications*, vol. 179, no. 4, pp. 12-18, 2017.
  - [144] M. Freiman, O. Eliassaf, Y. Taieb, L. Joskowicz, Y. Azraq, and J. Sosna, "An iterative Bayesian approach for nearly automatic liver segmentation: algorithm and validation," *International Journal of Computer Assisted Radiology and Surgery*, vol. 3, pp. 439-446, 2008.
  - [145] E. Kyrimi *et al.*, "Bayesian networks in healthcare: What is preventing their adoption?," *Artificial Intelligence in Medicine*, vol. 116, p. 102079, 2021.
  - [146] S. Luo, J. S. Jin, S. K. Chalup, and G. Qian, "A liver segmentation algorithm based on wavelets and machine learning," in *2009 international conference on computational intelligence and natural computing*, 2009, vol. 2: IEEE, pp. 122-125.
  - [147] N. N. Reddy and G. Ramkumar, "Liver segmentation and classification in computed tomography images using convolutional neural network and comparison of accuracy with support vector machine," in *AIP Conference*

- Proceedings*, 2023, vol. 2821, no. 1: AIP Publishing.
- [148] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," *Psychological review*, vol. 65, no. 6, p. 386, 1958.
  - [149] B. Widrow and M. A. Lehr, "30 years of adaptive neural networks: perceptron, madaline, and backpropagation," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1415-1442, 1990.
  - [150] F. Rosenblatt and S. Papert, *Perceptron*. April, 2021.
  - [151] I. Goodfellow *et al.*, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139-144, 2020.
  - [152] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.
  - [153] S. Suthaharan and S. Suthaharan, "Support vector machine," *Machine learning models and algorithms for big data classification: thinking with examples for effective learning*, pp. 207-235, 2016.
  - [154] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527-1554, 2006.
  - [155] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *The Annals of Mathematical Statistics*, pp. 462-466, 1952.
  - [156] B. T. Polyak, "Some methods of speeding up the convergence of iteration methods," *Ussr computational mathematics and mathematical physics*, vol. 4, no. 5, pp. 1-17, 1964.
  - [157] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of machine learning research*, vol. 12, no. 7, 2011.
  - [158] D. P. Kingma, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
  - [159] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.
  - [160] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.
  - [161] K. Simonyan, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
  - [162] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.
  - [163] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
  - [164] M. Yani, S. Budhi Irawan, Si, MT, and S. Casi Setiningsih, MT, "Application of transfer learning using convolutional neural network method for early detection of terry's nail," in *Journal of Physics: Conference Series*, 2019, vol. 1201, no.

- 1: IOP Publishing, p. 012052.
- [165] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431-3440.
  - [166] A. Ben-Cohen, I. Diamant, E. Klang, M. Amitai, and H. Greenspan, "Fully convolutional network for liver segmentation and lesions detection," in *Deep Learning and Data Labeling for Medical Applications: First International Workshop, LABELS 2016, and Second International Workshop, DLMIA 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 21, 2016, Proceedings 1*, 2016: Springer, pp. 77-85.
  - [167] D. Ciresan, A. Giusti, L. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," *Advances in neural information processing systems*, vol. 25, 2012.
  - [168] P. F. Christ *et al.*, "Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks," *arXiv preprint arXiv:1702.05970*, 2017.
  - [169] C. Sun *et al.*, "Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on FCNs," *Artificial intelligence in medicine*, vol. 83, pp. 58-66, 2017.
  - [170] X. Han, "Automatic liver lesion segmentation using a deep convolutional neural network method," *arXiv preprint arXiv:1704.07239*, 2017.
  - [171] S. Li and G. Tso, "Bottleneck Supervised U-Net for Pixel-wise Liver and Tumour Segmentation," *Bottleneck Supervised U-Net for Pixel-wise Liver and Tumour Segmentation*, 1810.
  - [172] L. Jiang *et al.*, "RMAU-Net: residual multi-scale attention U-Net for liver and tumor segmentation in CT images," *Computers in Biology and Medicine*, vol. 158, p. 106838, 2023.
  - [173] Y. Chen *et al.*, "A deep residual attention-based U-Net with a biplane joint method for liver segmentation from CT scans," *Computers in Biology and Medicine*, vol. 152, p. 106421, 2023.
  - [174] Q. Jin, Z. Meng, C. Sun, H. Cui, and R. Su, "RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans," *Frontiers in Bioengineering and Biotechnology*, vol. 8, p. 605132, 2020.
  - [175] H. Jiang, T. Shi, Z. Bai, and L. Huang, "Ahcnet: An application of attention mechanism and hybrid connection for liver tumor segmentation in ct volumes," *Ieee Access*, vol. 7, pp. 24898-24909, 2019.
  - [176] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, 2018: Springer, pp. 3-11.

