Automatic liver tumor segmentation in computed tomography (CT) imaging

Mr. Kasun Gayashan Hettihewa

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Engineering in Computer Engineering

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2023

การแบ่งส่วนเนื้องอกในตับอัตโนมัติด้วยการคำนวณจากถ่ายภาพเอกซเรย์คอมพิวเตอร์ (CT)

นายกาซัน คยาชาน เฮตติเทวา

Thesis Title   Automatic liver tumor segmentation in computed tomography (CT) imaging

By   Mr. Kasun Gayashan Hettihewa

Field of Study   Computer Engineering

Thesis Advisor   Associate Professor Thanarat Chalidabhongse, PhD

Accepted by the Faculty of Engineering ,Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Engineering

-- Dean of the Faculty of Engineering

(Professor Supot Teachavorasinskun, Ph.D.)

THESIS COMMITTEE

-- Chairman

(Associate Professor Duangdao Wichadakul, Ph.D.)

-- Thesis Advisor

(Associate Professor Thanarat Chalidabhongse, PhD)

-- External Examiner

(Professor Kosin Chamnongthai, Ph.D.)

กาซัน คยาชาน เฮตติเทวา :
การแบ่งส่วนเนื้องอกในตับอัตโนมัติด้วยการคำนวณจากถ่ายภาพเอกซเรย์คอมพิวเตอร์ (CT). (Automatic liver tumor segmentation in computed tomography (CT) imaging) อ.ที่ปรึกษาหลัก : ธนารัตน์ ชลิดาพงศ์

การแบ่งส่วนภาพของเนื้องอกในตับจากภาพถ่ายรังสีแบบอัตโนมัติ เป็นวิธีการที่สำคัญในการวินิจฉัยและรักษาโรคที่เกี่ยวกับเนื้องอกในตับ แต่ด้วยความที่ภาพถ่ายรังสีของเนื้องอกมีรูปร่างและความเข้มแสงที่หลากหลายมากจึงทำให้การแบ่งส่วนภาพนั้นมีความท้าทายอย่างยิ่ง

หลายปีมานี้ โมเดลโครงข่ายประสาทเทียมเชิงลึกได้ถูกนำมาใช้ในการแยกส่วนภาพทางการแพทย์โดยนำโมเดลมาใช้เพื่อการสกัดคุณลักษณะเด่นจากภาพ รวมถึงการเรียนรู้คุณลักษณะเด่นเหล่านั้น แต่อย่างไรก็ดีโมเดลโครงข่ายประสาทเทียมเชิงลึกเหล่านี้ ยังไม่แม่นยำเท่าสายตาและความเชี่ยวชาญของรังสีแพทย์ เนื่องจากความซับซ้อนของภาพ แต่ด้วยการกลไกการทำงานของเทคนิค Attention mechanisms ในการ optimize การเลือกคุณลักษณะภาพ ได้รับการพัฒนาอย่างต่อเนื่อง จนระบบมีความสามารถในการทำ Visual attention ที่มีประสิทธิภาพและใกล้เคียงมนุษย์มากขึ้น

ในงานวิจัยนี้ เราจึงเสนอเครือข่ายใหม่ที่เรียกว่า Multi Attention Network หรือ MANet ซึ่งเป็นการผสมผสานเทคนิค Attention เพื่อเรียนรู้และเน้นคุณลักษณะที่สำคัญในขณะเดียวกันก็ตัดคุณลักษณะที่ไม่เกี่ยวข้องกับการแบ่งส่วนภาพเนื้องอก ซึ่ง MANet ใช้สถาปัตยกรรม U-Net เป็นฐานและยังมีตัวเข้ารหัสที่ใช้กลไก Residual mechanism ด้วย อีกทั้งมีการใช้ร่วมกับ convolutional block attention module (CBAM) หรือโมเดลที่มีการแบ่งโมดูล ออกเป็นโมดูลที่สนใจฟีเจอร์ที่สำคัญและโมดูลที่สนใจพื้นที่ที่สำคัญ โดยจะนำไปใช้ในตัวเข้ารหัสและตัวถอดรหัสแยกกัน สรุปได้ว่าเราได้นำ Attention U-Net มาใช้ร่วมกับ CBAM เพื่อรวมคุณสมบัติเด่นของทั้งสองเข้าด้วยกัน โดยสถาปัตยกรรมการเรียนรู้เชิงลึกที่เราสร้างขึ้นหรือ MANet ได้รับการฝึกฝนและประเมินประสิทธิภาพโดยใช้ตัววัดการประเมินหลายรายการ โดยใช้ชุดข้อมูล MICCAI 2017 Liver Tumor Segmentation (LiTS17) และชุดข้อมูล DIRCADb ซึ่งผลจากการทดสอบพบว่า MANet นั้นใช้พารามิเตอร์น้อยกว่าแต่ยังให้ผลลัพธ์ที่น่าพึงพอใจเมื่อเปรียบเทียบกับวิธีการที่ใช้ในปัจจุบัน (State-of-the- art)

สาขาวิชา	วิศวกรรมคอมพิวเตอร์	ลายมือชื่อนิสิต ....
ปีการศึกษา	2566	ลายมือชื่อ อ.ที่ปรึกษาหลัก ....

6470369721 : MAJOR COMPUTER ENGINEERING

KEYWORD: Deep learning, UNet, convolutional neural network (CNN), Channel Attention, Spatial Attention, Liver Tumor Segmentation

Kasun Gayashan Hettihewa : Automatic liver tumor segmentation in computed tomography (CT) imaging. Advisor: Assoc. Prof. Thanarat Chalidabhongse, PhD

Automatic liver tumor segmentation is a highly important application for diagnosing and treating liver tumors. Due to the diversity of tumor shape and intensity alteration, it has become an extremely challenging procedure. Automatic liver tumor segmentation has the potential to establish a diagnostic standard for providing important radiological information to physicians.

Recently, deep convolutional neural networks have shown numerous benefits in feature extraction and learning in terms of medical image segmentation. However, the model can be inconsistent in imitating visual attention as well as awareness of radiological expertise for tumor recognition and segmentation tasks due to multi-layer dense feature stacking. Attention mechanisms for optimized feature selection have evolved to bridge that gap in visual attention capabilities.

In this research, we propose a novel network called Multi Attention Network (MANet) as a fusion of attention techniques to learn and emphasize significant features while suppressing irrelevant features for liver tumor segmentation. The proposed deep learning network is based on the U-Net architecture. Furthermore, the encoder has a residual mechanism. The convolutional block attention module (CBAM) has been divided into channel attention and spatial attention modules to be implemented in the encoder and decoder separately. The spatial attention mechanism in Attention U-Net has been integrated into the proposed network to capture low-level features to combine with high-level ones. The constructed deep learning architecture is trained and evaluated using multiple evaluation metrics using the publically available MICCAI 2017 Liver Tumor Segmentation (LiTS17) dataset and 3DIRCADb dataset. MANet produced promising results when compared to state-of-the-art methods with relatively low parameter overhead.

Field of Study:	Computer Engineering	Student's Signature ....
Academic Year:	2023	Advisor's Signature ....

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

LIST OF TABLES

LIST OF FIGURES

1. Introduction

One of the major cancer forms with the highest number of reported mortalities worldwide is liver cancer (Siegel et al., 2019, 2020). Early diagnosis of liver cancers is crucial for rapid clinical care to be successful in achieving survival. To assess the severity of the disease, tumor burden analysis, which includes key elements of quantifying the size and location of the tumor, is crucial. Medical imaging is a noninvasive method for classifying and assessing the severity of malignancy. Prior to the pathological examination, radiologists mostly rely on Computed Tomography (CT) scans for the diagnosis and clinical care. The contrast enhancement on CT scans allows doctors to differentiate the tumor location from the liver parenchyma. However, radiologists still face difficulties in identifying tumor locations because of the significant interclass similarity, intraclass variance, and fussy tumor boundaries. Computer-aided detection systems are extremely helpful in establishing diagnostic criteria to bridge the knowledge gap in all levels of radiological competence in order to address these difficulties.

The development of computer-aided automatic liver tumor segmentation methods still faces significant difficulties. The high expense of data collection for experiments is a main challenge. To create a suitable medical dataset to train and evaluate the model, data labeling is a time-consuming and tedious activity. Tumor diversity is a significant factor that contributes to the incorrect classification of tumor areas. Tumors may appear in many forms, at various sites, and in varying numbers. It becomes more difficult to distinguish tumor regions from healthy liver due to the intensity dissimilarity in tumor regions.

1.1. Liver

As the largest organ of the human body, the liver accounts for approximately 2% to 3% of the average body weight (Sherif & Mark, 2010). The liver lies underneath the right lung where is the upper right abdominal cavity under the right hemidiaphragm. Mainly, the liver is divided into two lobes which are described by the morphological anatomy and functional anatomy illustrated in Figure 1. Hepatocytes are the main cell type that forms the liver. There are other cells that line its blood vessels and small tubes called bile ducts (Society, 2019). It secretes a clear yellow or orange fluid called bile that helps with digestion. Moreover, the liver breaks down and stores nutrients that are required to maintain the function of the body. Apart from that some nutrients change (metabolize) in the liver to form energy or repair the tissues in the body. More importantly, the liver clears the blood of harmful substances and fights against infections by producing immune factors to remove bacteria from the bloodstream (HEALTH, 2023).
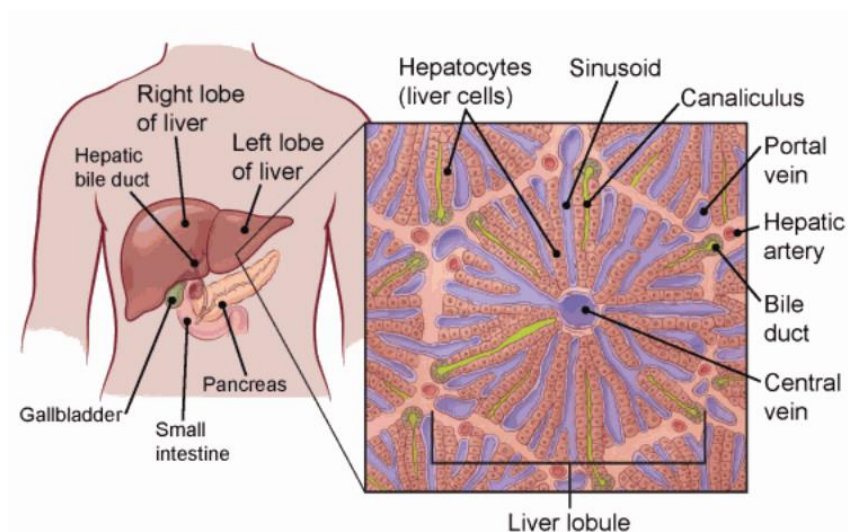
Figure 1: Anatomy of the liver (HEALTH, 2023)

1.2. Liver Cancers

Liver cancer has become the 6th most common cancer worldwide. It is the 5th most common cancer type among men while the 9th most common cancer time in women (Cancer, 2022). According to the study conducted in 2022, they have reported that 905,700 people were diagnosed with liver cancer and 830,200 people lost their lives due to liver cancer globally in 2020 (Harriet et al., 2022). Furthermore, they have predicted the number of new liver cancer cases and deaths can increase by more than 55% between 2020 to 2040, 1.4 million people can be diagnosed in 2040 and 1.3 million people could die due to liver cancer in 2040 based on the predictive analysis (Harriet et al., 2022). A summary of the study is graphically illustrated in Figure 2.

Liver cancers can be divided into two main categories, primary liver cancers and secondary liver cancers (Society, 2019). Primary liver cancers refer to the abnormal growth of cells within the liver that lead to form tumors in the liver. Secondary liver cancers are initiated in other parts of the body and spread into the liver. Hepatocellular carcinoma (HCC), which is also called hepatoma, is the most prevalent primary liver cancer type and forms in hepatocellular cells. It is the main type of cells in the liver (Hope, 2023). HCC is directly associated with chronic liver diseases such as hepatitis B, hepatitis C infections, and cirrhosis. High alcohol consumption could also increase the risk of HCC. Intrahepatic cholangiocarcinoma is a primary liver cancer type that originates in bile ducts within the liver. Cholangiocarcinoma appears in approximately 10% to 20% of liver cancers which is almost identical to 8,000 patients per year, according to the liver cancer statistics of the American Cancer Society (Society, 2019).
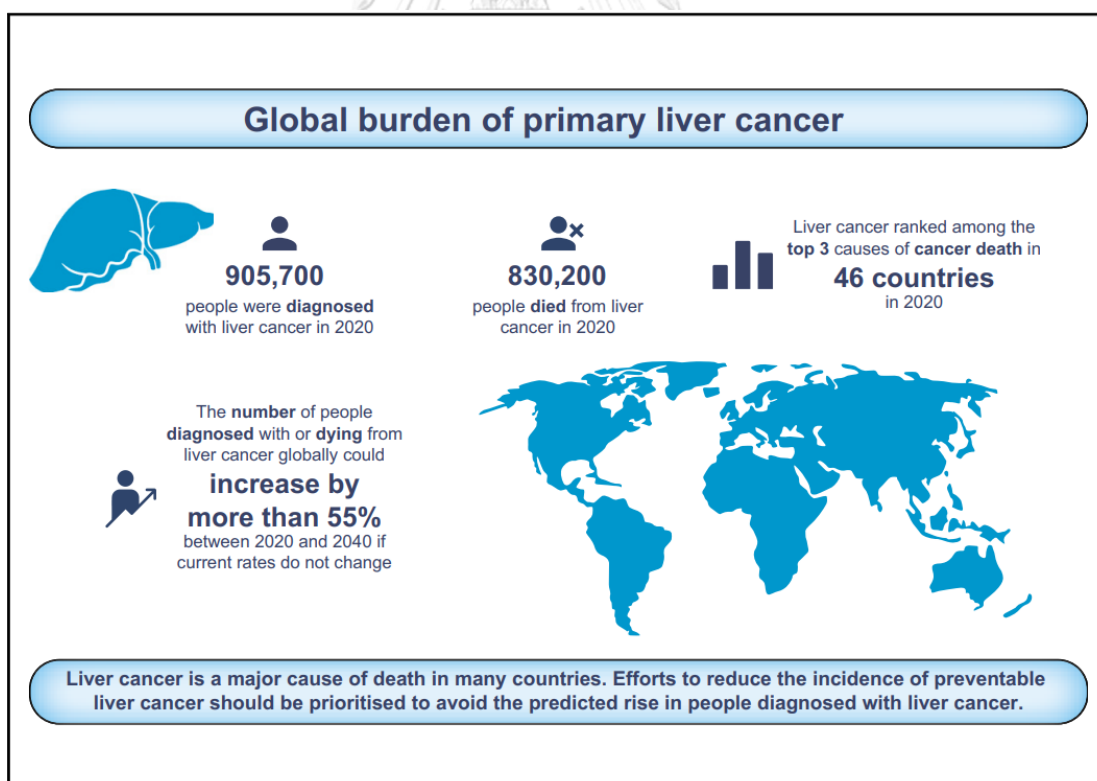
Figure 2: The graphical abstract of the global burden of primary liver cancer in 2020 and predictions to 2040 (Harriet et al., 2022).

Non-cancerous growths that can occur in the liver are known as benign liver tumors. These tumors do not spread to other parts of the body or invade tissues nearby. Most benign liver tumors do not produce symptoms and are frequently identified by chance during medical imaging for unrelated conditions. Some examples of benign liver cancers include Hemangiomas that made up of blood vessels, Hepatic adenomas develop from liver cells and Focal nodular hyperplasia (FNH) that mostly found in women. In contrast, malignant liver tumors are cancerous growths in the liver. Hepatocellular carcinoma (HCC), which develops from the liver's hepatocytes, is the most frequent type of malignant liver tumor. Cholangiocarcinoma (cancer of the bile ducts within the liver) and less common kinds such as angiosarcoma and fibrolamellar carcinoma are examples of malignant liver tumors (Society, 2019). Moreover, liver tumor classification is illustrated in Figure 3.
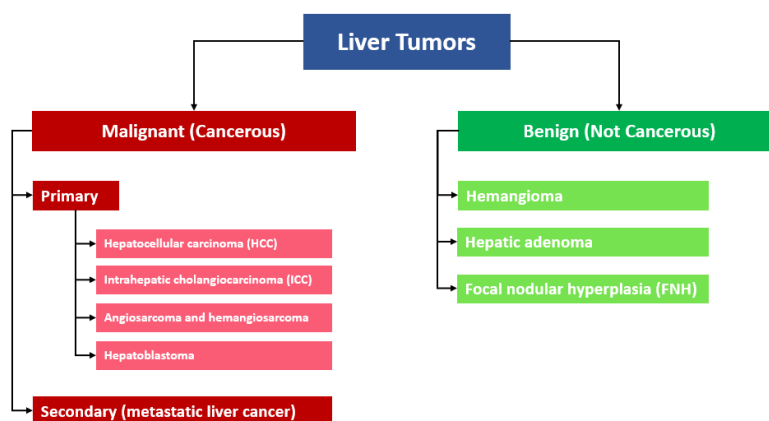
Figure 3: Liver tumor classification.

1.3. Radiological Imaging for Liver Cancers

Radiological imaging plays a vital role in diagnosing and treatment monitoring for liver diseases. They allow healthcare professionals to visualize the liver to identify abnormalities and diseases. Medical experts always recommend detecting liver diseases in earlier stages to achieve better survival from liver cancer. Recognizing liver diseases or liver cancers is challenging since liver diseases are often asymptomatic until the final stage. The most traditional way to diagnose liver cancer is liver biopsy. However, radiological imaging has become a crucial technical development to monitor liver health in a non-invasive manner.

1.3.1. Ultrasonography (US)

A hepatic ultrasound is a noninvasive examination that captures images of the liver and its blood arteries. It can aid in the diagnosis of a variety of conditions of the liver, including fatty liver, liver cancer, and gallstones. The scanning process utilizes sound waves to generate the ultrasound image. This scanning procedure is safer and usually does not require a longer time for the scanning process. At the preliminary stage of liver disease diagnosis, healthcare experts recommend a US liver scan. It is a comparatively cheaper and highly available liver scanning tool for various liver conditions, such as cancer, hepatitis, fatty liver disease, and cirrhosis (Hennedige & Venkatesh, 2013). US images can be obtained with more details by using contrast materials. Injecting contrast dye can improve the visualization of the liver structure. However, some people might get allergic reactions due to contrast materials in very rare cases (Today, 2022).

Figure 4: Ultrasonography (US) in liver diagnosis (Today, 2022).

1.3.2. Computed Tomography (CT)

Computed Tomography is a radiological imaging technique that combines with an X-ray imaging procedure. The X-ray is aimed at the patient and rotates around the body while capturing X-ray images. Once the number of X-rays (slices) were obtained, the computer system in CT stacked all the slices to generate a three-dimensional (3D) image of the patient that can visualize body structures to perform diagnosis. The CT scan shows detailed images or 3D visualization of body parts including the bone, muscles, fat, organs, and blood vessels (Medicine, 2023).

Figure 5: Computed Tomography (CT) in liver diagnosis.

CT scans are immensely useful to use as a diagnostic tool in life-threatening conditions such as blood clots, hemorrhage, and cancers. Early detection and faster diagnosis are highly important in such conditions to save lives. However, CT scan is developed based on X-ray technology that produces ionizing radiation. Ionizing radiation is not suitable for patients with kidney disease.

CT scan is capable of visualizing body structures that can differentiate dense body structures such as bones. However, the soft tissues of the organs might not appear very clearly in a detailed manner. In such situations, contrast agents can be used to enhance the visibility of soft tissues by injecting contrast-based iodine into the bloodstream. However, contrast agents may cause allergic reactions or temporary kidney failures in rare cases (Medicine, 2023).

1.3.3. Magnetic Resonance Imaging (MRI)

Magnetic Resonance Imaging can produce three-dimensional anatomical imaging that can help healthcare professionals diagnose disease. MRI is a sophisticated technique that can visualize the liver with a very high spatial and temporal resolution. Powerful magnetic that produces a strong magnetic field is employed to simulate protons in the body to generate the visualization of the anatomical structures (Bioengineering, 2022). Due to the strong magnetic field, patients with metal implants are not suitable to undergo MRI scans. To enhance the visualization, contrast agents are given intravenously before or during the MRI scans (Lencioni et al., 2004).

Figure 6: Magnetic Resonance Imaging (MRI) in liver diagnosis.

1.4. Artificial Intelligence in Liver Diagnosis

Researchers have explored developing segmentation methods using different deep learning approaches in an effort to overcome such difficulties. Convolutional neural networks (CNN), which have dramatically increased performance on a wide range of computer vision applications by automatically learning multi-level feature representations, have significantly advanced medical image analysis.
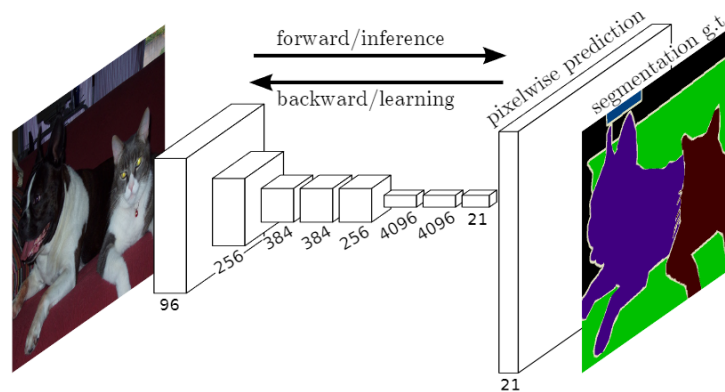
Figure 7: Graphical illustration of Fully convolutional network for semantic segmentation (Long et al., 2015).

Fully convolutional networks (FCN), which have developed quickly in deep learning, have displayed incredible pixel-level classification accuracy (Long et al., 2015). The formation of the FCN architecture is illustrated in Figure 7. The fully connected layer of the convolutional neural network follows a deconvolutional process with up-sampling to generate the pixel-level classification. Bilinear interpolation is used to up-sample the final pool layers before performing the fusion to generate the final prediction (See Figure 8).
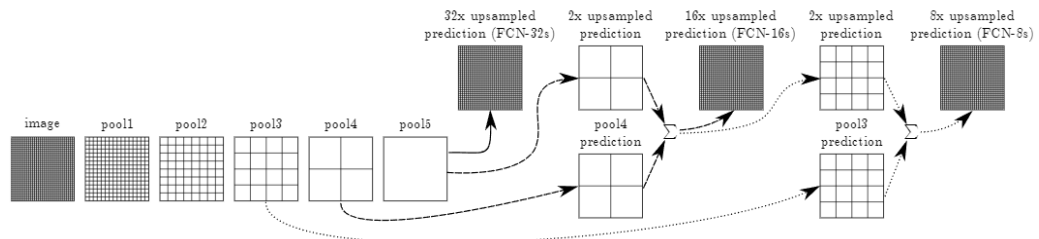
Figure 8: Fully convolutional network structure development (Long et al., 2015)

Ronneberger et al. (2015) proposed U-Net in 2015 based on the FCN and has demonstrated enormous potential for success in the segmentation of medical images. The U-Net architecture is designed with a contracting path and expansion path denoted by the encoder path and decoder path respectively (see Figure 9). Each path contains four blocks comprised of two convolutional layers and an activation function. In particular, the accuracy of the semantic segmentation depends on the contextual and location information. The skip connection is implemented in all the levels of the network that concatenate

corresponding location information in the encoder path and contextual information in the decoder path to get better segmentation performance with precise localization.
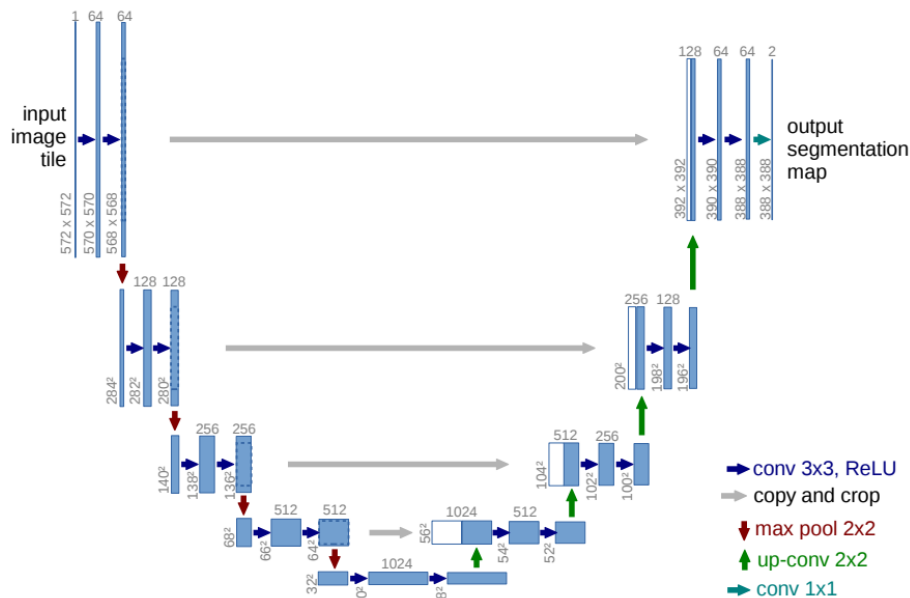
Figure 9: The design of the U-net architecture (Ronneberger et al., 2015).

In history, convolutional neural network (CNN) model developers commonly suggested that the performance of the model is directly proportional to the number of layers in the model (Krizhevsky et al., 2017; Zisserman, 2015). In real applications, the performance was not increased with the depth of the network since gradient vanishment degraded the performance of the network. To address the gradient vanishing issue, He et al. (2016) proposed a Residual Network (ResNet) with an identity map that can bypass convolutional layers in the block and combine with the final output of the convolutional block as illustrated in Figure 10. The residual learning concept that is designed with the bypass pathway concept could lessen errors in the learning process.
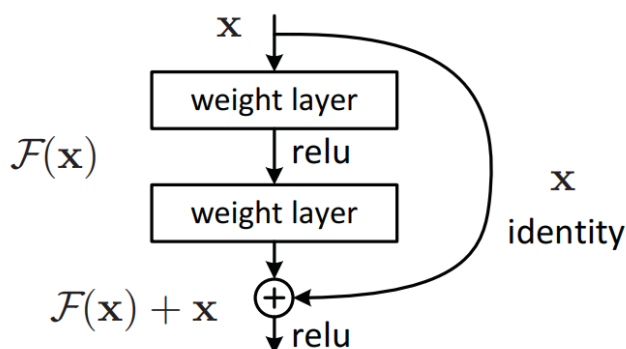
Figure 10: Residual learning concept: a building block (He et al., 2016).

The evolution of FCN-based architectures, H-DenseUnet (Li et al., 2018) and U-Net++ (Zhou et al., 2018) as the extended versions of the U-Net architecture. Furthermore, Zhang et al. (2018) proposed Deep Residual U-Net (ResUNet) which is constructed by combining the advantages of U-Net (Ronneberger et al., 2015) and deep residual learning (He et al., 2016). Residual learning is further utilized to improve the UNet-based networks (Devidas et al., 2023; Z. Li et al., 2022; Thomas et al., 2021; Xiwang et al., 2022). Z. Li et al. (2022) proposed Residual attention unet++. In the development, the UNet++ network is further developed utilizing the strengths of residual learning and spatial attention mechanism in order to decrease learning errors and improve the semantic gap between the encoder and decoder of the network. RA-UNet demonstrated an attention method that uses the max-pooling operation between the encoder and decoder to emphasize key features while decreasing noise (Jin et al., 2020). Residual connections are used to keep the original features while highlighting the salient ones.

Recently, ResUNet++ (Zhang et al., 2018) has been constructed as a more complex version of ResUNet. Dense predictions in multi-stage Cascaded CNNs serve to illustrate how well the majority of the aforementioned architectures work. This method uses unnecessary and redundant computational requirements for feature processing during the segmentation task. Researchers have developed spatial modules with attention methods to suppress unnecessary features while highlighting the most significant spatial information for the segmentation task.

Attention U-Net (Oktay et al., 2018) is a U-Net and end-to-end-trainable attention module (Jetley, 2018) based architecture proposed for image classification. Skip connection uses an attention method that Attention U-Net is designed to retrieve salient features for fusing with high-level semantic features. The attention mechanism is illustrated in Figure 11. The enhancement of the crucial features and reduction of redundant regions for the segmentation task could significantly improve learning.
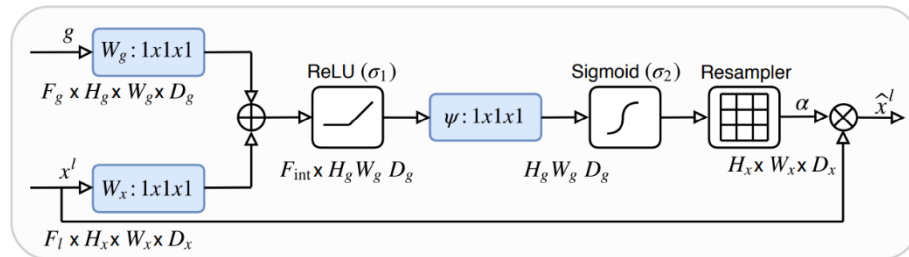
Figure 11: The proposed additive attention gate (AG) (Oktay et al., 2018)

Moreover, ResUNet++ architecture incorporates a squeeze-and-excitation attention mechanism (Hu et al., 2018) to recalibrate channels in each level of the encoder path. Woo et al. (2018) proposed an attention mechanism that is a combination of the channel attention mechanism and the spatial attention mechanism that creates the Convolutional Block Attention Module (CBAM) as illustrated in Figure 12. CBAM is a simple, lightweight technique that has been effectively applied in recent research developments (Chaoqun et al., 2021; Chen et al., 2020; Lee et al., 2020). It is very simple to include in neural networks. Also, researchers who created attention mechanisms have observed that attention mechanisms considerably increased the accuracy and sensitivity of the prediction with comparatively fewer parameters.
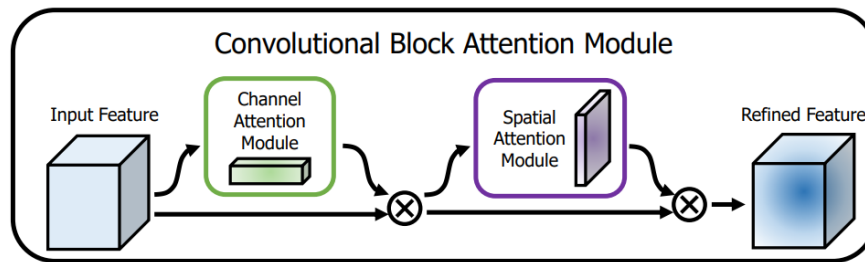
Figure 12: The CBAM overview. The module is divided into two sub-modules: channel and spatial. At each convolutional block of deep networks, the intermediate feature map is adaptively refined by the module (CBAM) (Woo et al., 2018)

Recent developments in fully convolutional networks still need performance improvement for liver tumors with fuzzy boundaries. There is a higher possibility of overfitting the training set and less generalizable when the model is heavy. To address these issues, we investigate the potential of attention mechanisms and residual learning by integrating it with U-Net-based architecture to improve the sensitivity of the model to recognize tumors with fuzzy boundaries. Furthermore, we investigate how well attention-based architecture can minimize the computational cost of the liver tumor segmentation task.

1.5. Objectives and Contribution

The objective of our study is to develop a Computer-Aided Diagnosis (CAD) tool to segment liver tumors from computed tomography (CT) imaging by utilizing less computational cost. Assess and validate the applicability of the developed liver tumor segmentation tool with a radiologist.

The contribution of the study is summarized as follows,

- We propose a novel semantic segmentation neural network design that exploits the strengths of residual blocks, channel attention, and spatial attention methods developed in CBAM. The spatial attention mechanism is implemented to extract spatial features from the encoder and combine them with corresponding high-level semantic features in the decoder. Channel attention and spatial attention are implemented in the encoder and decoder paths. The proposed design uses the U-Net architecture as the base for development.

- The integration of attention techniques in the encoder path, skip connection, and decoder path significantly improved the target segmentation's focus on the region of interest, obtaining the best sensitivity score in all experiments.

1.6. Scope of Study

The liver tumor segmentation tool is trained and evaluated with two publicly available datasets and the scope, and the constraints are as follows,

- The proposed model can correctly segment liver tumors compared to the baseline methods.

- The proposed model utilizes less computational cost compared to baseline methods.

- The study is conducted using scans (slices) with liver tumors.

2. Literature Review

2.1. Background

The automatic segmentation of liver tumors has become a popular topic in the field of deep learning-based medical research. Deep learning breakthroughs provide substantial contributions to the quality and accuracy of the diagnosis by providing quick and helpful opinions for therapeutic intervention. The recent advancements in machine learning technology in liver tumor segmentation have reached the diagnostic capabilities of radiologists (Kim et al., 2020; Zhen et al., 2020).

Kim et al. (2020) proposed a deep learning system to detect hepatocellular carcinoma (HCC) using contrast-enhanced magnetic resonance imaging (CE-MRI). They have designed and experimented with a simple convolution neural network (CNN) based architecture that is stacked with four convolutional layers as shown in Figure 13. The architecture is trained and tested with over 50,000 images obtained from 549 patients.
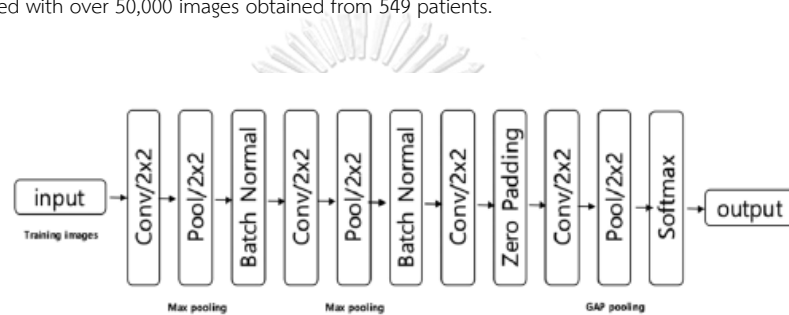
Figure 13: The CNN structure of the model for detecting HCC in MR images (Kim et al., 2020).

All the images are preprocessed by applying normalization. ROI is generated to recognize lesions to create the mask. Data augmentation includes rotating and shifting to increase the training samples and minimize the overfitting. The network training optimization is tested with four optimizers, Adam optimizer could demonstrate better performance in network optimization. The process of the deep learning system is illustrated in Figure 14. The deep learning system demonstrated 6 times faster HCC detection compared to junior radiologists.
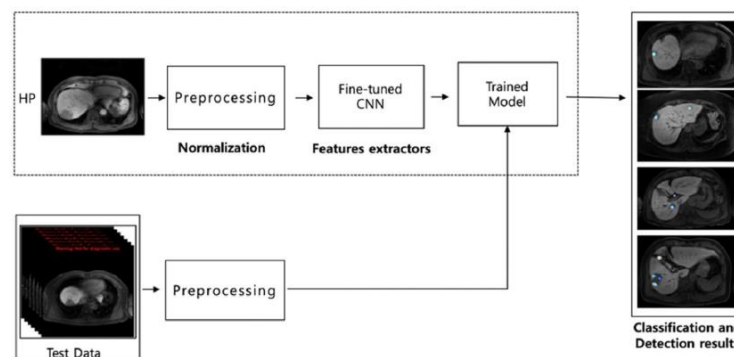
Figure 14: Proposed Deep Learning system for HCC detection (Kim et al., 2020).

A multi-phase contrast-enhanced magnetic resonance imaging (CE-MRI)-based computer-aided diagnosis method for identifying and classifying liver cancers proposed by (Alksas et al., 2021). Support vector machine (SVM), Naive Bayes

classifier (NB), k-nearest neighbors (KNN), and linear discriminant analysis (LDA) have all been used to evaluate the effectiveness of imaging markers. Spherical harmonics that can describe morphological complexity, are used to generate morphological markers. Functional markers are generated based on the characteristics enhanced in different phases in MRI. Textural markers that are considered histogram markers are calculated based on the gray-level co-occurrence matrix (GLCM) and gray-level run-length matrix (GLRLM) to recognize the texture patterns according to the liver tumor grades in the final classification. Random forest classifier demonstrated more enhanced diagnostic performance among other classifiers experimented in the study. The proposed computer-aided diagnosis (CAD) system is visualized in Figure 15.
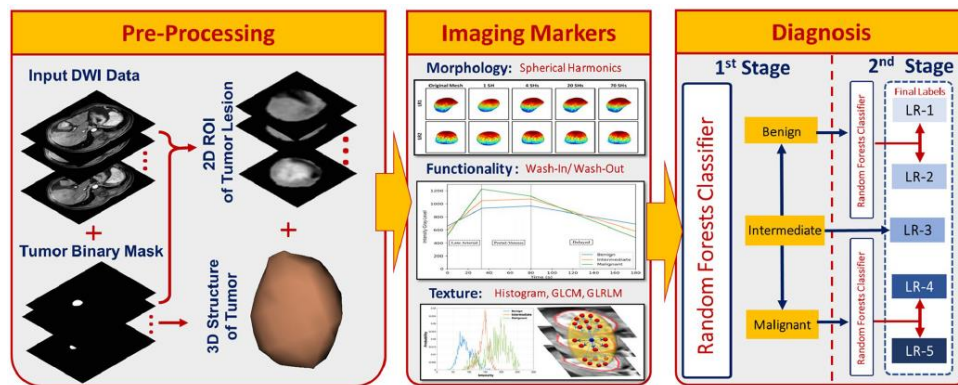
Figure 15: The computer-aided diagnostic (CAD) method for detecting and classifying cancers in the liver (Alksas et al., 2021).

Seven different types of liver lesions and clinical data were used in a large-scale deep learning-based investigation (Zhen et al., 2020). Seven models have tested their abilities to classify liver lesions using clinical information, unenhanced MRI, and enhanced MRI. Two models that were created using MRI imaging and clinical data demonstrated superiority in terms of diagnosis compared to experienced radiologists. The graphical summary of the study is illustrated in Figure 16.

(Hamm et al., 2019) proposed a deep learning system for the purpose of evaluating the viability of classifying liver lesions. It is a classifier with a multi-phasic MRI-based convolutional neural network. The designed custom CNN is comprised of three convolutional layers integrated with rectifier linear units, two maximum pooling layers, and two fully connected layers for final lesion classification. Imaging Reporting and Data System (LI-RADS), experts in radiology developed generic classification guidelines that were referred to in the study. They have obtained MRI scans from a heterogeneous collection of scanners and multi-phasic scans including the Arterial phase, Venous phase, and Equilibrium phase to train and evaluate the model. The sample image categorization followed by three MRI phases and based on LI-RADS categories is illustrated in Figure

17. They recommended using automatic categorization criteria based on deep learning for systems like LI-RADS in order to reduce the diversity in picture interpretation and improve the quality assurance of the research.
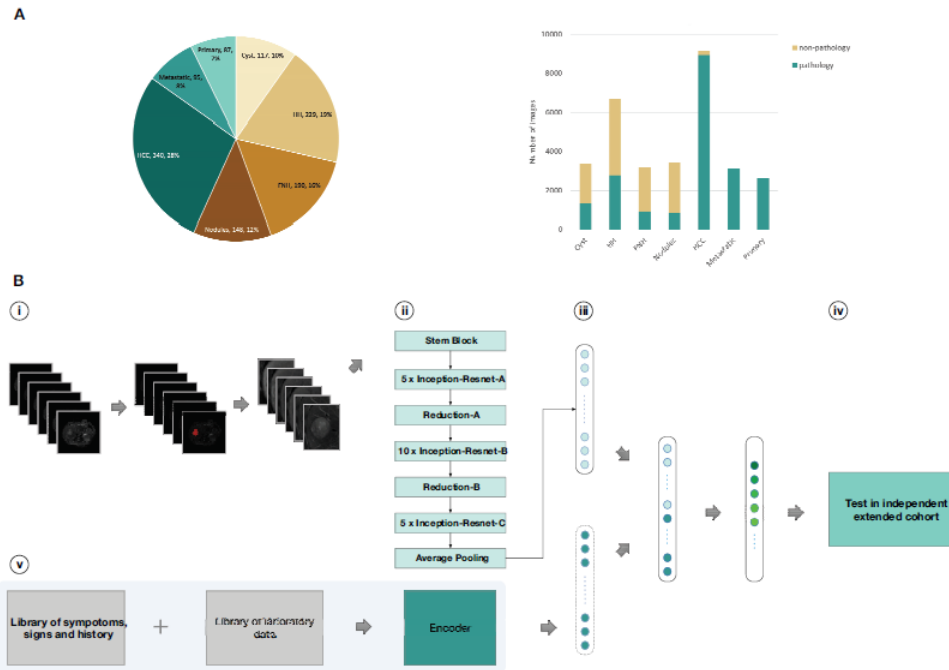
Figure 16: Overview of the study (Zhen et al., 2020).
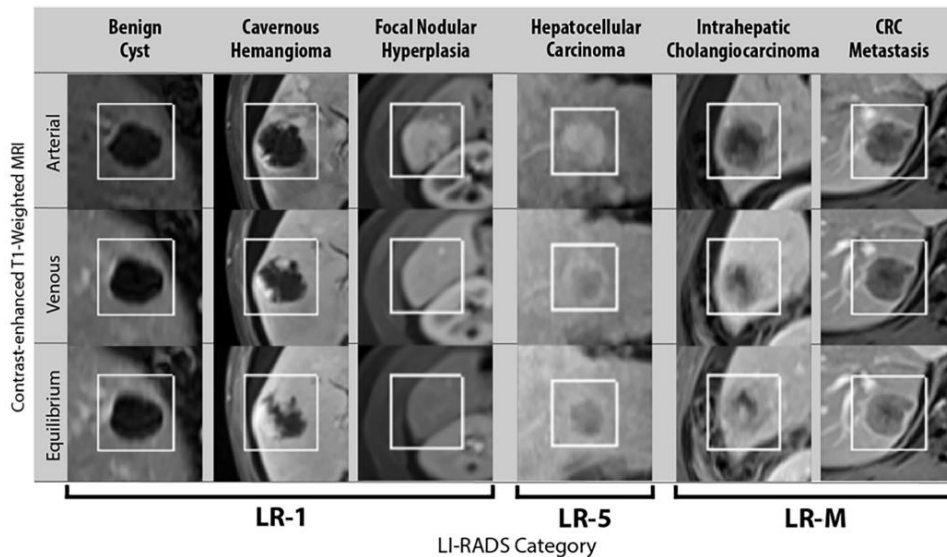
Figure 17: Sample images of lesion classes and corresponding derived LI-RADS categories (Hamm et al., 2019).

2.2. Fully Convolutional Network Based Approaches

Due to the creation of the fully convolutional network (FCN), the research area took a different turn away from deep learning-based classification (Long et al., 2015). End-to-end training using FCN pixel-level classification could produce segmentation output. U-Net architecture is becoming more popular in the field of medical research which is an extension of FCN. Several adaptations based on the U-Net architecture (He et al., 2016; Huimin Huang et al., 2020; Jha et al., 2019; Li et al., 2018; Zhou et al., 2018) have become popular deep learning architectures to segment liver and tumors.

Alirr (2020) proposed an architecture based on U-Net, an automated approach for segmenting tumors and the liver was suggested. Preprocessing procedures involve the use of HU windowing and median filtering and a tensor-based 3D edge enhancing diffusion (EED) to improve training data for the model training. Ayalew et al. (2021) developed a U-Net-based liver and tumor segmentation approach with parameter reduction. When compared to the original U-Net, the class imbalance method and data refinement techniques are used to improve segmentation performance while minimizing computing costs.

U-Net++ is designed with nested and dense skip connections (Zhou et al., 2018). UNet 3+ is a more advanced version with deep supervision at each stage of the decoder (Huimin Huang et al., 2020). B. Li et al. (2022) improved the UNet++ architecture design by implementing a channel attention mechanism to the long-hop connections. Channel attention could help to reduce eigenvalue loss. Moreover, CE-Net (Gu et al., 2019) and DefED-Net (Lei et al., 2022) architectures are developed to improve the feature representation using multi-scale feature extractors and deformable convolutions. However, both networks needed more parameters to improve performance and require comparatively higher computational power for the segmentation task.

2.3. Attention Mechanism Based Approaches

In recent developments in attention mechanisms for liver and tumor segmentation studies, Attention U-Net gained popularity which was proposed by Oktay et al. (2018). For the segmentation task, the attention mechanism investigates the potential to highlight critical information while suppressing unimportant ones in spatial dimensions. Attention UNet++ (Li et al., 2020) employs the same spatial attention approach for upgrading UNet++ architecture. The modified architecture shows a significantly superior focus on target regions while suppressing irrelevant regions. By using spatial attention gates to highlight important features, liver segmentation has improved recently by Wang et al. (2022).

Channel attention, in addition to the spatial attention process, is crucial for improved feature propagation. The mechanisms of global attention and hybrid attention are created to effectively focus on both local and global features of the segmentation (Huaxiang et al., 2023; Jiang et al., 2019). To improve the feature representation for liver and tumor segmentation, self-attention-based architecture is constructed (Fan et al., 2020). However, the encoder path and skip connections of the network are not improved by implementing feature recalibration utilizing the strengths of the attention mechanisms.

Lately, the inter-channel connection has been created utilizing global average-pooled features to determine channel-wise attention. Squeeze-and-excitation (Hu et al., 2018) is a method designed to take advantage of this relationship. In order to boost its sensitivity to important features while suppressing irrelevant features, several deep learning architectures have successfully segmented medical images using the squeeze-and-excitation module (Jha et al., 2019; Wang et al., 2022). The SE module is incorporated into MS-UNet's design to enhance channel-wise feature recalibration (Devidas & Sanjay, 2021). HFRU-Net is designed by implementing an SE module in skip connection to recalibrate encoder features to fuse with deep features in the decoder (Devidas & Sanjay, 2022). Eventually, Woo et al. (2018) proposed and demonstrated better channel

attention utilizing average-pooling and max-pooling, which is named Convolutional Block Attention Module (CBAM), a sequential combination of channel attention with spatial attention.