- [177] J. Li *et al.*, "Eres-UNet++: Liver CT image segmentation based on high-efficiency channel attention and Res-UNet++," *Computers in Biology and Medicine*, vol. 158, p. 106501, 2023.
- [178] H. Huang *et al.*, "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)*, 2020: IEEE, pp. 1055-1059.
- [179] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961-2969.
- [180] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1137-1149, 2016.
- [181] M. N. U. Haq, A. Irtaza, N. Nida, M. A. Shah, and L. Zubair, "Liver tumor segmentation using resnet based mask-R-CNN," in *2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST)*, 2021: IEEE, pp. 276-281.
- [182] X. Chen *et al.*, "Liver segmentation in CT imaging with enhanced mask region-based convolutional neural networks," *Annals of translational medicine*, vol. 9, no. 24, 2021.
- [183] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE transactions on medical imaging*, vol. 37, no. 12, pp. 2663-2674, 2018.
- [184] J. Chi, X. Han, C. Wu, H. Wang, and P. Ji, "X-Net: Multi-branch UNet-like network for liver and tumor segmentation from 3D abdominal CT scans," *Neurocomputing*, vol. 459, pp. 81-96, 2021.
- [185] Z. Bai, H. Jiang, S. Li, and Y.-D. Yao, "Liver tumor segmentation based on multi-scale candidate generation and fractal residual network," *Ieee Access*, vol. 7, pp. 82122-82133, 2019.
- [186] T. Zhou, Q. Li, H. Lu, Q. Cheng, and X. Zhang, "GAN review: Models and medical image fusion applications," *Information Fusion*, vol. 91, pp. 134-148, 2023.
- [187] J. J. Jeong, A. Tariq, T. Adejumo, H. Trivedi, J. W. Gichoya, and I. Banerjee, "Systematic review of generative adversarial networks (GANs) for medical image classification and segmentation," *Journal of Digital Imaging*, vol. 35, no. 2, pp. 137-152, 2022.
- [188] X. Xiao *et al.*, "Radiomics-guided GAN for segmentation of liver tumor without contrast agents," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*, 2019: Springer, pp. 237-245.
- [189] Y. Enokiya, Y. Iwamoto, Y.-W. Chen, and X.-H. Han, "Automatic liver segmentation using U-Net with Wasserstein GANs," *Journal of Image and Graphics*, vol. 6, no. 2, pp. 152-159, 2018.