Shuchao et al. (2021) developed TA-Net, which improves performance in medical image segmentation by achieving improved feature representation through the use of multiple deep learning techniques such as inception blocks, context blocks, and attention blocks. When employed in the shallow feature extraction path and the deep feature extraction path independently, they have found that channel attention with both average pooling and max pooling performs better than channel attention with merely average pooling. In order to highlight significant features in input-level shallow features and semantically high-level features in output, Zhao et al. (2020) used CBAM by separating it to channel attention for the extraction of deep features at the bottom of the architecture and spatial attention for both encoder and decoder at the top level. CBAM is used in Small Attention-UNet (SmaAt-UNet) to skip connection and bottleneck of the network. Important encoder features are amplified using CBAM in order to concatenate them with deep features in the decoder (Kevin et al., 2021). The depthwise-separable convolutions have significantly contributed to minimizing the training parameters of the network. Additionally, S-Net uses CBAM to increase spatial features and optimize channel weights at a deeper layer of the network, resulting in a notable improvement in liver tumor segmentation (Luan et al., 2021).

3. Proposed Method

In this chapter, we present a novel network named Multi Attention Network (MANet) for the liver tumor segmentation task. The network is developed as a fusion of attention mechanisms to extract important features while suppressing irrelevant features for liver tumor segmentation. The proposed deep learning network has followed U-Net as the basic architecture. Moreover, the residual learning and attention mechanisms are implemented in the network.

3.1. MANet Architecture

We present the MANet design, which has been tested for segmenting liver tumors. Figure 18 displays the proposed architecture's block diagram. The Table 1 illustrates the structure of the proposed network. By exploiting the benefits of attention mechanisms (Oktay et al., 2018; Woo et al., 2018) and deep residual learning (He et al., 2016), MANet architecture is an upgraded version of U-Net (Ronneberger et al., 2015). Encoder, bridge, and decoder are the primary elements of architecture. The encoder utilizes channel attention and residual blocks to propagate information by the input image. In order to calculate semantic segmentation, the decoder creates pixel-wise classifications. The deepest level feature propagation across the residual unit, channel, and spatial level is determined in the bridge.

We have used deep residual learning to construct encoder residual blocks to solve the degradation issue in the network since deep neural networks often suffer from degradation due to erroneous feature learning. The encoder residual block is coupled with two convolution layers, batch normalization, and Rectified Linear Unit (ReLU) activation, as can be seen in the diagram. To overcome the gradient vanishing issue and accelerate the network's convergence, ReLU activation is used after batch normalization. Each encoder block's feature propagation output is sent to the channel attention module, which calibrates the channel weights for a more advantageous inter-channel relationship to improve semantic feature extraction. The first convolution layer in each residual block is applied with a stride of 2, which is used for the spatial dimension down-sampling procedure. Using the gate signal, which is obtained by deep features in the lower stage, the spatial attention mechanism is used in the skip connection to extract significant shallow features. Transposed convolution is used to up-sample lower-stage deep features, which are then combined with matching shallow features that were retrieved using the attention mechanism.

Spatial attention is applied to the input of the decoder block in order to highlight semantic information and keep it throughout the decoding process. Batch normalization and ReLU activation are implemented after each convolution layer in decoder blocks, which is similar to the encoder block. The decoder block uses two succussive convolution layers with 3 × 3 kernels for feature propagation. The segmentation output was generated using sigmoid activation after 1 × 1 convolution to the output of the decoder path.
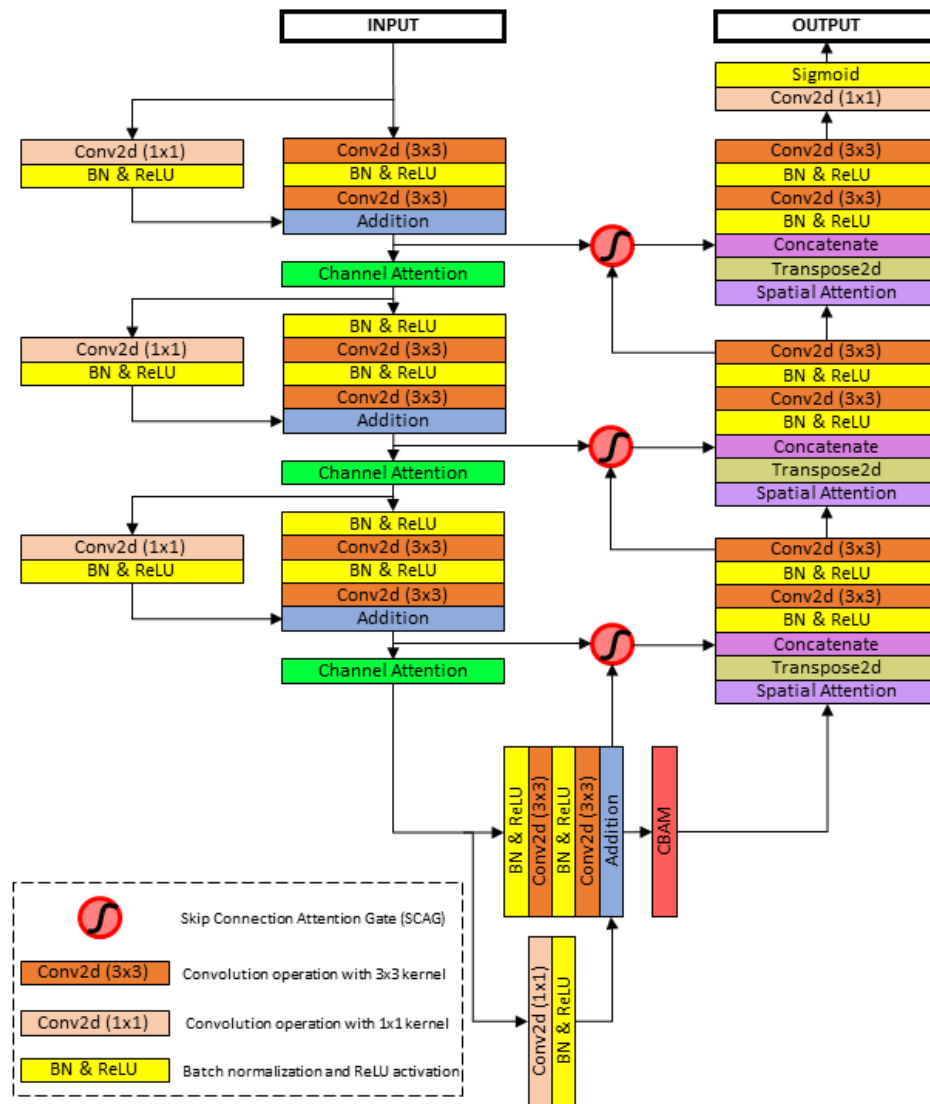
Figure 18: Block diagram of the proposed MANet architecture.

Block name	Operation	Filter size	Number of filters	Stride	Output size
Input Image					512 × 512 × 3
Encoder 1	Conv 1	3 × 3	68	1	512 × 512 × 68
	Conv 2	3 × 3	68	1	512 × 512 × 68
Encoder 2	Conv 3	3 × 3	136	2	256 × 256 × 136
	Conv 4	3 × 3	136	1	256 × 256 × 136
Encoder 3	Conv 5	3 × 3	272	2	128 × 128 × 272
	Conv 6	3 × 3	272	1	128 × 128 × 272
Bridge	Conv 7	3 × 3	544	2	64 × 64 × 544
	Conv 8	3 × 3	544	1	64 × 64 × 544
Decoder 1	Conv 9	3 × 3	136	1	128 × 128 × 136
	Conv 10	3 × 3	136	1	128 × 128 × 136
Decoder 2	Conv 11	3 × 3	68	1	256 × 256 × 68
	Conv 12	3 × 3	68	1	256 × 256 × 68
Decoder 3	Conv 13	3 × 3	68	1	512 × 512 × 68
	Conv 14	3 × 3	34	1	512 × 512 × 34
Output	Conv 15	1 × 1	3	1	512 × 512 × 3

Table 1: The network structure of the proposed MANet architecture.

3.2. U-Net and Residual Blocks

To improve segmentation performance in semantic segmentation, high-resolution low-level features, and high-level semantic features must be combined (Long et al., 2015; Ronneberger et al., 2015). The U-Net skip connection, which is used at each level of the network, has the potential to improve segmentation performance and be successful in the field of medical image segmentation. In deep residual learning (He et al., 2016), which suggested reducing the training errors in deep neural networks, it was further emphasized that using skip connections at each level of the network improves information propagation without degradation. And confirmed with state-of-the-art approaches (Jha et al., 2019; Zhang et al., 2018). We created an encoder using residual blocks, which consists of two 3 × 3 convolution blocks and one residual connection, inspired by residual learning. Because of the memory restrictions, the input and output of the residual block are added using a convolution block with a 1 × 1 kernel applied to control channels. In the residual block with skip connection, batch normalization, and ReLU activation are used to reduce performance deterioration, and gradient vanishment, and help accelerate the propagation of features.

3.3. Attention Mechanisms

Attention processes are crucial in the segmentation task because they enable the extraction of more accurate contextual information. We presented a novel MANet that was inspired by two attention methodologies (Oktay et al., 2018; Woo et al., 2018). By using a very limited number of parameters, attention mechanisms enable the potential to improve feature representations. Generally, there are two types of attention mechanisms: channel attention and spatial attention. The statistical weight of each channel is determined by global average pooling, which is performed through channel attention. The global pooling operation across the channel dimension is performed by spatial attention to retrieve contextual data. Moreover, channel attention instructs the network to concentrate on "what" salient features to represent, whereas spatial attention investigates "where" significant elements are situated in the feature map. Spatial attention, channel attention, skip connection attention gate, and convolutional block attention module (CBAM) are the four attention mechanisms included in the proposed MANet.
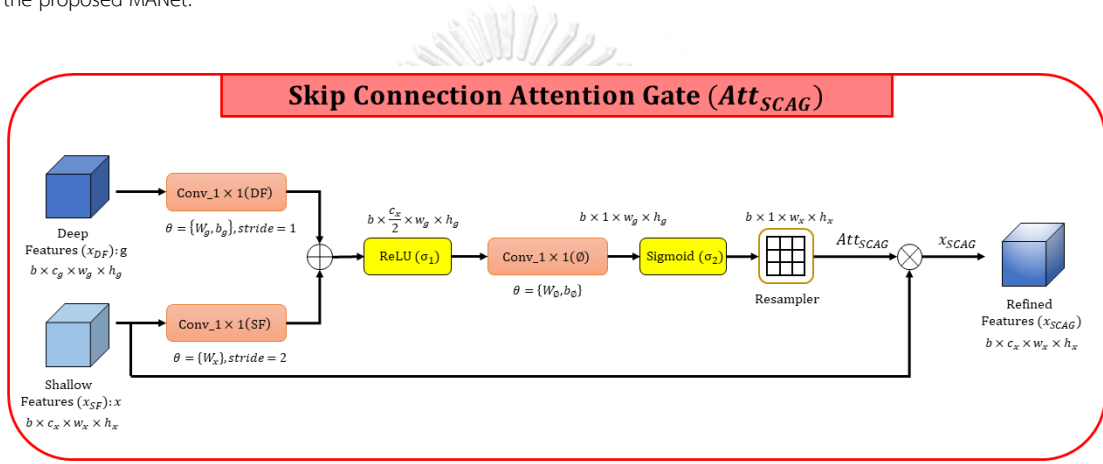
Figure 19: Schematic diagram of Skip Connection Attention Gate (SCAG).

The skip connection attention gate is aimed at capturing crucial shallow encoder features and concatenating them with semantically high-level decoder information. The attention is calculated by combining shallow features x_{SF} and deep features x_{DF} from the network's previous decoder block. Figure 19 shows the block diagram of the skip connection attention gate, which can be stated as follows:

Equation 1

$$x_{conv\_1\times1(SF)} = W_x . x_{SF} \, , \, x_{SF} \, \epsilon \, \mathbb{R}^{b\times c_x\times w_x\times h_x}, x_{conv\_1\times1(SF)} \, \epsilon \, \mathbb{R}^{b\times c_x/2\times w_g\times h_g}$$

$$x_{conv\_1\times1(DF)} = W_g . x_{DF} + b_g \, , \, x_{DF} \, \epsilon \, \mathbb{R}^{b\times c_g\times w_g\times h_g}, x_{conv\_1\times1(DF)} \, \epsilon \, \mathbb{R}^{b\times c_x/2\times w_g\times h_g}$$

$$Att_{SCAG}(x_{SF}, x_{DF}; \, \theta_{SCAG}) = \sigma_2(W_\emptyset . \sigma_1(x_{conv\_1\times1(SF)} + x_{conv\_1\times1(DF)}) + b_\emptyset)$$

Thus, in order to produce the final attention map, σ_1 represents for the ReLU activation function and σ_2 for the sigmoid activation function. The number of channels is c, the size of the feature maps is $w \times h,$ and the batch size is b. The parameters for the attention mechanism $(Att_{SCAG}(x_{SF}, x_{DF}; \, \theta_{SCAG}))$ are $\theta_{SCAG} = \{W_x, W_g, b_g, W_\emptyset, b_\emptyset\}$, where W and b represent the weights and bias terms of the convolutions respectively. As stated in Equation 1, input feature maps (x_{SF}, x_{DF}) are linearly mapped to half of the shallow feature mappings $(\mathbb{R}^{b\times c_x/2\times w_g\times h_g})$ in the dimensional space. The refined feature representation x_{SCAG} is formulated as follows:

Equation 2

$$x_{SCAG} = x_{SF} \otimes Att_{SCAG}(x_{SF}, x_{DF}; \theta_{SCAG}), x_{SCAG} \in \mathbb{R}^{b \times c_x \times w_x \times h_x}$$

Where, element-wise multiplication denotes by $\otimes$
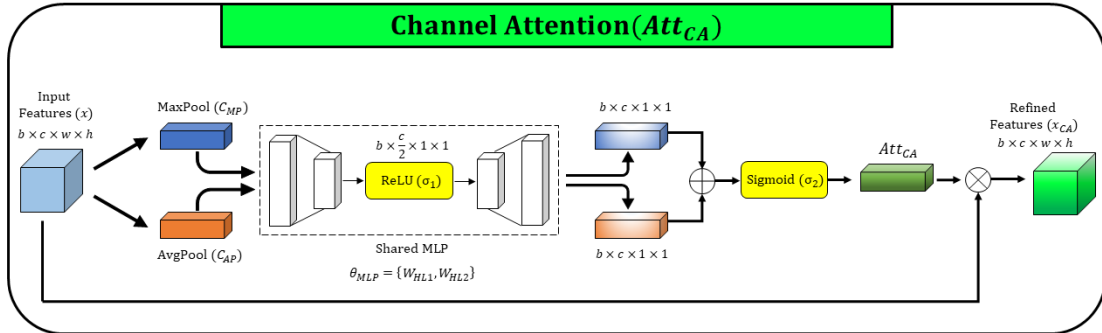
Figure 20: Schematic diagram of Channel Attention (CA).

The inter-channel relationship is captured by channel attention, which then recalibrates it to improve segmentation performance. When compared to the "Squeeze and Excitation" channel attention technique, which is based solely on global average pooling, the channel attention that comprises both global max pooling and global average pooling demonstrated better performance. The feature fusion of Max-pooling and average-pooling improves inter-channel interactions compared to average-pooling feature extraction alone when it comes to channel recalibration (Shuchao et al., 2021; Woo et al., 2018). Due to the benefits, we employ the channel attention mechanism in the proposed architecture, which is strengthened by average pooling and max-pooling techniques. As shown in Figure 20, input feature maps (x) are subjected to global pooling procedures to determine the global maximum and average pooling feature descriptors C_{MP} and C_{AP} respectively $(C_{MP}, C_{AP} \in \mathbb{R}^{b \times c \times 1 \times 1})$, formulated in Equation 3.

Equation 3

$$C_{MP} = MaxPool(x)$$

$$C_{AP} = AvgPool(x)$$

The result of the global polling operations is used to create a shared multi-layer perceptron that captures channel-wise correlation (MLP). MLP is built using two hidden layers and ReLU, which was formulated in Equation 5. The first hidden layer's output size is set to half of the input channels $(\mathbb{R}^{b \times c/2 \times 1 \times 1})$ in order to decrease the total amount of parameters. Then, the summation of feature descriptors generated through MLP is subjected to sigmoid activation (σ_2).

Equation 4

$$Att_{CA}(x; \theta_{CA}) = \sigma_2(MLP(C_{MP}) + MLP(C_{AP}))$$

Where MLP is formulated as follows,

Equation 5

$$MLP(x; \theta_{MLP}) = W_{HL2}.\sigma_1(W_{HL1}.x)$$

The formula for the channel attention mechanism is Equation 4, where σ_1 and σ_2 stand for the ReLU and sigmoid activation functions respectively. The channel attention parameters are W_{HL1} and W_{HL2} $(\theta_{CA} = \theta_{MLP} =$

$\{W_{HL1}, W_{HL2}\})$, which are used in two hidden layers of the MLP. Following the channel attention computation, element-wise multiplication is used to determine the calibrated feature representation x_{CA} (see Equation 6).

Equation 6

$$x_{CA} = x \otimes Att_{CA}(x; \theta_{CA})$$

$$x_{CA} \in \mathbb{R}^{b \times c \times w \times h}$$
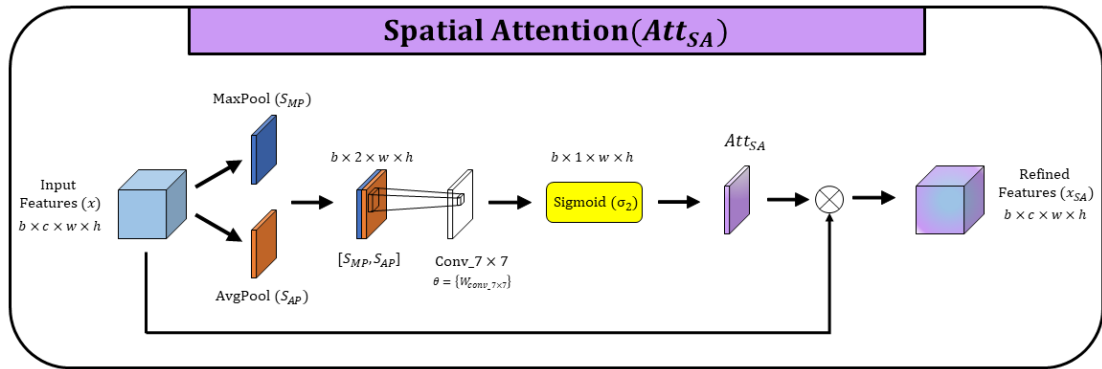
Figure 21: Schematic diagram of Spatial Attention (SA).

In order to improve segmentation performance, spatial attention is designed to gather significant spatial information. To take advantage of significant tumor features while suppressing non-tumor features in the decoder path, a spatial attention technique is used in the decoder block. To determine the spatial feature descriptors $S_{MP}, S_{AP} \in \mathbb{R}^{b \times 1 \times w \times h}$ of global max pooling and average pooling are carried out together with the channel axis for the input features (x), as shown in Figure 21. The two spatial feature maps are then concatenated. The contextual tumor feature aggregation across the spatial locations is represented by those feature maps with two channels. To further extract crucial contextual data that is highly relevant to segmenting liver tumors, convolution with a 7 × 7 kernel is performed. The spatial attention map produced by the 7 × 7 convolution layer is later subjected to the sigmoid activation function (σ_2). As seen below, a spatial attention mechanism has been formulated in Equation 7.

Equation 7

$$Att_{SA}(x; \theta_{SA}) = \sigma_2(W_{conv\_7 \times 7}.([S_{MP}, S_{AP}]))$$

Where $([S_{MP}, S_{AP}])$ stands for the concatenation of global maximum and average pools, and $\theta_{SA} = \{W_{conv\_7 \times 7}\}$ represents the parameters of the convolution operation in the spatial attention mechanism. By elementwise multiplying the spatial attention map produced by $Att_{SP}(x; \theta_{SP})$ with the input feature maps, features that correlate to the attention feature map are recalibrated. The following formula represents the enhanced feature representation x_{SA} (see Equation 8).

Equation 8

$$x_{SA} = x \otimes Att_{SA}(x; \theta_{SA})$$

$$x_{SA} \in \mathbb{R}^{b \times c \times w \times h}$$

Convolutional Block Attention Module(Att_{CBAM})
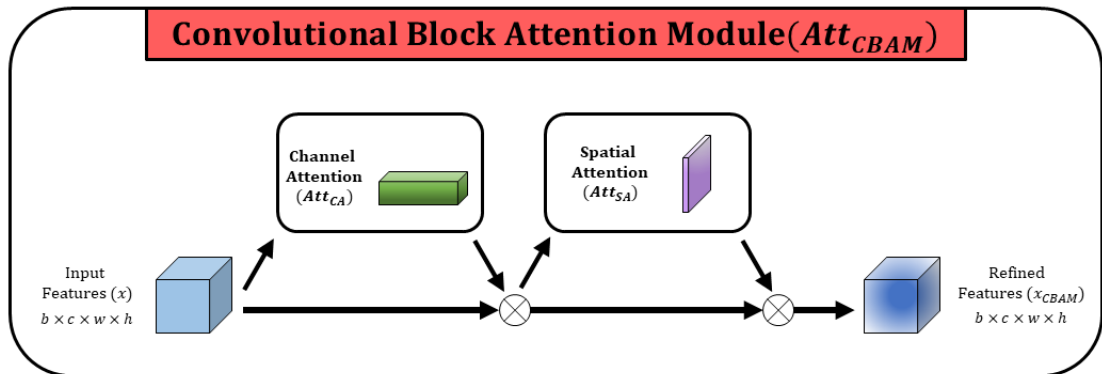
Figure 22: Schematic diagram of Convolutional Block Attention Module (CBAM).

The Convolutional Block Attention Module (CBAM) is created by sequentially applying 1D channel attention and 2D spatial attention illustrated in Figure 22. In the proposed network, CBAM is used in the bridge that connects the network's encoder path and decoder path. In order to improve segmentation performance, CBAM is used at the deepest level of the network to extract significant features in the channel dimension and sequentially acquire spatial feature representations in the spatial dimension. Channel attention (see Equation 4) and spatial attention (see Equation 7) are the basis for the calibrated feature maps x_{CBAM}, which is the convolutional block attention mechanism formulated in Equation 9.

Equation 9

$$x_{CBAM} = x \otimes Att_{CA}(x; \theta_{CA}) \otimes Att_{SA}(x \otimes Att_{CA}(x; \theta_{CA}); \theta_{SA})$$

$$x_{CBAM} \in \mathbb{R}^{b \times c \times w \times h}$$

4. Experimental Setup

4.1. Dataset and Preprocessing

The proposed architecture is tested and evaluated with the LiTS17 dataset, which is a publicly available dataset used for the MICCAI 2017 Liver Tumor Segmentation challenge (Bilic et al., 2023) and 3DIRCADb dataset (Luc Soler, 2010). The LiTS17 dataset contains 131 CT scans in the training set and 70 CT scans in the testing set, however, ground truth is available only for the 131 CT scans which are under the training set. The dataset is gathered from various clinical sites around the world. The collection includes a variety of liver tumor diseases that were found using various CT scanners. Each clinical location has trained radiologists to annotate segmentation masks, and three more experienced radiologists then confirm these annotations. The CT scan volume has slices with an image size of 512 × 512 and a range of 42 to 1026. The 3DIRCADb dataset consists of 20 CT volumes with liver tumors in 15 CT volumes. The LiTS dataset is created by including 20 CT volumes of the 3DIRCADb dataset, those from volume 28 to volume 47 (Jiang et al., 2019). There are typically 0 to 75 tumors that appear in the scan, ranging in size from $38\ mm^3$ to $349\ cm^3$.

The CT scan slices use a wide range of intensity values, from -1000 to 3000, to represent various organs and regions. The image intensities of all scans are clipped to the range of [-150, 250] Hounsfield Units (HU) to enhance the liver area from the abdominal scan and then follow histogram equalization and normalization to feed into the training procedure. CT slices with tumor annotations are chosen for the experiments in order to assess the effectiveness of the proposed liver tumor segmentation approach. 130 CT scans (7050 slices) in total have been utilized for the experiments, with one scan being discarded due to abnormalities. To assess the models, we have carried out several studies based on volumes and slices. For the training set and test set, all the data were randomly divided into a 4:1 ratio. All of the slices in the slice-based experiment were randomly divided into 1410 slices for the test set and 5640 slices for the training set. All the scans were randomly divided into 104 volumes (5408 slices) for the training set and 26 volumes (1642 slices) for the test set to conduct the volume-based experiment. The training and test sets for the volume-based experiment include 8 and 7 volumes, respectively, of the 3DIRCADb dataset. Due to the limited amount of biomedical data, we limited the data split only to the training set and test set, excluding the validation set.

A real-time data augmentation method Albumentations (Buslaev et al., 2020) is employed, which randomly changes the batch of data without increasing the number of slices, to minimize the possibility of overfitting. There is no data duplication among training processes in different epochs thanks to the random transformations. During the random transformations, operations including vertical flip, shift, scale, and rotation are used. The process flow of the proposed liver tumor segmentation method is illustrated in Figure 23.
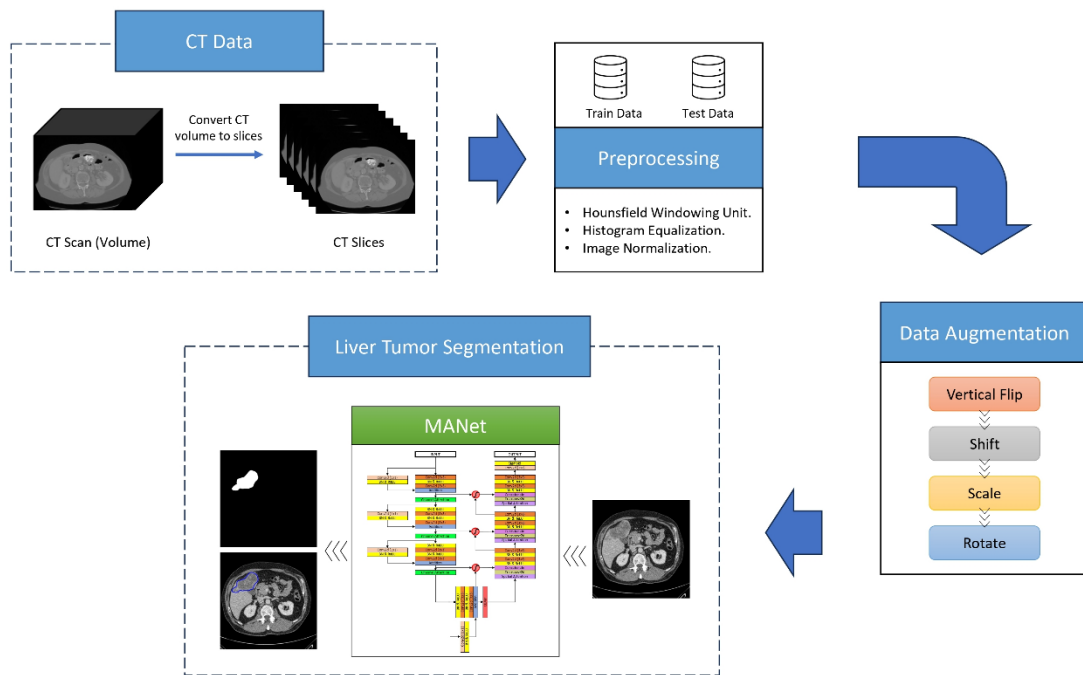
Figure 23: Flow diagram of the liver tumor segmentation method.

4.2. Implementation Details

We conducted all the experiments on a workstation that operates Windows 11. Furthermore, the workstation consists of RTX2070 GPU with 8 GB memory, 32GB of RAM, and Intel(R) Core (TM) i7-9750H CPU @ 2.60GHz 2.59 GHz (6 cores). PyTorch 1.9 is the deep learning framework we executed all of the experiments.

4.3. Evaluation Metrics

The seven most common evaluation metrics are computed to accurately assess the tumor segmentation performance of the experimental models. In general, the two categories of selected evaluation metrics are overlap-based methods and boundary-distance-based approaches. One of the most popular evaluation metrics is the Dice score (also known as the F1 score), while the Jaccard index is also known as the intersection over union (IoU), the volume overlap error (VOE) is the corresponding error metric for the Jaccard index (1 - Jaccard index), and accuracy, sensitivity (recall), and specificity are methods that use overlap. The average symmetric surface distance (ASSD), which is determined using a boundary-distance-based technique, is the average distance between points in the predicted binary mask and the ground truth binary mask (Yeghiazaryan & Voiculescu, 2018). The formulation of the evaluation metrics is as follows:

$$\text{Dice score} = \frac{2|A \cap B|}{|A|+|B|} = \frac{2TP}{2TP+FP+FN}$$

$$\text{Jaccard index} = \text{IoU} = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{TP+FP+FN}$$

$$\text{VOE} = 1 - \frac{|A \cap B|}{|A \cup B|} = 1 \text{ - Jaccard index}$$

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

$$\text{Sensitivity (Recall)} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{TN+FP}$$

$$\text{ASSD} = \frac{\sum_{x \in \partial B} \text{d}(x, \partial A) + \sum_{y \in \partial A} \text{d}(y, \partial B)}{|\partial B| + |\partial A|}$$

Where A and B, respectively, stand for the predicted binary mask and the ground truth binary mask. True positives, true negatives, false positives, and false negatives are represented by the TP, TN, FP, and FN, respectively.

5. Experiments, Results and Discussion

The proposed architecture is evaluated with several experiments described in the Experiments section. The novel architecture is contrasted among four other comparison models based on the UNet architecture. In addition, the proposed model is compared with state-of-the-art methods with the same experimental setup. All the experiments were conducted with CT imaging data of the hepatobiliary phase. Moreover, an experienced radiologist validated the proposed architecture's segmentation performance.

5.1. Experiments

The proposed architecture is designed based on UNet architecture. To evaluate the effectiveness of the architectural design, UNet based comparison architectures are used. Those architectures were created with the techniques and methodologies utilized to develop the proposed architecture. Quantitative analysis is conducted to analyze the proposed model among comparison methods and state-of-the-art methods. Mainly, two experiments are conducted based on volumes and slices of CT scans. In the first experiment, CT volumes are randomly split into a train set and a test set to evaluate the performance of the liver tumor segmentation. Secondly, slices of the CT scans are split into a train set and test set to assess the liver tumor segmentation performance. These experiments evaluated the segmentation performance under the seven most common evaluation metrics explained in the Evaluation Metrics section under the Experimental Setup chapter. The result of the quantitative analysis is discussed in the Quantitative Analysis of Segmentation Performance section.

5.2. Results and Discussion

Quantitative experiment is conducted with two methods, volume-based segmentation, and slice-based segmentation. Seven common evaluation metrics are used to evaluate segmentation performance. The result of the qualitative analysis is discussed in the Quantitative Analysis of Segmentation Performance section. The quantitative analysis is further continued in two perspectives. Performance analysis is conducted based on the number of tumors and the total area of tumors in the CT slice which are discussed in Quantitative Analysis based on the Number of Tumors section and Quantitative Analysis based on the Total Area of Tumors section respectively.

The qualitative analysis provides crucial insights into the segmentation performance of the proposed model. Segmentation output of the proposed model is visualized with comparison models used in the study. Liver tumor segmentation is classified into four main categories based on the size of the tumor and performance of the segmentation, those are large tumors, small tumors, poor segmentations, and over/non-segmentations. The result of the qualitative analysis is discussed in the Qualitative Analysis of Segmentation Mask section.

To validate the segmentation performance of the model, some of the samples with abnormal tumor labels or significant deviations compared to other tumor labels were selected to further verify with an experienced radiologist. The detailed discussion with an experienced radiologist is included in Liver Tumor Segmentation Validation with a Radiologist section.

Feature maps of the proposed network are visualized to demonstrate the effectiveness of the attention mechanisms applied in the network. Feature visualization is further continued for the comparison networks to compare the effectiveness of the proposed network. A detailed discussion of the feature visualization is explained in the Model Feature Visualization section. The computational cost of the proposed network is highlighted among comparison networks by emphasizing segmentation performance and computational cost with computational complexity, total parameters, and

Inference time under the Computational Cost Analysis section. Ultimately, the ablation study is conducted in 8 steps to demonstrate the effectiveness of the proposed design. A detailed discussion of the ablation study is in the Ablation Analysis section.

5.3. Quantitative Analysis of Segmentation Performance

The proposed MANet architecture is compared with the comparison models under seven evaluation metrics that are commonly used in liver tumor segmentation. The formation of the proposed architecture is strengthened with attention mechanisms and residual learning that can be considered as seven convolutional blocks architecture. The design of the architecture is optimized by setting the depth of the architecture to three, to minimize the parameter utilization and the complexity of the architecture. We have conducted experiments with four comparison networks, including three comparison networks with the same depth as the proposed network. Attention UNet, UNet+Resnet18, and UNet+CBAM are comparison networks with the same depth (7 blocks). UNet is the only comparison network that is deeper than the proposed network design. The original UNet architecture consists of nine convolutional blocks and the depth of the network is five. UNet architecture is the baseline for all the networks experimented on in the study.

Furthermore, the Attention UNet is designed by following the UNet architecture, and the spatial attention mechanism is implemented in skip connection to extract low-level shallow features that are enriched with localization information by following high-level contextual features in the decoder path. The UNet+Resnet18 is formed using the Resnet18 backbone with the UNet architecture. The UNet architecture is integrated with CBAM to each stage of the network as the proposed architecture is designed with CBAM and the submodules (channel attention: Att_{CA} and spatial attention: Att_{SA}) of it, to create the UNet+CBAM comparison network.

We contrasted the effectiveness and robustness of the proposed network to existing state-of-the-art approaches. The UNet architecture is the basis of all the comparison networks. In the comparison models, UNet 3+ is the most recent advancement based on the UNet architecture. Other comparison networks are created by combining the advantages of attention mechanisms with multi-level feature extraction methods. Other comparative architectures used to compare the overall performance of the proposed network include ResUNet++, SmaAt-UNet, and TA-Net. For the purpose of carrying out the experiments, we referred to original articles and baseline architecture codes.

5.3.1. Quantitative Analysis of Comparison Networks

Two experiments were conducted to assess the performance of the liver tumor segmentation of the proposed model and the comparison models including the based model UNet. The LiTS dataset was used to conduct volume-based and slice-based segmentation experiments. Furthermore, volume-based segmentation was conducted with the 3DIRCADb dataset to prove the generalizability of the network. The segmentation performance was evaluated under the seven most common evaluation metrics used in liver tumor segmentation, as explained in the Experimental setup section. Table 2 and Table 4 show the quantitative analysis results for the liver tumor segmentation on the LiTS dataset and 3DIRCADb dataset respectively. The test performance in terms of the dice score of the proposed MANet model and the baseline models on the test set for 80 epochs of the training process is illustrated in Figure 24. To visualize the performance variation of the proposed model and the baseline models, the evaluation results of volume-based and slice-based segmentation experiments are illustrated in Figure 25 and Figure 26 respectively.

5.3.1.1. Quantitative Analysis with the LiTS Dataset
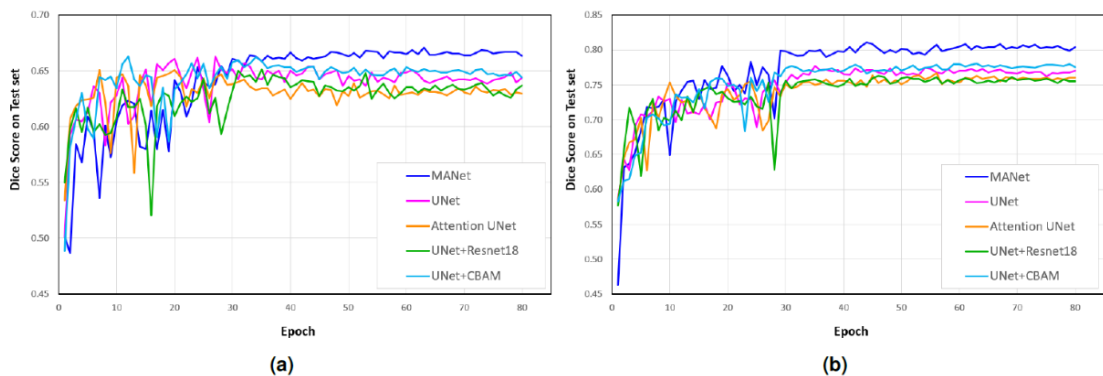
Figure 24: Test performance of the proposed MANet model and the baseline models on the test set for 80 epochs of the training process. (a) The segmentation performance of the volume-based experiment. (b) The segmentation performance of the slice-based experiment.

The proposed MANet could improve the average dice score by more than 3% in slice-based segmentation and 1% in volume-based segmentation while maintaining a similar performance gap in the Jaccard index. In volume-based segmentation, UNet+CBAM performed approximately identically to the proposed model, the MANet is lower by 1% in dice score and Jaccard index. However, MANet could boost performance in slice-based segmentation by approximately 3% in the dice score and Jaccard index. Furthermore, MANet got the lowest volume overlap error (VOE) in the two experiments by obtaining the highest overlapping rate, which can also be considered as the Jaccard index error metric. In terms of ASSD, the proposed MANet outperformed baseline models. It was further verified by possessing the greatest overlapping rate with superiority in liver tumor segmentation. Attention UNet has failed to demonstrate a significant performance improvement over the UNet model. However, it could achieve the same level of segmentation performance as UNet+Resunet18 while utilizing substantially fewer amount of parameters. In particular, the Dice score, Jaccard index, VOE, and Sensitivity plots of both volume-based segmentation and slice-based segmentation demonstrate a better correlation in all comparison models calculated based on foreground pixels (see (a), (b), (c), (e) plots of Figure 25 and Figure 26. Furthermore, the performance

comparison of volume-based and slice-based segmentation is plotted in Figure 27 to get a better insight into the proposed network in terms of effectiveness and generalizability.

Task	Method	Dice score	ASSD	Jaccard index (IoU)	VOE	Accuracy	Sensitivity (Recall)	Specificity
Volume-based Segmentation	UNet	0.6612 ± 0.277	1.0843 ± 1.425	0.5469 ± 0.266	0.4530 ± 0.266	**0.9950 ± 0.004**	0.6394 ± 0.285	0.9987 ± 0.002
	Attention UNet	0.6505 ± 0.278	1.2551 ± 1.338	0.5356 ± 0.263	0.4643 ± 0.263	0.9945 ± 0.006	0.6250 ± 0.292	0.9984 ± 0.002
	UNet + Resnet18	0.6560 ± 0.281	**0.9321 ± 0.960**	0.5433 ± 0.268	0.4566 ± 0.268	**0.9950** ± 0.005	0.6108 ± 0.294	**0.9991 ± 0.001**
	UNet + CBAM	0.6635 ± 0.271	1.2795 ± 1.638	0.5487 ± 0.261	0.4512 ± 0.261	0.9946 ± 0.005	0.6678 ± 0.283	0.9981 ± 0.002
	MANet (Proposed model)	**0.6735 ± 0.267**	1.2049 ± 1.356	**0.5590 ± 0.258**	**0.4409 ± 0.258**	**0.9950 ± 0.004**	**0.7426 ± 0.283**	0.9978 ± 0.002
Slice-based segmentation	UNet	0.7790 ± 0.208	0.9009 ± 1.020	0.6744 ± 0.217	0.3255 ± 0.217	0.9940 ± 0.006	0.7476 ± 0.237	0.9982 ± **0.001**
	Attention UNet	0.7676 ± 0.195	0.9188 ± 0.783	0.6550 ± 0.208	0.3449 ± 0.208	0.9935 ± 0.006	0.7423 ± 0.231	0.9978 ± 0.002
	UNet + Resnet18	0.7686 ± 0.211	1.0037 ± 1.4291	0.6619 ± 0.223	0.3380 ± 0.223	0.9934 ± 0.007	0.7342 ± 0.245	**0.9984 ± 0.001**
	UNet + CBAM	0.7784 ± 0.202	0.8241 ± 0.810	0.6720 ± 0.214	0.3279 ± 0.214	0.9941 ± 0.005	0.7439 ± 0.234	0.9982 ± 0.002
	MANet (Proposed model)	**0.8145 ± 0.150**	**0.7084 ± 0.701**	**0.7084 ± 0.171**	**0.2915 ± 0.171**	**0.9947 ± 0.004**	**0.8723 ± 0.173**	0.9970 ± 0.002

Table 2: The quantitative liver tumor segmentation performance comparison of the proposed MANet model and the baseline models for volume-based and slice-based segmentation experiments (mean ± standard deviation) on the LiTS dataset. The best values are in bold.

Figure 25: The quantitative liver tumor segmentation performance analysis of the proposed MANet model and the baseline models for volume-based segmentation on the LiTS dataset for the measures:(a) Dice score, (b) Jaccard index (IoU), (c) Sensitivity (Recall), (d) ASSD, (e) VOE, (f) Specificity, (g) Accuracy.
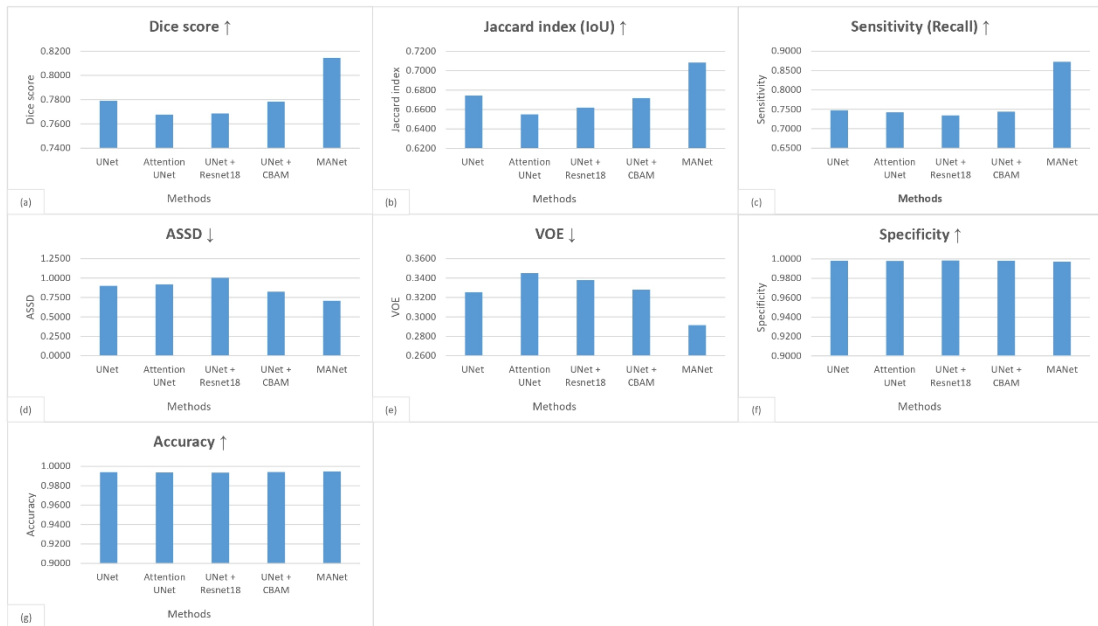
Figure 26: The quantitative liver tumor segmentation performance analysis of the proposed MANet model and the baseline models for slice-based segmentation on the LiTS dataset for the measures:(a) Dice score, (b) Jaccard index (IoU), (c) Sensitivity (Recall), (d) ASSD, (e) VOE, (f) Specificity, (g) Accuracy.

The proposed model failed to produce the best specificity performance, which two perspectives could potentially explain. According to the verification of an experienced radiologist in our research team, we discovered that several ground truth mask regions were smaller than actual tumor regions, and the proposed approach could segment and recognize tumor boundaries more effectively with greater accuracy. The liver tumor segmentation results verification with an experienced radiologist is comprehensively discussed in the following section. In terms of evaluation metrics, those specific cases are classified as false positives (over-segmentation) that cause decreased specificity. Apart from that, the limited parameters of the model could lead to over-segmentation (approximately half of the parameters compared to the UNet). It has been shown by UNet+Resunet18 by obtaining the best specificity in both slice-based segmentation and volume-based segmentation tests with the greatest number of parameters among all experimental models.

It is worth highlighting that the proposed approach demonstrated the highest sensitivity among comparison models, it has outperformed all baseline models in terms of sensitivity by approximately 8% in volume-based segmentation and 13% in slice-based segmentation. When comparing the performance of the proposed model to UNet+CBAM, it is clear that using channel attention in encoder blocks and spatial attention in decoder blocks is far more efficient than using CBAM in both encoder and decoder network paths in each block. Furthermore, the proposed model outperformed Attention UNet by using attention mechanisms to extract features in every stage of the network (i.e., encoder, decoder, skip connection), whereas Attention UNet solely extracts features in the skip connection through attention gates.

Figure 27: The quantitative liver tumor segmentation performance analysis of the proposed MANet model and the baseline models for volume-based and slice-based segmentation on the LiTS dataset for the measures:(a) Dice score, (b) Jaccard index (IoU), (c) Sensitivity (Recall), (d) ASSD, (e) VOE, (f) Specificity, (g) Accuracy.

5.3.1.2. Quantitative Analysis based on the Performance Rating

The quantitative analysis is further continued based on the segmentation performance rating with the liver tumor segmentation results of volume-based and slice-based segmentation experiments on the LiTS dataset. The segmentation rating is categorized based on the liver tumor segmentation performance as shown in Table 3. Under the performance classification, segmentation rating 5 is regarded as excellent performance, and segmentation rating 1 gives poor performance.

Segmentation Rating	Performance Range (%)
1	0 % - 20 %
2	20 % - 40 %
3	40 % - 60 %
4	60 % - 80 %
5	80 % - 100 %

Table 3: The quantitative segmentation rating. The segmentation rating is based on the liver tumor segmentation performance.