- [190] Z. Zhang, B. Li, and J. Xu, "Automatic three-dimensional segmentation of liver and tumors regions based on conditional generative adversarial networks," *Sheng wu yi xue Gong Cheng xue za zhi= Journal of Biomedical Engineering= Shengwu Yixue Gongchengxue Zazhi*, vol. 38, no. 1, pp. 80-88, 2021.
- [191] M. Heker and H. Greenspan, "Joint liver lesion segmentation and classification via transfer learning," *arXiv preprint arXiv:2004.12352*, 2020.
- [192] S. Chen, K. Ma, and Y. Zheng, "Med3d: Transfer learning for 3d medical image analysis," *arXiv preprint arXiv:1904.00625*, 2019.
- [193] H. Cao *et al.*, "Swin-unet: Unet-like pure transformer for medical image segmentation," in *European conference on computer vision*, 2022: Springer, pp. 205-218.
- [194] L. Li and H. Ma, "Rdetrans u-net: A hybrid variable architecture for liver ct image segmentation," *Sensors*, vol. 22, no. 7, p. 2452, 2022.
- [195] P. E. Sijens, M. A. Edens, S. J. Bakker, and R. P. Stolk, "MRI-determined fat content of human liver, pancreas and kidney," *World journal of gastroenterology: WJG*, vol. 16, no. 16, p. 1993, 2010.
- [196] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.
- [197] F. He, T. Liu, and D. Tao, "Control batch size and learning rate to generalize well: Theoretical and empirical evidence," *Advances in neural information processing systems*, vol. 32, 2019.
- [198] D. Masters and C. Luschi, "Revisiting small batch training for deep neural networks," *arXiv preprint arXiv:1804.07612*, 2018.
- [199] A. Gotra *et al.*, "Comparison of MRI-and CT-based semiautomated liver segmentation: a validation study," *Abdominal radiology*, vol. 42, pp. 478-489, 2017.
- [200] R. Susomboon, D. S. Raicu, and J. Furst, "A hybrid approach for liver segmentation," in *Proceedings of MICCAI workshop on 3D segmentation in the clinic: a grand challenge*, 2007: Citeseer, pp. 151-160.
- [201] K. Roth, T. Konopczyński, and J. Hesser, "Liver lesion segmentation with slice-wise 2d tiramisu and tversky loss function," *arXiv preprint arXiv:1905.03639*, 2019.
- [202] O. I. Alirr, "Deep learning and level set approach for liver and tumor segmentation from CT scans," *Journal of Applied Clinical Medical Physics*, vol. 21, no. 10, pp. 200-209, 2020.
- [203] K. Han *et al.*, "An effective semi-supervised approach for liver CT image segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 8, pp. 3999-4007, 2022.
- [204] X. Xie *et al.*, "Dynamic adaptive residual network for liver CT image segmentation," *Computers & Electrical Engineering*, vol. 91, p. 107024, 2021.
- [205] X. Xie, X. Pan, F. Shao, W. Zhang, and J. An, "Mci-net: multi-scale context integrated network for liver ct image segmentation," *Computers and Electrical*

- Engineering*, vol. 101, p. 108085, 2022.
- [206] L. Bi, J. Kim, A. Kumar, and D. Feng, "Automatic liver lesion detection using cascaded deep residual networks," *arXiv preprint arXiv:1704.02703*, 2017.
  - [207] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3-19.
  - [208] P. Bilic *et al.*, "The liver tumor segmentation benchmark (lits)," *arXiv preprint arXiv:1901.04056*, 2019.
  - [209] M. G. Linguraru *et al.*, "Tumor burden analysis on computed tomography by automated liver and tumor segmentation," *IEEE transactions on medical imaging*, vol. 31, no. 10, pp. 1965-1976, 2012.
  - [210] K. C. Kaluva, M. Khened, A. Kori, and G. Krishnamurthi, "2D-densely connected convolution neural networks for automatic liver and tumor segmentation," *arXiv preprint arXiv:1802.02182*, 2018.
  - [211] A. A. Albishri, S. J. H. Shah, and Y. Lee, "CU-Net: Cascaded U-Net model for automated liver and lesion segmentation and summarization," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2019: IEEE, pp. 1416-1423.
  - [212] S. Saha Roy, S. Roy, P. Mukherjee, and A. Halder Roy, "An automated liver tumour segmentation and classification model by deep learning based approaches," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 11, no. 3, pp. 638-650, 2023.
  - [213] Z. Zhao, Z. Ma, Y. Liu, Z. Zeng, and P. K. Chow, "Multi-slice dense-sparse learning for efficient liver and tumor segmentation," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2021: IEEE, pp. 3582-3585.
  - [214] G. Chlebus, A. Schenk, J. H. Moltz, B. van Ginneken, H. K. Hahn, and H. Meine, "Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing," *Scientific reports*, vol. 8, no. 1, p. 15497, 2018.
  - [215] L. Han, Y. Chen, J. Li, B. Zhong, Y. Lei, and M. Sun, "Liver segmentation with 2.5 D perpendicular UNets," *Computers & Electrical Engineering*, vol. 91, p. 107118, 2021.