The analysis is conducted based on the Dice score, Jaccard index (IoU), and Sensitivity (Recall) that are used to evaluate the segmentation performance with the foreground pixels. Those are widely used evaluation metrics to analyze the performance of liver tumor segmentation. The analysis has taken into account the liver tumor segmentation results from both volume-based and slice-based experiments as visualized in Figure 28 and Figure 29 respectively.
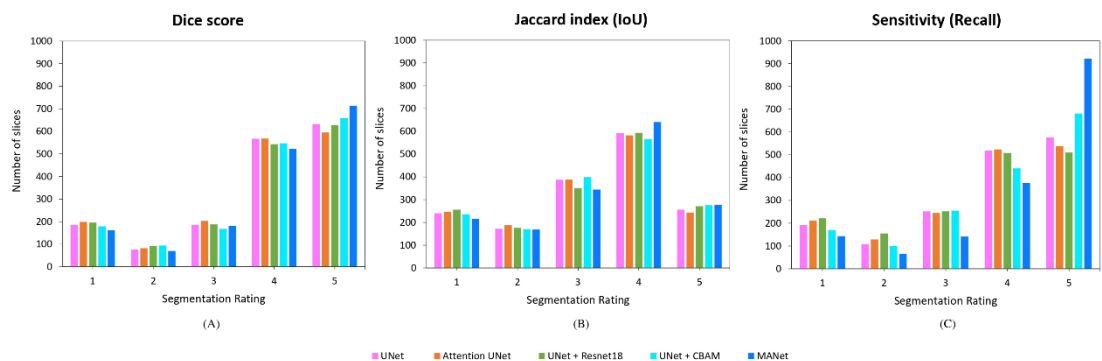
Figure 28: Histogram of the quantitative segmentation rating for volume-based liver tumor segmentation experiment on LiTS dataset. (A) Dice score, (B) Jaccard index (IoU), (C) Sensitivity (Recall)
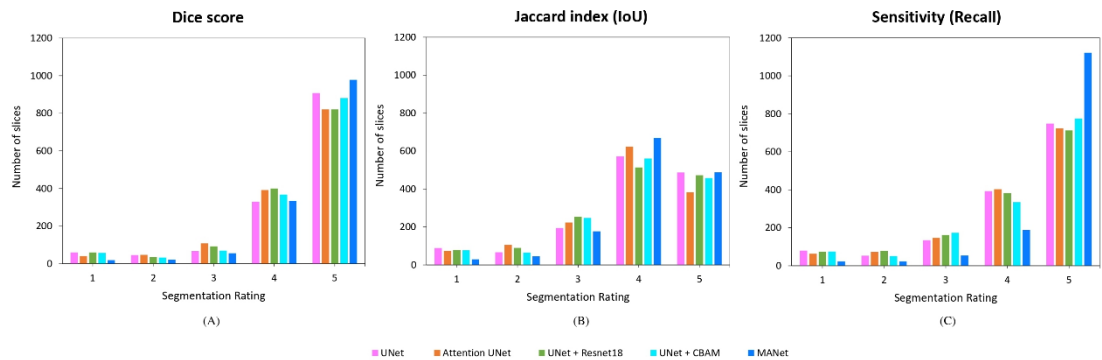
Figure 29: Histogram of the quantitative segmentation rating for slice-based segmentation experiment on the LiTS dataset. (A) Dice score, (B) Jaccard index (IoU), (C) Sensitivity (Recall)

The visualization clearly demonstrated that the proposed MANet has maximized the slice count under the segmentation rating 5 which is regarded as the excellent performance range (80% - 100%). Furthermore, the proposed model has minimized the slice count in both segmentation ratings 1 and 2 which presents the performance range between 0% and 40% in all the evaluation metrics. It is worth highlighting that the sensitivity of the proposed model has maintained a significant performance gap compared to other comparison models in both volume-based and slice-based experiments. Furthermore, the slice count of segmentation rating from 1 to 4 is minimized in the sensitivity performance histogram in both volume-based and slice-based experiments.

In common, the proposed model has demonstrated superiority in sensitivity compared to comparison models in each quantitative analysis. It is noticeable that the slice-based segmentation has demonstrated a significant performance gap compared to the volume-based segmentation. Because the model is trained with a wide range of scans (without increasing slice count) compared to the volume-based experiment, it proves that the performance of liver tumor segmentation can be improved by increasing the number of training samples with a wide range of liver tumor cases.

5.3.1.3. Quantitative Analysis with the 3DIRCADb Dataset

The MANet architecture is further assessed using the 3DIRCADb dataset, which is capable of showing the network generalization. Table 4 shows the quantitative analysis results for the liver tumor segmentation on the 3DIRCADb dataset.

Method	Dice score	ASSD	Jaccard index (IoU)	VOE	Accuracy	Sensitivity (Recall)	Specificity
UNet	0.5767 ± 0.282	1.2578 ± 1.199	0.4534 ± 0.246	0.5466 ± 0.246	0.9942 ± 0.006	0.4813 ± **0.253**	0.9996 ± 0.001
Attention UNet	0.5863 ± 0.281	1.4189 ± 1.288	0.4629 ± 0.245	0.5371 ± 0.245	0.9943 ± 0.006	0.4954 ± 0.259	0.9995 ± 0.001
UNet + Resnet18	0.5941 ± **0.270**	**1.2051 ± 1.038**	0.4681 ± **0.241**	0.5319 ± **0.241**	0.9944 ± 0.006	0.4956 ± 0.256	**0.9997 ± 0.001**
UNet + CBAM	0.5763 ± 0.278	1.5157 ± 1.458	0.4521 ± 0.246	0.5479 ± 0.246	0.9941 ± 0.006	0.4909 ± 0.257	0.9995 ± 0.001
MANet (Proposed model)	**0.6400** ± 0.279	1.3492 ± 1.362	**0.5227** ± 0.258	**0.4773** ± 0.258	**0.9947 ± 0.006**	**0.6240** ± 0.298	0.9990 ± 0.002

Table 4: The quantitative liver tumor segmentation performance comparison of the proposed MANet model and the baseline models for volume-based segmentation experiment (mean ± standard deviation) on the 3DIRCADb dataset. The best values are in bold.

In terms of dice score, the proposed network enhanced performance by almost 5%. It can be proved by achieving the highest possible overlapping rate by minimizing volume overlap error (VOE). It is noteworthy that the network's sensitivity has maintained a significant performance margin (about 13%) even in the 3DIRCADb dataset. The volume-based experiment results with the LiTS and 3DIRCADb datasets showed a strong correlation in all evaluation metrics. Nevertheless, in terms of dice score and sensitivity in the 3DIRCADb dataset, the proposed MANet could outperform the comparison networks substantially. Moreover, it is easily noticeable that the proposed architecture has maintained a minimum performance gap in most of the measures as illustrated in Figure 30. In the performance analysis, the proposed network outperformed in both the LiTS and 3DIRCADb datasets, demonstrating the network's higher generalizability.
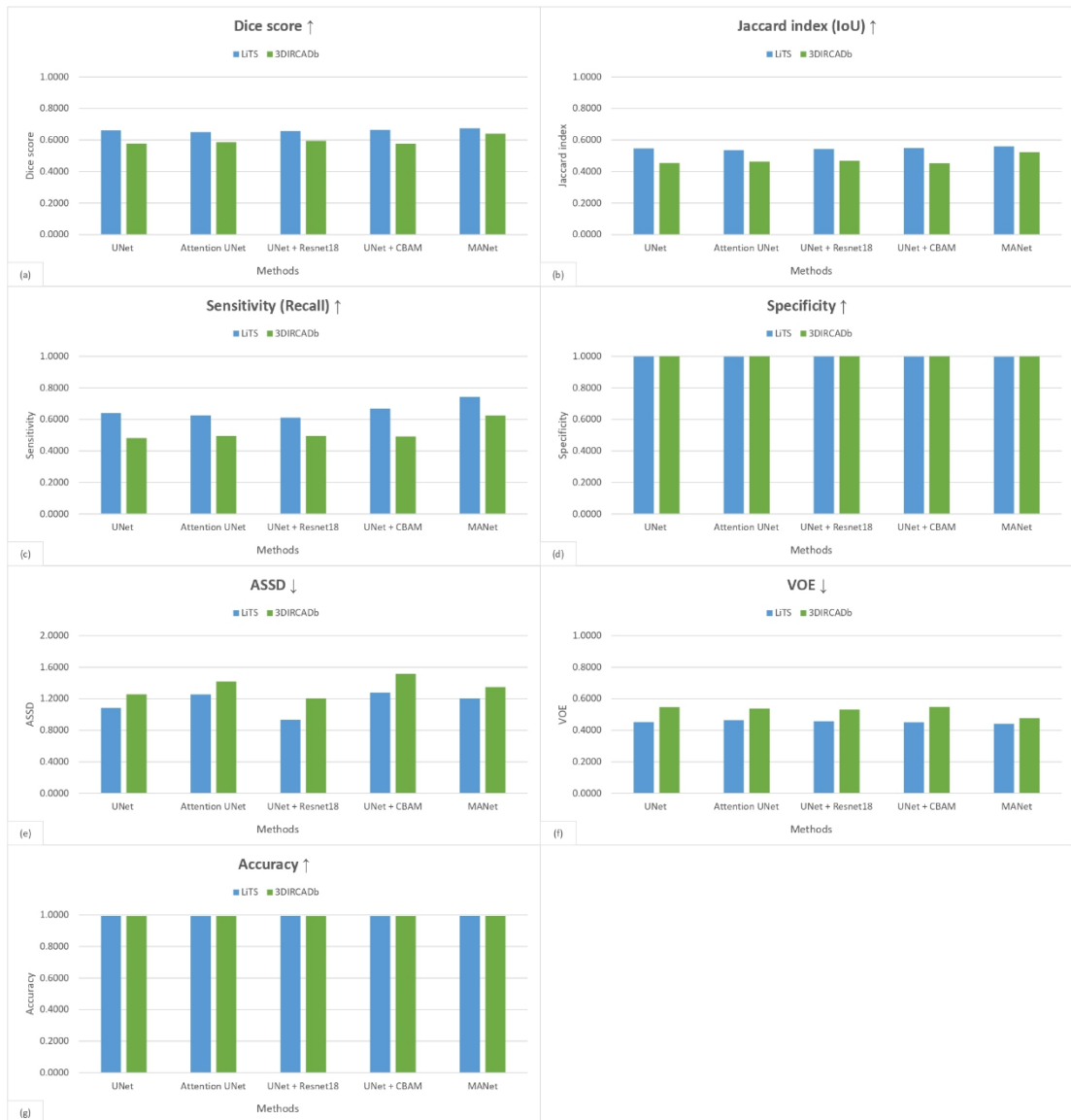
Figure 30: The quantitative liver tumor segmentation performance analysis of the proposed MANet model and the baseline models for volume-based segmentation on the LiTS and 3DIRCADb datasets for the measures:(a) Dice score, (b) Jaccard index (IoU), (c) Sensitivity (Recall), (d) ASSD, (e) VOE, (f) Specificity, (g) Accuracy

5.3.2. Quantitative Analysis of state-of-the-art Networks

The slice-based segmentation experiment is carried out to compare the proposed network's robustness and effectiveness with other state-of-the-art approaches. To evaluate performance with the proposed MANet architecture, the most recently presented UNet-based developments of UNet 3+, ResUNet++, SmaAt-UNet, and TA-Net were used. Table 5 and Figure 31 illustrate the quantitative analysis of the slice-based segmentation experiment of the state-of-the-art approaches. It is easily noticeable that the proposed MANet development outperformed other state-of-the-art methods in the majority of evaluation metrics.

It is worth noting that the architectures with attention mechanisms performed substantially better in the experiment. ResUNet++ has come closer to the proposed MANet model in terms of dice score, but it has not demonstrated notable performance similarity in sensitivity. While both ResUNet++ and SmaAt-UNet had higher dice scores with lower parameter overhead, the proposed MANet development showed its superiority by achieving the highest overlapping rate, which can be reflected by the VOE, ASSD, and Jaccard index.

Method	Dice score	ASSD	Jaccard index (IoU)	VOE	Accuracy	Sensitivity (Recall)	Specificity	Total training parameters (M)
UNet 3+	0.5036 ± 0.341	1.3994 ± 1.857	0.4054 ± 0.306	0.5946 ± 0.306	0.9893 ± 0.010	0.4696 ± 0.364	0.9977 ± 0.005	26.98
ResUNet++	0.8101 ± 0.175	1.0323 ± 0.950	0.6727 ± 0.191	0.3273 ± 0.191	0.9937 ± 0.006	0.8330 ± 0.205	0.9968 ± 0.003	4.06
SmaAt-UNet	0.7880 ± 0.185	0.8300 ± 0.955	0.6802 ± 0.202	0.3198 ± 0.202	0.9938 ± 0.007	0.7433 ± 0.218	0.9986 ± 0.002	4.03
TA-Net	0.7904 ± 0.172	0.9331 ± 0.974	0.6799 ± 0.190	0.3202 ± 0.190	0.9937 ± 0.007	0.7751 ± 0.209	0.9979 ± 0.003	29.57
MANet (Proposed model)	**0.8145 ± 0.150**	**0.7084 ± 0.701**	**0.7084 ± 0.171**	**0.2915 ± 0.171**	**0.9947 ± 0.004**	**0.8723 ± 0.173**	**0.9970 ± 0.002**	**7.83**

Table 5: The quantitative liver tumor segmentation performance comparison of the proposed MANet model and other state-of-the-art methods for slice-based segmentation experiment (mean ± standard deviation) on the LiTS dataset. The proposed model results are in bold.
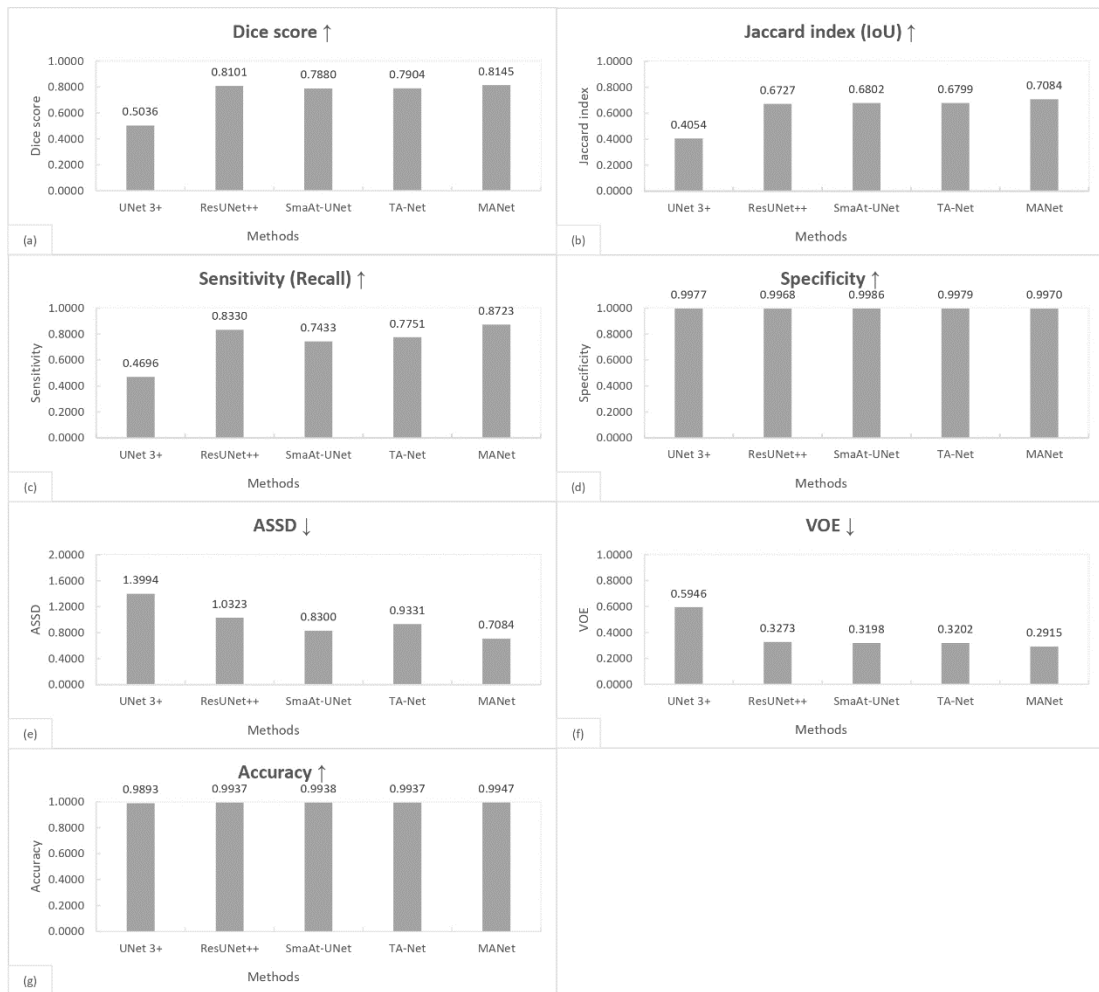
Figure 31: The quantitative liver tumor segmentation performance analysis of the proposed MANet model and other state-of-the-art methods for slice-based segmentation on the LiTS dataset for the measures:(a) Dice score, (b) Jaccard index (IoU), (c) Sensitivity (Recall), (d) ASSD, (e) VOE, (f) Specificity, (g) Accuracy

5.4. Quantitative Analysis based on the Number of Tumors

The quantitative analysis is further continued based on the number of tumors present in the CT slice. It is crucial to assess the performance of the model based on the number of tumors since focusing on multiple tumor cases can be challenging compared to single tumor cases even for radiologists. To deliver better insight into the model, segmentation performance is compared among comparison models based on the number of tumors present in the CT slice.

The total area of the tumor is calculated based on in-plane image resolution ranges given in the liver tumor benchmark (LiTS) publication (Bilic et al., 2023). The in-plane image resolution varies in the range from 0.56 mm to 1.0 mm. The average in-plane image resolution is 0.78 mm which is used as pixel spacing to calculate the total area of the tumors based on the number of foreground pixels. The OpenCV contour method is used to calculate the number of blobs in the tumor mask. The number of tumors (blobs) varies between 1 and 24. The analysis is conducted for the number of tumors from 1 to 15 and more than 15 (16 categories). According to the slice distribution based on the number of tumors, most of the tumor CT slices contained only one tumor while comparatively less amount of CT slices contained multiple tumors. The distribution of the liver tumor CT slices is shown in Table 6. The quantitative analysis based on the number of tumors is conducted with the proposed MANet and the comparison models as shown in Table 7 to Table 11. The analysis is visualized based on each evaluation metric to compare the performance of the proposed network with comparison networks (Dice score: Figure 32, Jaccard index (IoU): Figure 33, Sensitivity (Recall): Figure 34, ASSD: Figure 35, VOE: Figure 36, Specificity: Figure 37, Accuracy: Figure 38).

Number of Tumors	Number of Slices	Average Total Area of Tumors (mm^2)
1	432	1689
2	285	2842
3	150	3707
4	132	3131
5	96	3983
6	60	3635
7	46	2637
8	33	3495
9	25	3495
10	17	4179
11	14	2808
12	16	3219
13	13	4397
14	12	4537
15	26	5517
> 15	53	6262

Table 6: Liver tumor CT slice data distribution for categories based on the number of liver tumors in the CT slice.

Number of Tumors	Dice score	ASSD	Jaccard index (IoU)	VOE	Accuracy	Sensitivity (Recall)	Specificity
1	0.8011	0.5949	0.6997	0.3003	0.9970	0.8897	0.9985
2	0.8085	0.7777	0.7088	0.2912	0.9956	0.8516	0.9975
3	0.8308	0.7233	0.7341	0.2659	0.9954	0.8947	0.9969
4	0.8199	0.7394	0.7072	0.2928	0.9944	0.8931	0.9961
5	0.8378	0.8642	0.7293	0.2707	0.9933	0.9134	0.9954
6	0.8392	0.8850	0.7288	0.2712	0.9936	0.8914	0.9959
7	0.8404	0.7412	0.7285	0.2715	0.9950	0.8736	0.9973
8	0.8125	0.9539	0.6904	0.3096	0.9920	0.8372	0.9957
9	0.8153	1.0955	0.6969	0.3031	0.9922	0.8281	0.9965
10	0.8105	1.1173	0.6846	0.3154	0.9896	0.8441	0.9940
11	0.7765	0.8708	0.6410	0.3591	0.9924	0.8436	0.9957
12	0.7865	0.9254	0.6528	0.3472	0.9914	0.8700	0.9948
13	0.8042	1.0523	0.6753	0.3247	0.9889	0.8153	0.9943
14	0.8267	1.0841	0.7079	0.2921	0.9898	0.8418	0.9946
15	0.8206	1.0343	0.6982	0.3019	0.9883	0.7937	0.9955
> 15	0.8231	1.1431	0.7018	0.2982	0.9869	0.7909	0.9951

Table 7: Quantitative analysis based on the number of tumors for liver tumor segmentation performance of the MANet. The analysis is conducted for the slice-based segmentation experiment on the LiTS dataset.

Number of Tumors	Dice score	ASSD	Jaccard index (IoU)	VOE	Accuracy	Sensitivity (Recall)	Specificity
1	0.7722	0.6853	0.6861	0.3139	0.9962	0.7499	0.9994
2	0.7776	0.7979	0.6746	0.3254	0.9951	0.7302	0.9988
3	0.7741	1.0236	0.6768	0.3232	0.9950	0.7493	0.9980
4	0.7604	0.9761	0.6410	0.3590	0.9930	0.7385	0.9976
5	0.7768	1.1425	0.6622	0.3378	0.9915	0.7555	0.9970
6	0.8204	0.9193	0.7075	0.2925	0.9934	0.7982	0.9973
7	0.8128	0.8973	0.6937	0.3063	0.9947	0.7758	0.9983
8	0.8052	1.0232	0.6819	0.3181	0.9926	0.7825	0.9972
9	0.8107	0.9678	0.6933	0.3067	0.9928	0.7609	0.9983
10	0.7745	1.3685	0.6411	0.3589	0.9889	0.7145	0.9979
11	0.7705	1.1647	0.6335	0.3665	0.9925	0.7411	0.9976
12	0.7634	1.2743	0.6209	0.3791	0.9904	0.7782	0.9958
13	0.7781	1.4391	0.6397	0.3603	0.9885	0.7324	0.9963
14	0.7986	1.1357	0.6745	0.3255	0.9895	0.7431	0.9970
15	0.7889	1.3525	0.6573	0.3427	0.9869	0.7209	0.9966
> 15	0.7984	1.3773	0.6659	0.3341	0.9854	0.7367	0.9956

Table 8: Quantitative analysis based on the number of tumors for liver tumor segmentation performance of the UNet. The analysis is conducted for the slice-based segmentation experiment on the LiTS dataset.

Number of Tumors	Dice score	ASSD	Jaccard index (IoU)	VOE	Accuracy	Sensitivity (Recall)	Specificity
1	0.7450	0.6932	0.6473	0.3527	0.9957	0.7204	0.9994
2	0.7695	0.7760	0.6622	0.3378	0.9949	0.7340	0.9985
3	0.7913	0.9715	0.6886	0.3114	0.9947	0.7817	0.9974
4	0.7828	0.9964	0.6621	0.3379	0.9933	0.7850	0.9968
5	0.7930	1.1575	0.6702	0.3298	0.9916	0.7856	0.9959
6	0.8003	1.0385	0.6773	0.3227	0.9919	0.7921	0.9960
7	0.7808	1.0736	0.6496	0.3504	0.9935	0.7571	0.9975
8	0.7504	1.3164	0.6116	0.3884	0.9904	0.7202	0.9960
9	0.7490	1.3784	0.6135	0.3865	0.9904	0.6997	0.9968
10	0.7626	1.3453	0.6201	0.3799	0.9895	0.6989	0.9976
11	0.7516	0.9951	0.6063	0.3937	0.9915	0.7306	0.9972
12	0.7529	1.1617	0.6063	0.3937	0.9908	0.7404	0.9965
13	0.7519	1.4564	0.6065	0.3935	0.9879	0.6829	0.9965
14	0.7747	1.2310	0.6352	0.3648	0.9889	0.6892	0.9976
15	0.7754	1.2046	0.6356	0.3644	0.9861	0.7000	0.9964
> 15	0.7709	1.4435	0.6296	0.3704	0.9836	0.6996	0.9954

Table 9: Quantitative analysis based on the number of tumors for liver tumor segmentation performance of the Attention UNet. The analysis is conducted for the slice-based segmentation experiment on the LiTS dataset.

Number of Tumors	Dice score	ASSD	Jaccard index (IoU)	VOE	Accuracy	Sensitivity (Recall)	Specificity
1	0.7517	0.9523	0.6683	0.3317	0.9958	0.7289	0.9995
2	0.7890	0.6995	0.6869	0.3131	0.9953	0.7484	0.9989
3	0.8038	0.8909	0.7041	0.2959	0.9949	0.7762	0.9983
4	0.7403	1.0909	0.6168	0.3833	0.9916	0.7242	0.9975
5	0.7634	1.3811	0.6421	0.3579	0.9902	0.7375	0.9974
6	0.8025	0.9987	0.6812	0.3188	0.9924	0.7743	0.9974
7	0.7951	0.9999	0.6688	0.3312	0.9941	0.7645	0.9979
8	0.7939	1.0078	0.6681	0.3319	0.9923	0.7684	0.9972
9	0.7942	1.1530	0.6681	0.3319	0.9918	0.7461	0.9978
10	0.7561	1.4690	0.6182	0.3818	0.9880	0.6960	0.9971
11	0.7415	1.0301	0.5982	0.4018	0.9914	0.6819	0.9977
12	0.7114	1.4865	0.5561	0.4439	0.9887	0.6852	0.9962
13	0.7232	1.5897	0.5726	0.4274	0.9861	0.6580	0.9960
14	0.7654	1.2658	0.6271	0.3729	0.9884	0.6741	0.9978
15	0.7382	1.5766	0.5919	0.4082	0.9851	0.6214	0.9983
> 15	0.7452	1.6290	0.5967	0.4033	0.9830	0.6335	0.9975

Table 10: Quantitative analysis based on the number of tumors for liver tumor segmentation performance of the UNet+Resnet18. The analysis is conducted for the slice-based segmentation experiment on the LiTS dataset.

Number of Tumors	Dice score	ASSD	Jaccard index (IoU)	VOE	Accuracy	Sensitivity (Recall)	Specificity
1	0.7582	0.6306	0.6675	0.3325	0.9963	0.7329	0.9994
2	0.7758	0.7486	0.6726	0.3274	0.9951	0.7262	0.9988
3	0.8009	0.8116	0.7035	0.2965	0.9954	0.7731	0.9979
4	0.7953	0.7611	0.6838	0.3162	0.9945	0.7784	0.9975
5	0.8275	0.8952	0.7207	0.2793	0.9935	0.8242	0.9967
6	0.8116	0.9196	0.6971	0.3029	0.9930	0.7926	0.9966
7	0.7838	1.0562	0.6569	0.3431	0.9938	0.7442	0.9979
8	0.7423	1.3264	0.6042	0.3958	0.9898	0.7020	0.9957
9	0.7528	1.4092	0.6157	0.3843	0.9905	0.6958	0.9970
10	0.7570	1.2785	0.6156	0.3844	0.9885	0.7056	0.9961
11	0.7687	0.8803	0.6272	0.3728	0.9925	0.7224	0.9982
12	0.7693	1.0789	0.6286	0.3714	0.9912	0.7469	0.9969
13	0.7768	1.1617	0.6387	0.3613	0.9888	0.7180	0.9967
14	0.7840	1.0761	0.6476	0.3524	0.9890	0.7056	0.9975
15	0.7847	1.1990	0.6494	0.3506	0.9872	0.6959	0.9977
> 15	0.7638	1.4124	0.6217	0.3783	0.9844	0.6650	0.9974

Table 11: Quantitative analysis based on the number of tumors for liver tumor segmentation performance of the UNet+CBAM. The analysis is conducted for the slice-based segmentation experiment on the LiTS dataset.

Figure 32: Quantitative analysis based on the number of tumors for liver tumor segmentation performance of the proposed MANet model and the baseline models for slice-based segmentation on the LiTS dataset. The performance comparison is based on the Dice score.
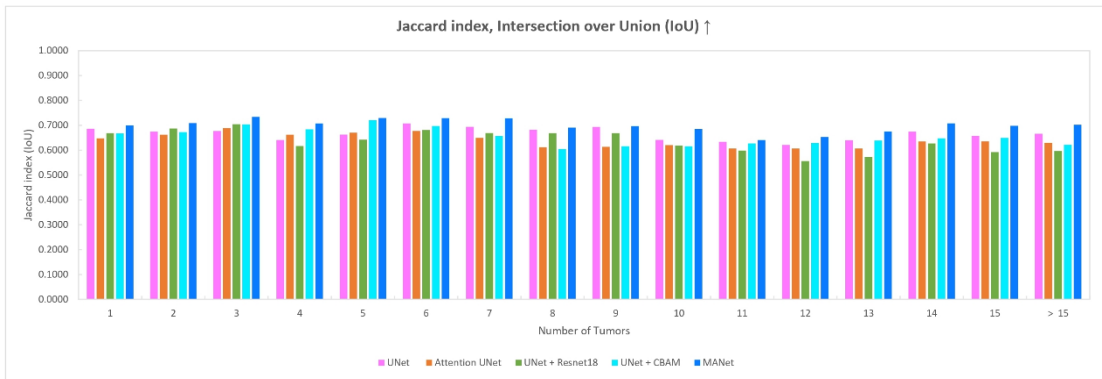
Figure 33: Quantitative analysis based on the number of tumors for liver tumor segmentation performance of the proposed MANet model and the baseline models for slice-based segmentation on the LiTS dataset. The performance comparison is based on the Jaccard index, Intersection over Union (IoU).
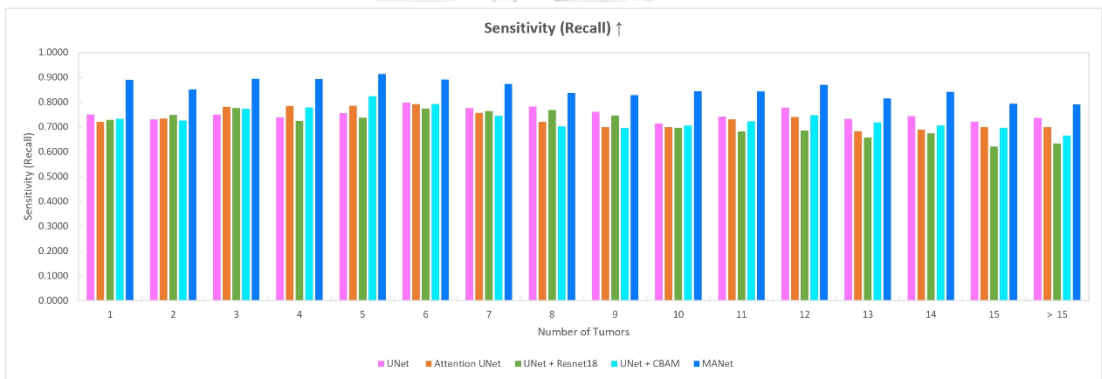
Figure 34: Quantitative analysis based on the number of tumors for liver tumor segmentation performance of the proposed MANet model and the baseline models for slice-based segmentation on the LiTS dataset. The performance comparison is based on the Sensitivity (Recall).
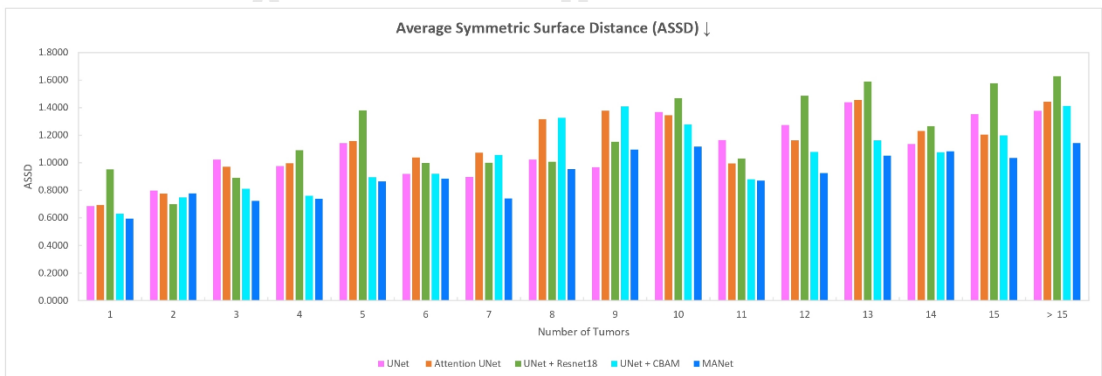
Figure 35: Quantitative analysis based on the number of tumors for liver tumor segmentation performance of the proposed MANet model and the baseline models for slice-based segmentation on the LiTS dataset. The performance comparison is based on the Average Symmetric Surface Distance (ASSD).
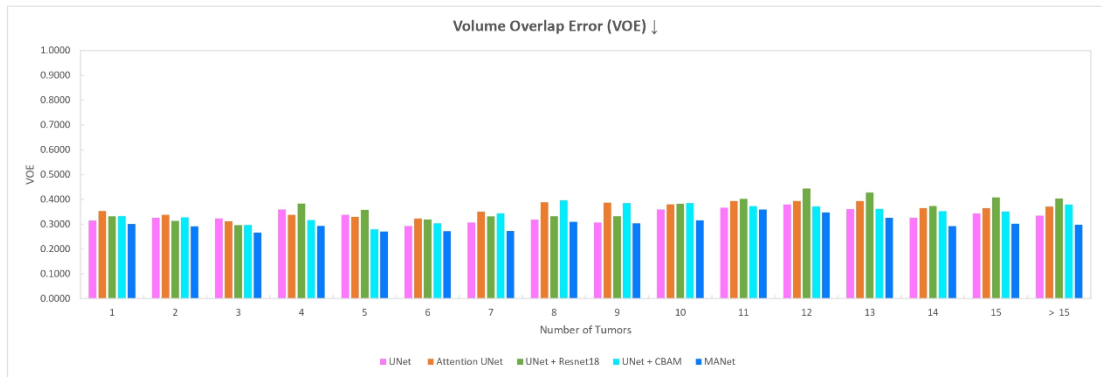
Figure 36: Quantitative analysis based on the number of tumors for liver tumor segmentation performance of the proposed MANet model and the baseline models for slice-based segmentation on the LiTS dataset. The performance comparison is based on the Volume Overlap Error (VOE).
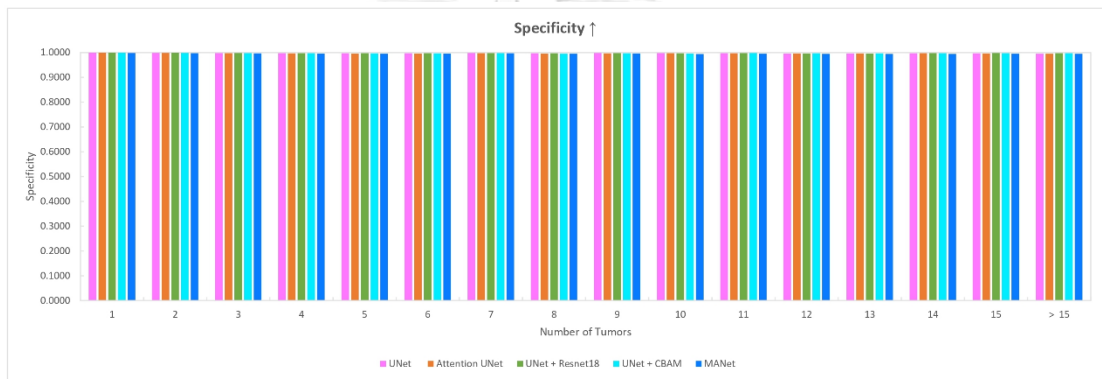
Figure 37: Quantitative analysis based on the number of tumors for liver tumor segmentation performance of the proposed MANet model and the baseline models for slice-based segmentation on the LiTS dataset. The performance comparison is based on the Specificity.
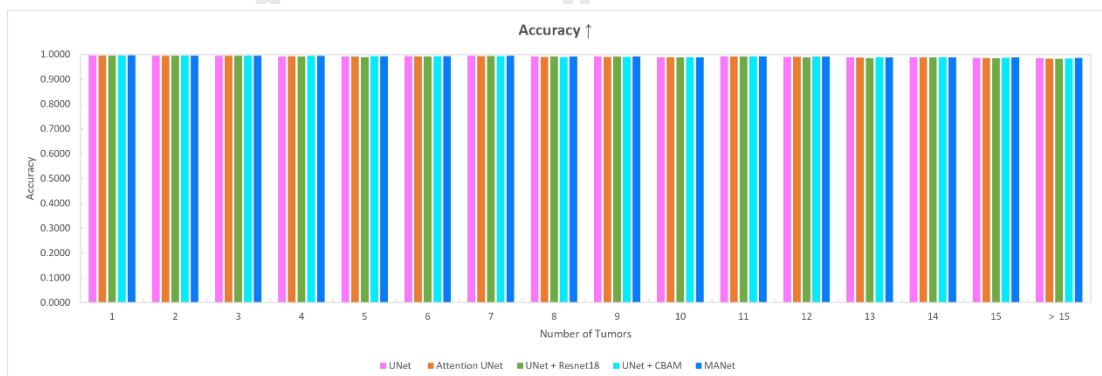
Figure 38: Quantitative analysis based on the number of tumors for liver tumor segmentation performance of the proposed MANet model and the baseline models for slice-based segmentation on the LiTS dataset. The performance comparison is based on the Accuracy.

The proposed network has demonstrated a significant performance boost in each category for most of the evaluation metrics. For instance, the proposed model could maintain a 3% average performance gap in the Dice score. In terms of the Jaccard index, the proposed model has maintained a 5% performance gap while achieving an 11% maximum performance gap compared to the UNet + Resnet18 model. The identical performance gap has been demonstrated by volume overlap error (VOE) in all the tumor categories which is the error metric of the Jaccard index. Average Symmetric Surface Distance (ASSD) has shown over 10% of average error minimization in the proposed network compared to comparison networks. The error rate has significantly risen when the number of tumors is more than eight. It is further discussed in the qualitative analysis section. In multiple tumor cases, tumors appear with fuzzy tumor boundaries, it is challenging to segment the tumor region with accurate edge precision. However, the proposed architecture could minimize the error rate in multiple tumor segmentations (see Figure 35).

It is worth highlighting that the proposed network depicts a significant performance gap in all the tumor categories (see Figure 34). Furthermore, the proposed model has maintained a 12% average performance gap in sensitivity. It has achieved an 18% maximum performance boost with UNet + Resnet18 while obtaining a 5% minimum performance boost with UNet in sensitivity. In general, all the models demonstrated almost similar performance in terms of accuracy and specificity. In summary, quantitative analysis based on the number of tumors further emphasizes the effectiveness and robustness of the proposed architecture.

5.5. Quantitative Analysis based on the Total Area of Tumors

The performance and the robustness of the model are further evaluated based on the total area of the tumors. The total area of the tumors is categorized into ten categories. To deliver better insight into the model, segmentation performance is compared among comparison models based on the total area of tumors visible in the CT slice.

The total area of the tumor is calculated based on in-plane image resolution ranges given in the liver tumor benchmark (LiTS) publication (Bilic et al., 2023). The in-plane image resolution varies in the range from 0.56 mm to 1.0 mm. The average in-plane image resolution is 0.78 mm which is used as pixel spacing to calculate the total area of the tumors based on the number of foreground pixels. The analysis is conducted for the total area of tumors from 0 to 10000 mm^2 (10 categories). According to the slice distribution based on the total area of tumors, most of the tumor CT slices are in the first category which has a total tumor area below 10000 mm^2. The distribution of the liver tumor CT slices is shown in Table 12. The quantitative analysis based on the total area of tumors is conducted with the proposed MANet and the comparison models as shown in Table 13 to Table 17. The analysis is visualized based on each evaluation metric to compare the performance of the proposed network with comparison networks (Dice score: Figure 39, Jaccard index (IoU): Figure 40, Sensitivity (Recall): Figure 41, ASSD: Figure 42, VOE: Figure 43, Specificity: Figure 44, Accuracy: Figure 45).

Total Tumor Area (mm^2)	Number of Slices
0 - 1000	410
1000 - 2000	212
2000 - 3000	154
3000 - 4000	174
4000 - 5000	182
5000 - 6000	100
6000 - 7000	58
7000 - 8000	74
8000 - 9000	42
9000 - 10000	4

Table 12: Liver tumor CT slice data distribution for categories based on the total area of liver tumors in the CT slice.

Total Tumor Area (mm²)	Dice score	ASSD	Jaccard index (IoU)	VOE	Accuracy	Sensitivity (Recall)	Specificity
0 - 1000	0.7302	0.5793	0.6074	0.3926	0.9983	0.8768	0.9986
1000 - 2000	0.8006	0.5885	0.6822	0.3178	0.9966	0.8819	0.9975
2000 - 3000	0.8562	0.6520	0.7547	0.2453	0.9954	0.8717	0.9974
3000 - 4000	0.8558	0.9316	0.7569	0.2431	0.9937	0.8748	0.9963
4000 - 5000	0.8511	0.9129	0.7612	0.2388	0.9925	0.8523	0.9966
5000 - 6000	0.8607	1.0253	0.7642	0.2358	0.9906	0.8646	0.9951
6000 - 7000	0.8832	0.9163	0.7954	0.2046	0.9905	0.8980	0.9945
7000 - 8000	0.8892	0.9904	0.8055	0.1945	0.9900	0.8822	0.9953
8000 - 9000	0.8848	1.1859	0.7963	0.2037	0.9882	0.8775	0.9943
9000 - 10000	0.8827	1.0510	0.7906	0.2094	0.9869	0.8643	0.9943

Table 13: Quantitative analysis based on the total area of tumors for liver tumor segmentation performance of the MANet. The analysis is conducted for the slice-based segmentation experiment on the LiTS dataset.

Total Tumor Area (mm²)	Dice score	ASSD	Jaccard index (IoU)	VOE	Accuracy	Sensitivity (Recall)	Specificity
0 - 1000	0.7147	0.5658	0.6113	0.3887	0.9985	0.6993	0.9995
1000 - 2000	0.7780	0.6266	0.6604	0.3396	0.9968	0.7463	0.9989
2000 - 3000	0.8082	0.8469	0.6955	0.3045	0.9947	0.7660	0.9983
3000 - 4000	0.8313	0.8483	0.7240	0.2760	0.9935	0.7898	0.9980
4000 - 5000	0.7734	1.4326	0.6902	0.3098	0.9907	0.7480	0.9978
5000 - 6000	0.8250	1.2719	0.7215	0.2785	0.9898	0.7754	0.9974
6000 - 7000	0.8316	1.1898	0.7264	0.2736	0.9880	0.7847	0.9966
7000 - 8000	0.8508	1.2396	0.7557	0.2443	0.9876	0.8181	0.9959
8000 - 9000	0.8183	1.6167	0.7041	0.2959	0.9830	0.7643	0.9953
9000 - 10000	0.6247	2.7438	0.4786	0.5214	0.9689	0.4997	0.9976

Table 14: Quantitative analysis based on the total area of tumors for liver tumor segmentation performance of the UNet. The analysis is conducted for the slice-based segmentation experiment on the LiTS dataset.

Total Tumor Area (mm²)	Dice score	ASSD	Jaccard index (IoU)	VOE	Accuracy	Sensitivity (Recall)	Specificity
0 - 1000	0.7115	0.5988	0.6039	0.3961	0.9985	0.7154	0.9994
1000 - 2000	0.7670	0.7159	0.6471	0.3529	0.9966	0.7447	0.9987
2000 - 3000	0.7728	0.9334	0.6473	0.3527	0.9938	0.7219	0.9982
3000 - 4000	0.7838	1.0785	0.6615	0.3385	0.9919	0.7309	0.9977
4000 - 5000	0.7755	1.1785	0.6766	0.3234	0.9900	0.7379	0.9974
5000 - 6000	0.8289	1.1762	0.7183	0.2817	0.9895	0.7873	0.9966
6000 - 7000	0.8355	1.2585	0.7239	0.2761	0.9874	0.8144	0.9948
7000 - 8000	0.8464	1.2450	0.7425	0.2575	0.9866	0.8227	0.9946
8000 - 9000	0.8320	1.4503	0.7187	0.2813	0.9833	0.8086	0.9931
9000 - 10000	0.6386	2.1729	0.4820	0.5181	0.9692	0.5046	0.9976

Table 15: Quantitative analysis based on the total area of tumors for liver tumor segmentation performance of the Attention UNet. The analysis is conducted for the slice-based segmentation experiment on the LiTS dataset.

Total Tumor Area (mm^2)	Dice score	ASSD	Jaccard index (IoU)	VOE	Accuracy	Sensitivity (Recall)	Specificity
0 - 1000	0.7430	0.4745	0.6354	0.3646	0.9986	0.7313	0.9995
1000 - 2000	0.7875	0.6127	0.6739	0.3261	0.9969	0.7637	0.9989
2000 - 3000	0.8154	0.7762	0.6994	0.3006	0.9947	0.7770	0.9981
3000 - 4000	0.7543	1.0594	0.6370	0.3630	0.9915	0.6913	0.9982
4000 - 5000	0.7373	1.9772	0.6613	0.3387	0.9899	0.7156	0.9979
5000 - 6000	0.8218	1.2078	0.7148	0.2852	0.9895	0.7672	0.9975
6000 - 7000	0.8009	1.3892	0.6945	0.3055	0.9867	0.7470	0.9970
7000 - 8000	0.8059	1.6395	0.7030	0.2970	0.9854	0.7414	0.9974
8000 - 9000	0.7271	2.2329	0.6048	0.3952	0.9781	0.6341	0.9974
9000 - 10000	0.6144	2.6847	0.4581	0.5419	0.9684	0.4646	0.9992

Table 16: Quantitative analysis based on the total area of tumors for liver tumor segmentation performance of the UNet+Resnet18. The analysis is conducted for the slice-based segmentation experiment on the LiTS dataset.

Total Tumor Area (mm^2)	Dice score	ASSD	Jaccard index (IoU)	VOE	Accuracy	Sensitivity (Recall)	Specificity
0 - 1000	0.7068	0.5528	0.6008	0.3992	0.9986	0.6872	0.9996
1000 - 2000	0.7760	0.6601	0.6602	0.3398	0.9969	0.7366	0.9991
2000 - 3000	0.8089	0.7595	0.6920	0.3080	0.9947	0.7567	0.9985
3000 - 4000	0.7997	0.9567	0.6883	0.3117	0.9926	0.7507	0.9980
4000 - 5000	0.7957	1.0882	0.7046	0.2954	0.9910	0.7572	0.9979
5000 - 6000	0.8163	1.1502	0.7044	0.2956	0.9885	0.7781	0.9960
6000 - 7000	0.8536	1.0339	0.7542	0.2458	0.9890	0.8281	0.9958
7000 - 8000	0.8699	1.0755	0.7804	0.2196	0.9891	0.8353	0.9967
8000 - 9000	0.8721	1.1833	0.7808	0.2192	0.9874	0.8624	0.9943
9000 - 10000	0.6649	2.4213	0.5232	0.4768	0.9718	0.5486	0.9977

Table 17: Quantitative analysis based on the total area of tumors for liver tumor segmentation performance of the UNet+CBAM. The analysis is conducted for the slice-based segmentation experiment on the LiTS dataset.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Figure 39: Quantitative analysis based on the total area of tumors for liver tumor segmentation performance of the proposed MANet model and the baseline models for slice-based segmentation on the LiTS dataset. The performance comparison is based on the Dice score.

Figure 40: Quantitative analysis based on the total area of tumors for liver tumor segmentation performance of the proposed MANet model and the baseline models for slice-based segmentation on the LiTS dataset. The performance comparison is based on the Jaccard index, Intersection over Union (IoU).

Figure 41: Quantitative analysis based on the total area of tumors for liver tumor segmentation performance of the proposed MANet model and the baseline models for slice-based segmentation on the LiTS dataset. The performance comparison is based on the Sensitivity (Recall).
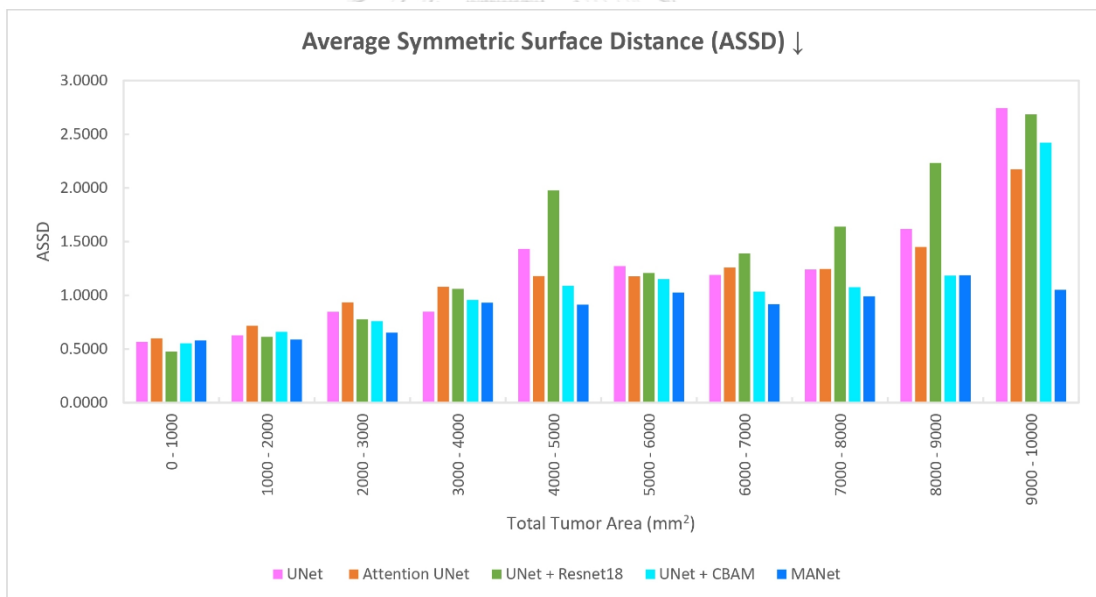
Figure 42: Quantitative analysis based on the total area of tumors for liver tumor segmentation performance of the proposed MANet model and the baseline models for slice-based segmentation on the LiTS dataset. The performance comparison is based on the Average Symmetric Surface Distance (ASSD).
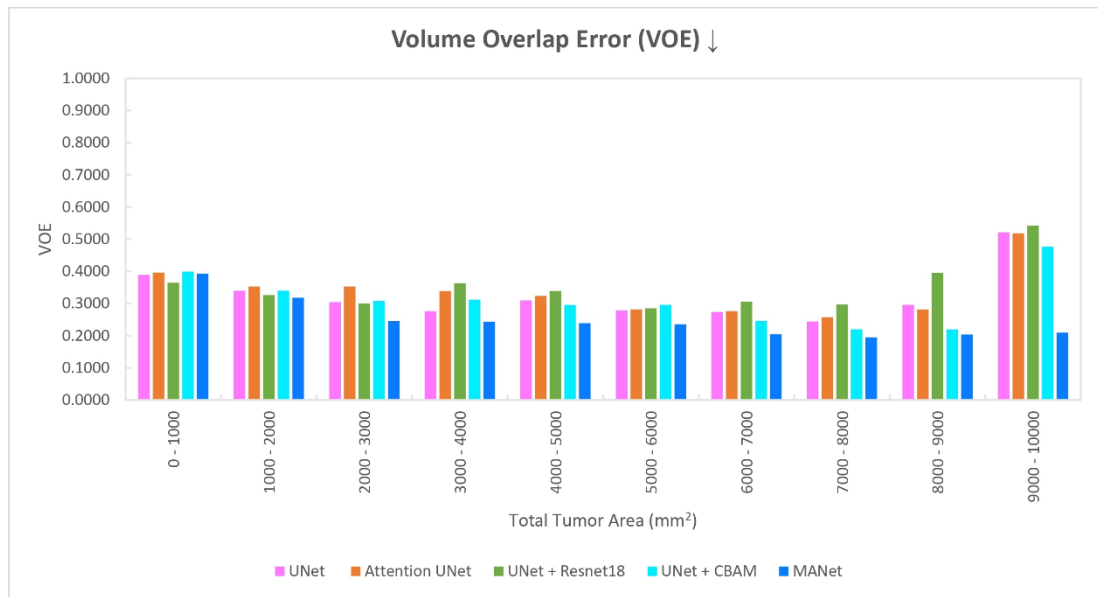
Figure 43: Quantitative analysis based on the total area of tumors for liver tumor segmentation performance of the proposed MANet model and the baseline models for slice-based segmentation on the LiTS dataset. The performance comparison is based on the Volume Overlap Error (VOE).
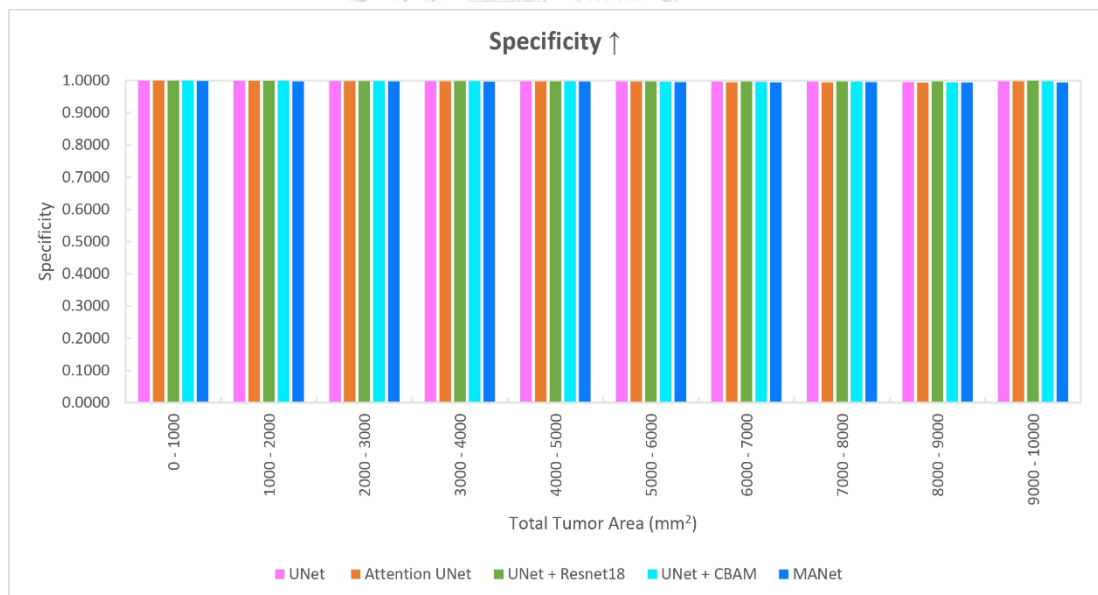
Figure 44: Quantitative analysis based on the total area of tumors for liver tumor segmentation performance of the proposed MANet model and the baseline models for slice-based segmentation on the LiTS dataset. The performance comparison is based on the Specificity.
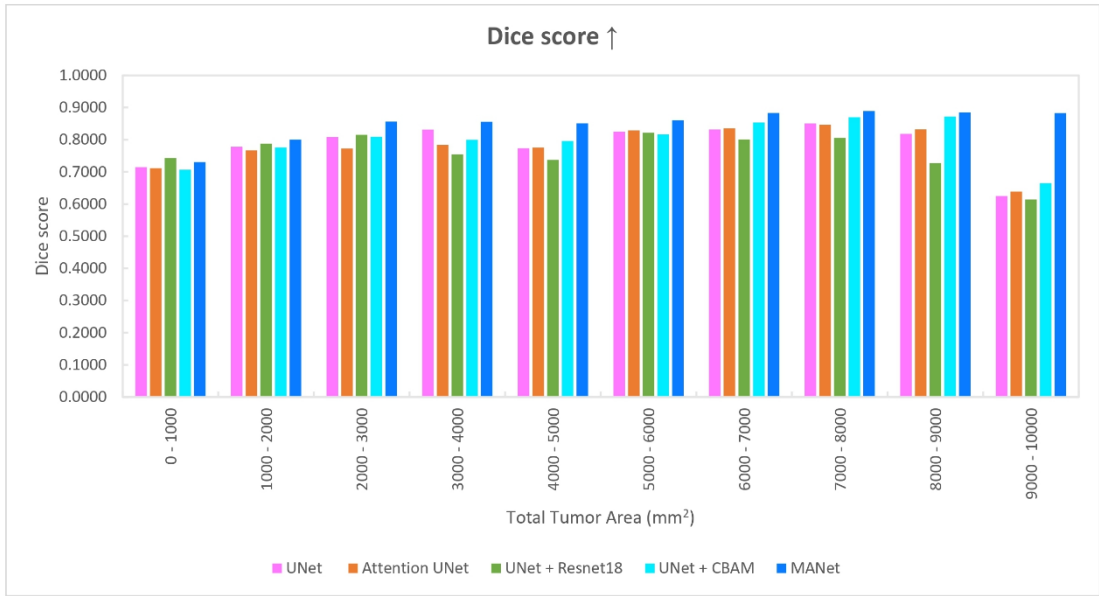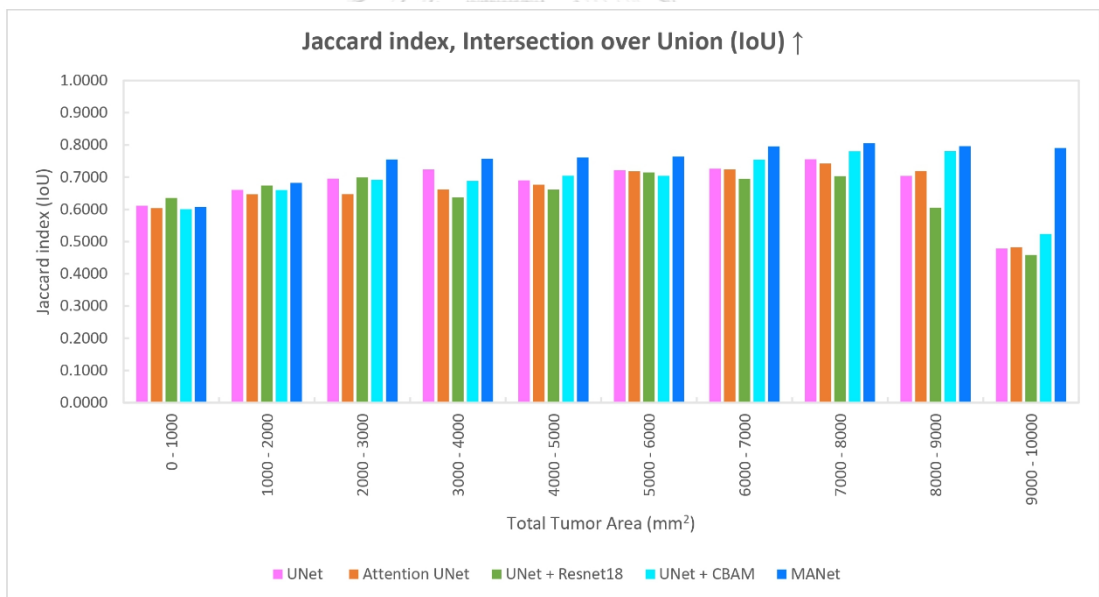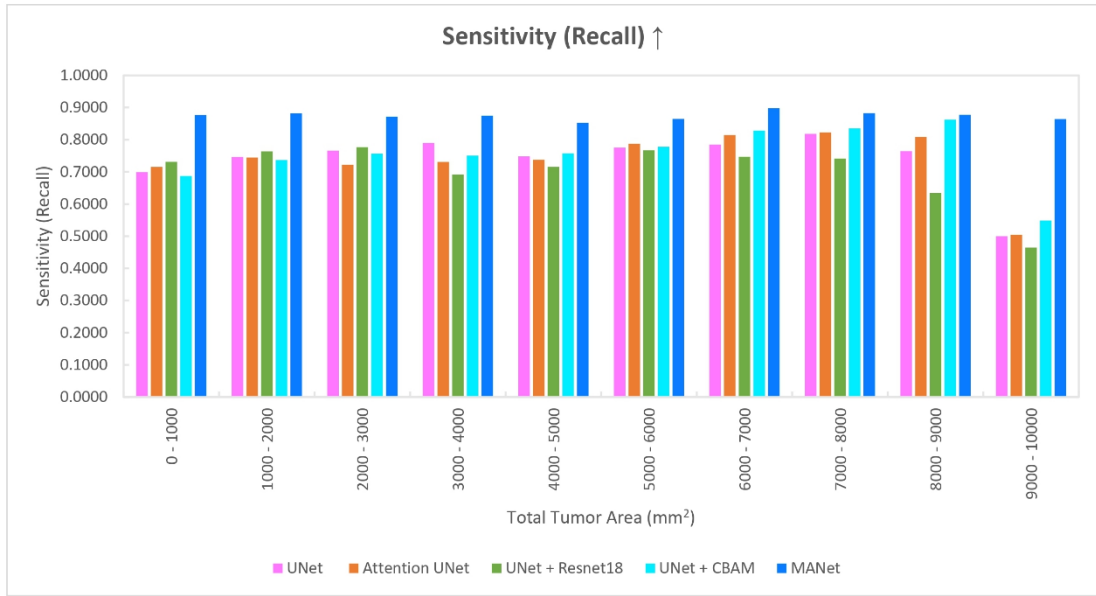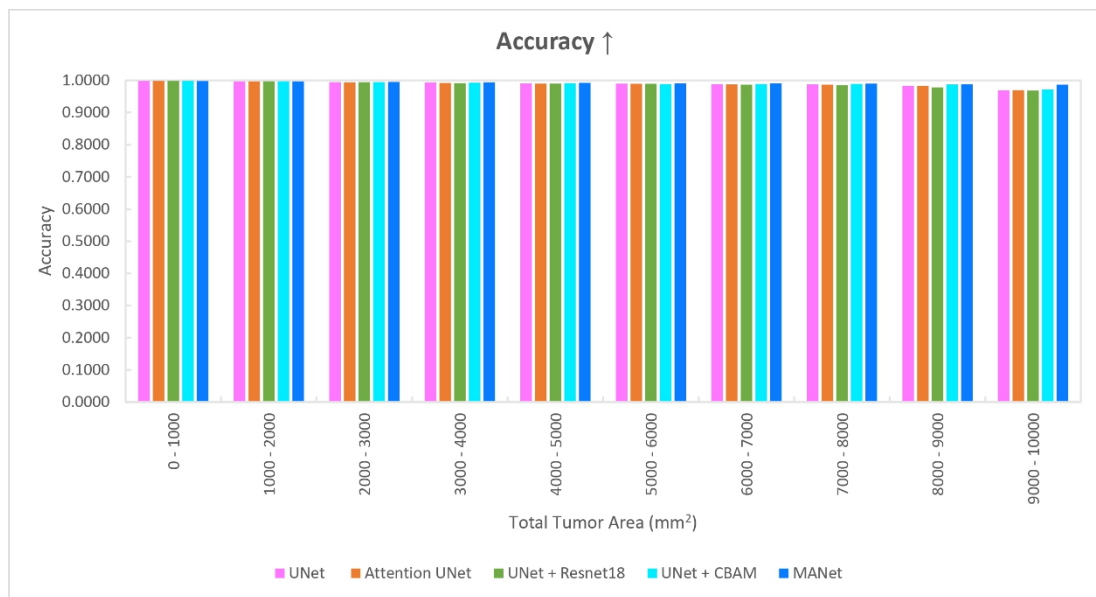
Figure 45: Quantitative analysis based on the total area of tumors for liver tumor segmentation performance of the proposed MANet model and the baseline models for slice-based segmentation on the LiTS dataset. The performance comparison is based on the Accuracy.

The proposed model has proven its superiority in most of the evaluation metrics. It is noticeable that the proposed model maintained an over 6% average performance gap in the Dice score. The performance gap is maximized in the last total tumor area category (almost 25% gap in the Dice score). The CT slices in this category can be multiple tumor cases since comparison models demonstrated poor segmentation performance with poor edge precision in multiple tumor cases according to the ASSD plot based on the number of tumors in the previous section (see Figure 35). We can see the same pattern in the last category in terms of Average Symmetric Surface Distance (ASSD). In terms of the Jaccard index, the proposed model has maintained an 8% average performance gap among comparison models. A 33% maximum performance gap was achieved with the UNet + Resnet18 model in the total tumor area between 9000 mm^2 and 10000 mm^2. The identical performance gap has been demonstrated by volume overlap error (VOE) in all the categories which is the error metric of the Jaccard index. Moreover, UNet + Resnet18 has significantly raised the error rate in most of the categories while the proposed model maintains a minimum error rate in most of the categories (see Figure 42).

It is worth highlighting that the proposed network depicts a significant performance gap in all the tumor categories (see Figure 41). Furthermore, the proposed model has maintained over 14% average performance gap in sensitivity. It has achieved a 40% maximum performance boost with UNet + Resnet18 (i.e., 9000 – 10000 mm^2 total tumor area range) while obtaining a 2% minimum performance boost with UNet + CBAM (i.e., 8000 – 9000 mm^2 total tumor area range) in sensitivity. In general, all the models demonstrated almost similar performance in terms of accuracy and specificity. In summary, quantitative analysis based on the total area of tumors further demonstrates the effectiveness and robustness of the proposed architecture.

5.6. Liver Tumor Segmentation Validation with a Radiologist

The segmentation accuracy of the proposed architecture is further validated by an experienced radiologist to ensure the effectiveness of the model in real clinical environments. During the visualization of the segmentation mask, we noticed that the ground truth mask of some samples did not correctly annotate the complete area of the tumor. And some tumor regions were not annotated in the ground truth. Furthermore, there were some tumor boundary inaccuracies in the tumor samples with fuzzy tumor boundaries. In those cases, evaluation metrics cannot provide correct evaluation referring to the ground truth due to erroneous ground truth masks.

An experienced radiologist in our research team participated in the liver tumor segmentation mask validation based on real clinical methods to recognize liver tumor regions. We have selected only 18 samples that represent all the confusion and issues encountered during the qualitative analysis to evaluate the segmentation performance. We have presented the liver tumor sample including the original CT image, ground truth region on the CT image, predicted liver tumor region by UNet on the CT image, and predicted liver tumor region by proposed model on the CT image. The comment of the radiologist is comprehensively explained under each sample and the comparison between UNet and the proposed model is shown in Table 18.
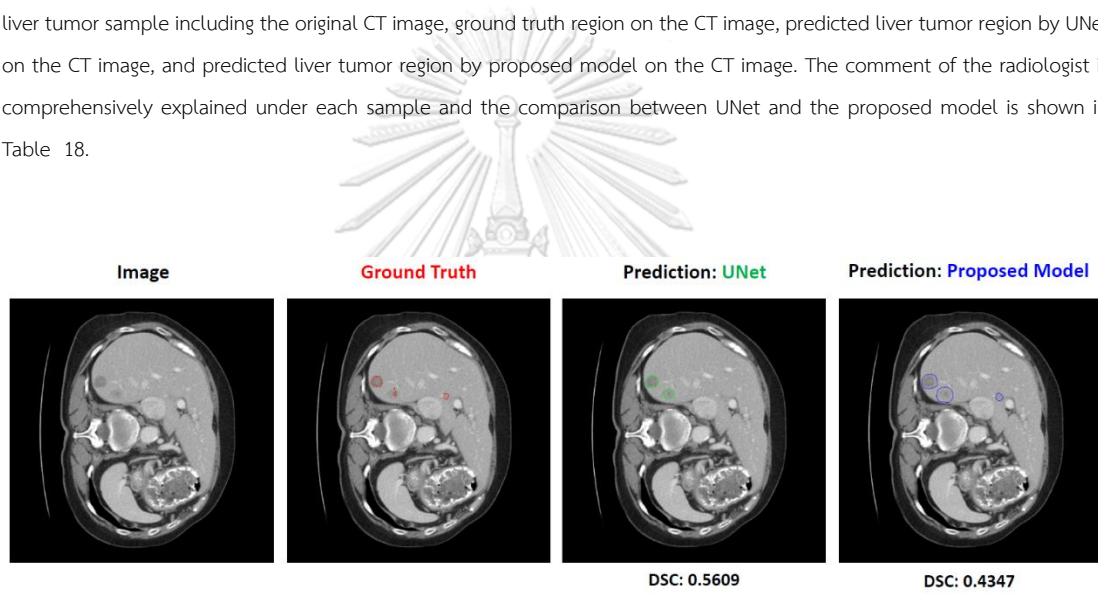
Figure 46: Liver tumor segmentation sample 1.

The comment of the radiologist for sample 1 (Figure 46): One tumor is not correctly annotated in the ground truth. But the UNet and proposed model have recognized the tumor which was partially annotated in the ground truth. the comparison between the UNet and the proposed model, the segmentation of the UNet is slightly smaller than the actual tumor region and the segmentation of the proposed model is slightly bigger than the actual tumor region. However, this variability can appear with the tumor segmentation by two radiologists (Interobserver variability). The tumor segmentation of the proposed model is better than UNet in recognizing tumor regions and recognizing all the tumors present in the CT image.

Figure 47: Liver tumor segmentation sample 2.

The comment of the radiologist for sample 2 (Figure 47): The segmentations from the UNet and the proposed model are acceptable. Those two segmentations are similar to interobserver variability between two radiologists. The proposed model has segmented the tumor region without error segmentations. The segmentation of the proposed model is comparatively better than UNet.
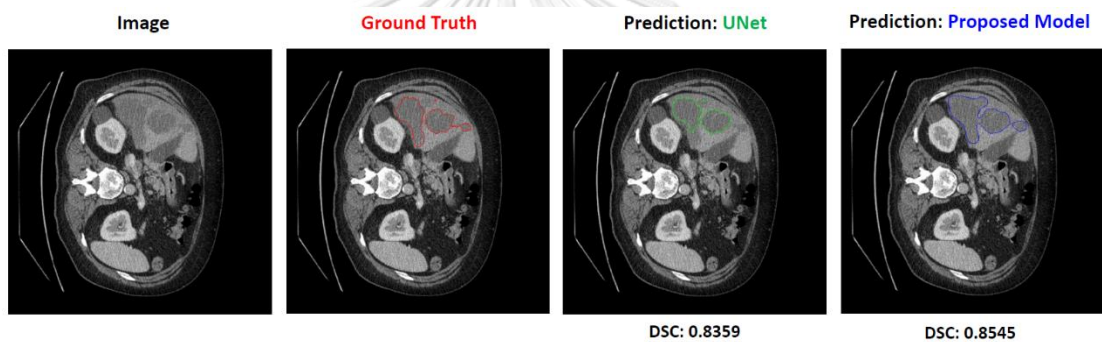
Figure 48: Liver tumor segmentation sample 3.

The comment of the radiologist for sample 3 (Figure 48): The ground truth does not completely cover the tumor region. However, the proposed model completely segmented all the regions of the tumor. The UNet partially segmented the liver tumor while the proposed model completely segmented all the tumor regions.
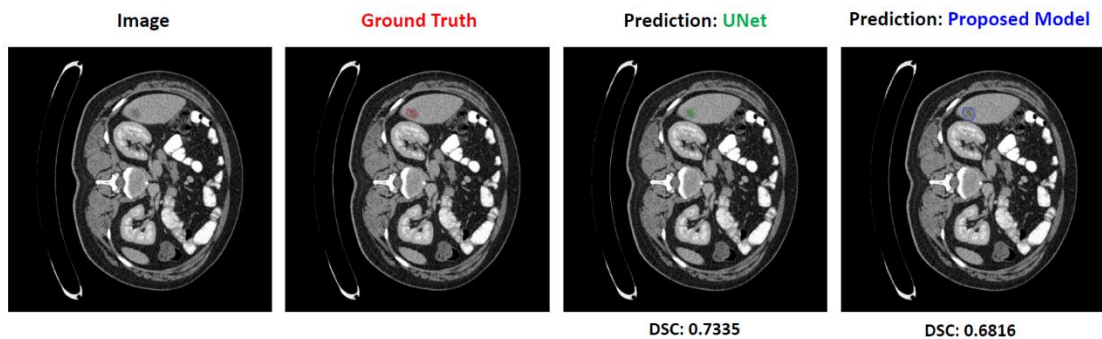
Figure 49: Liver tumor segmentation sample 4.

The comment of the radiologist for sample 4 (Figure 49): The ground truth does not include the fuzzy boundary of the tumor. And the UNet has followed the ground truth. However, the proposed model has segmented the whole area of the tumor including the fuzzy boundary of the tumor.

Figure 50: Liver tumor segmentation sample 5.

The comment of the radiologist for sample 5 (Figure 50): The segmentations from the UNet and the proposed model are acceptable. There is no significant difference in segmentation mask between the UNet and the proposed model. However, the proposed model has included the fuzzy boundary to the segmentation mask.

Figure 51: Liver tumor segmentation sample 6.

The comment of the radiologist for sample 6 (Figure 51): The ground truth does not include all the areas of the tumor. And the UNet has followed the ground truth. The proposed model has recognized all the tumor regions even better than the ground truth.
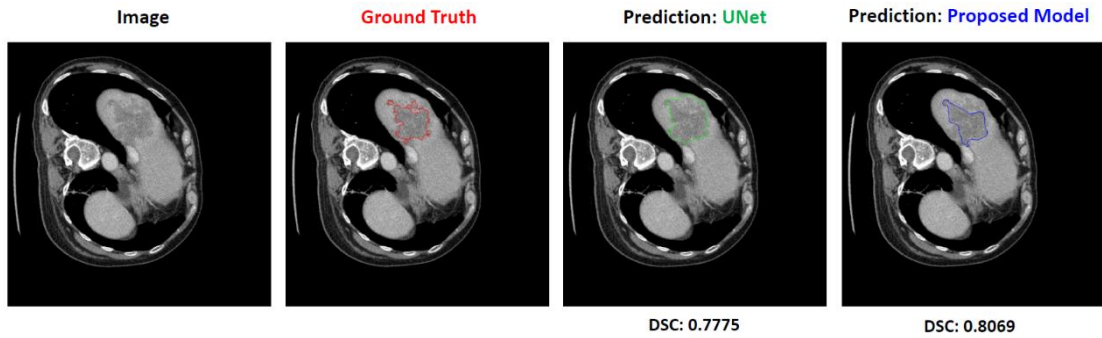
Figure 52: Liver tumor segmentation sample 7.

The comment of the radiologist for sample 7 (Figure 52): The ground truth does not cover the whole tumor region. The proposed model has partially segmented the tumor region. However, the UNet has segmented the whole area of the tumor.
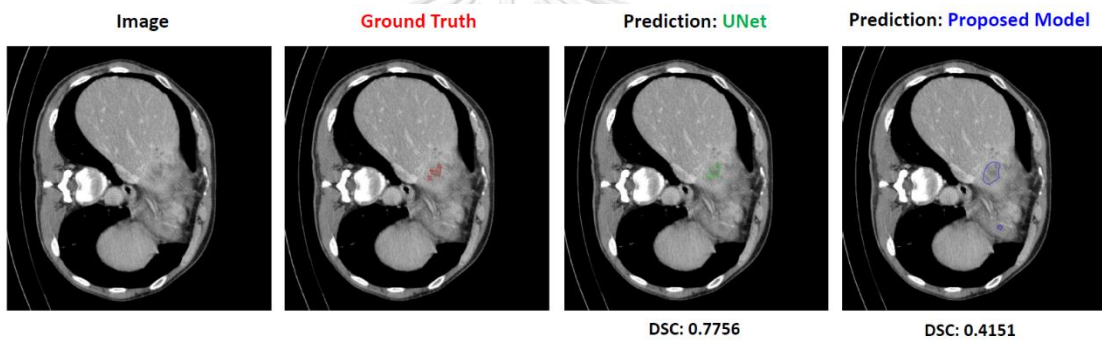
Figure 53: Liver tumor segmentation sample 8.

The comment of the radiologist for sample 8 (Figure 53): The ground truth does not include all the areas of the tumor. The UNet and the proposed model followed the ground truth without recognizing all the areas of the tumor.
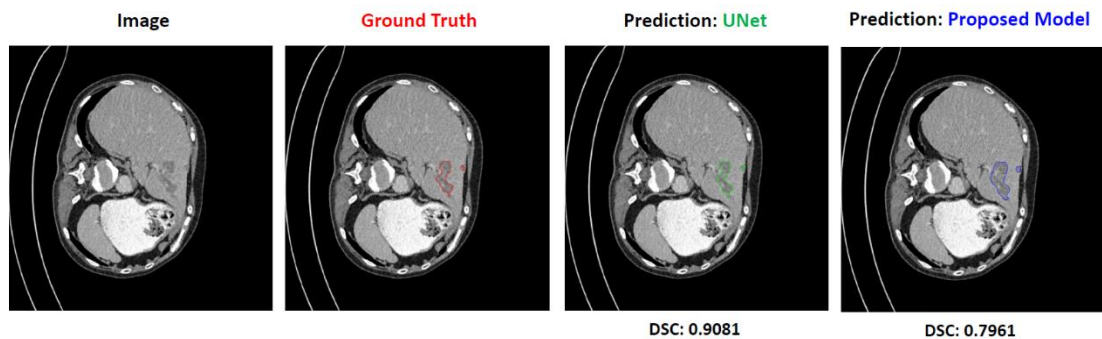
Figure 54: Liver tumor segmentation sample 9.

The comment of the radiologist for sample 9 (Figure 54): The proposed model has segmented the complete tumor region comparatively better than UNet.
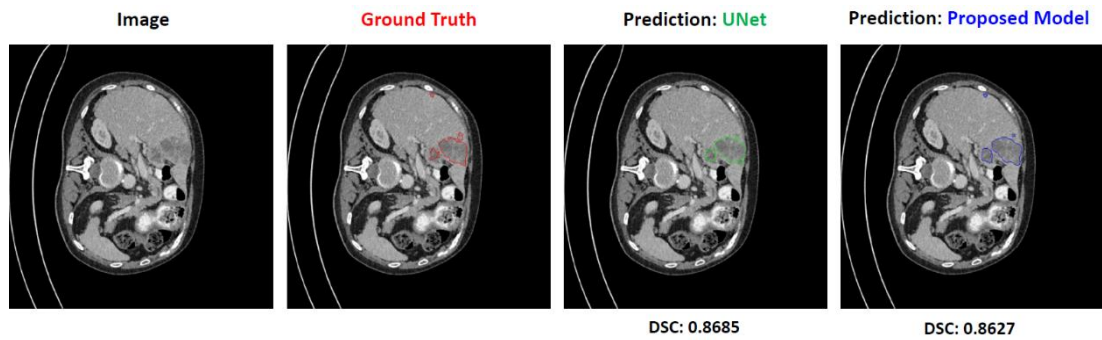
Figure 55: Liver tumor segmentation sample 10.

The comment of the radiologist for sample 10 (Figure 55): The proposed model has correctly segmented with the inclusion of all the tumors while UNet failed to capture all the tumors in the CT image.
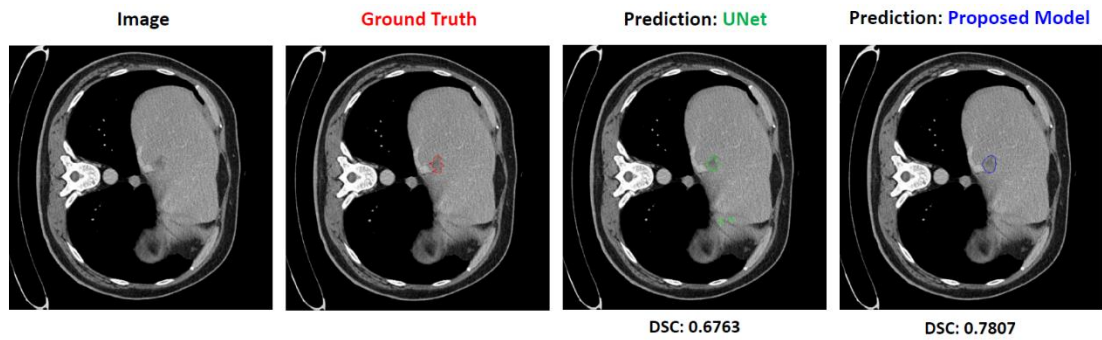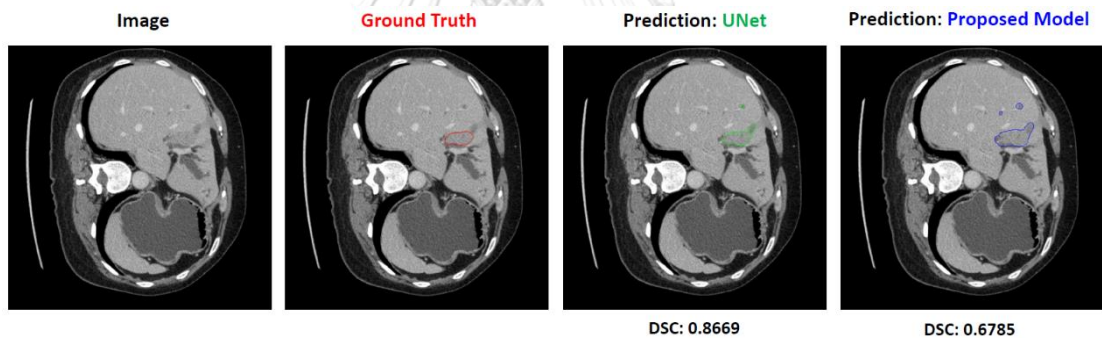

Figure 56: Liver tumor segmentation sample 11.

The comment of the radiologist for sample 11 (Figure 56): The UNet and the proposed model have recognized the tumor region. However, both models have recognized bile duct dilatation as a tumor region. The bile duct dilatation region cannot be recognized accurately by one CT image. A few consecutive slices should be observed to recognize the exact region of bile duct dilatation to exclude from the tumor region. Most probably the actual tumor area is similar to the tumor region in the ground truth.
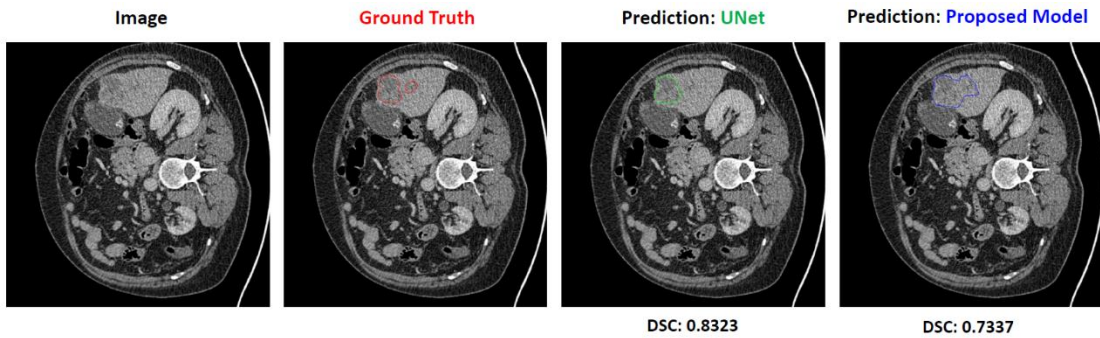
Figure 57: Liver tumor segmentation sample 12.

The comment of the radiologist for sample 12 (Figure 57): This tumor is quite contiguous. Two tumors can be connected. The proposed model has segmented the whole area of the tumor compared to UNet.
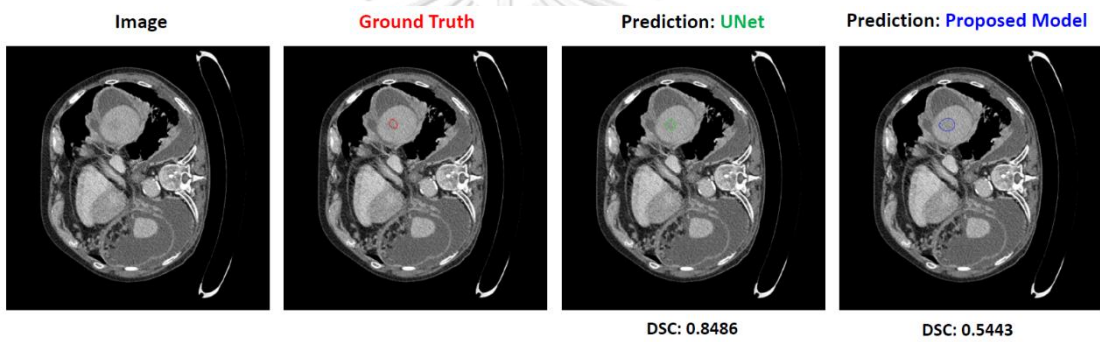
Figure 58: Liver tumor segmentation sample 13.

The comment of the radiologist for sample 13 (Figure 58): The tumor has spread to an area that is bigger than the ground truth. The proposed model has recognized the whole tumor area correctly while UNet follows the ground truth.
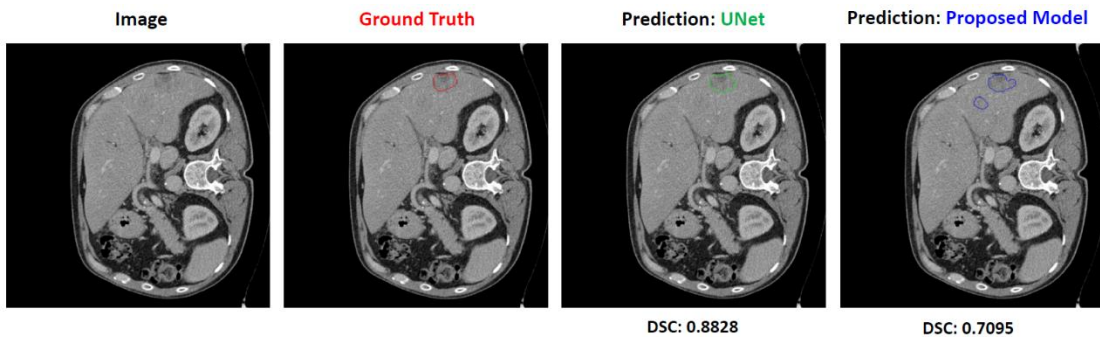
Figure 59: Liver tumor segmentation sample 14.

The comment of the radiologist for sample 14 (Figure 59): The UNet and the proposed model have recognized the tumor region annotated in the ground truth. There should be one more tumor region that has been recognized by the proposed model. The proposed model has accurately recognized all the tumor regions in the CT image.
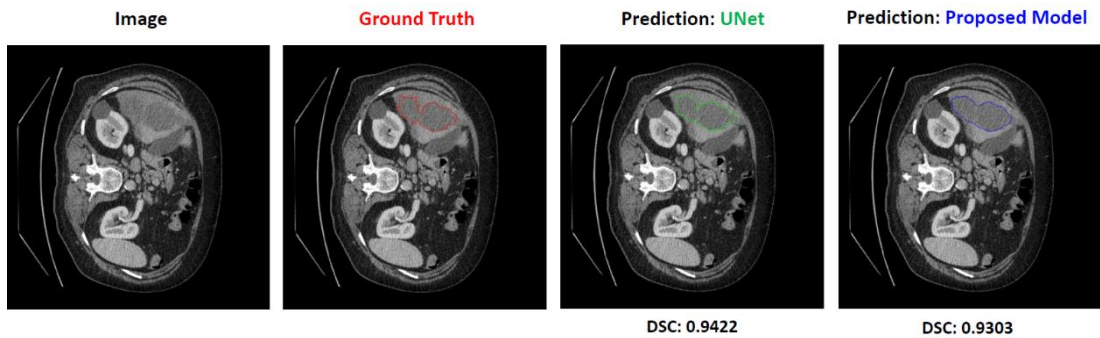
Figure 60: Liver tumor segmentation sample 15.

The comment of the radiologist for sample 15 (Figure 60): The proposed model has segmented the complete tumor region comparatively better than UNet.
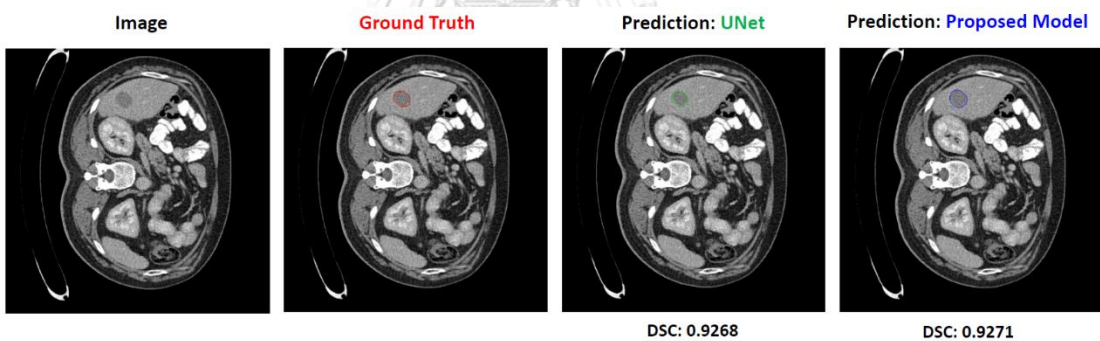
Figure 61: Liver tumor segmentation sample 16.

The comment of the radiologist for sample 16 (Figure 61): The segmentations from the UNet and the proposed model are acceptable. There is no significant difference in segmentation mask between the UNet and the proposed model.
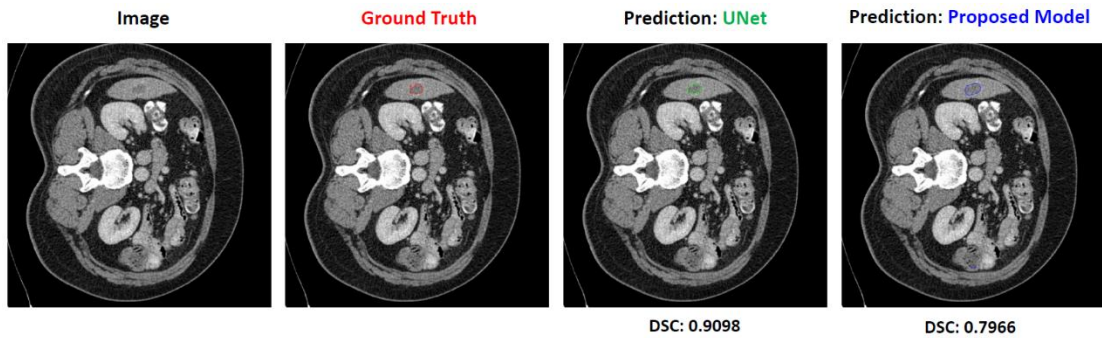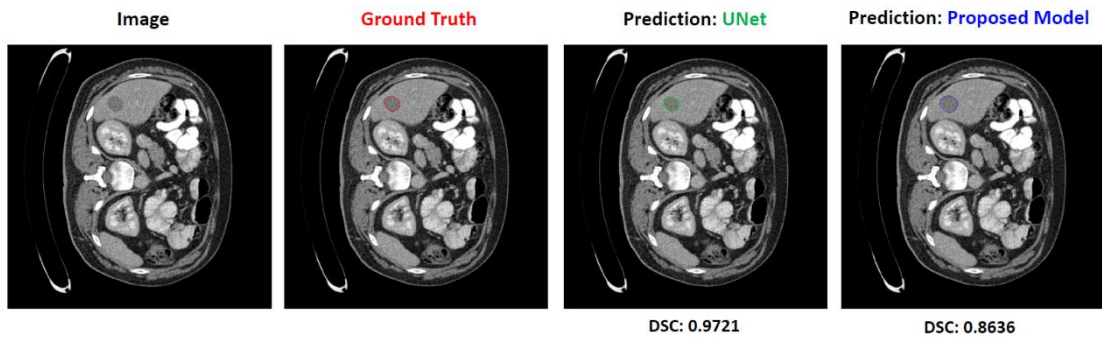
Figure 62: Liver tumor segmentation sample 17.

The comment of the radiologist for sample 17 (Figure 62): The segmentations from the UNet and the proposed model are acceptable. There is no significant difference in segmentation mask between the UNet and the proposed model.
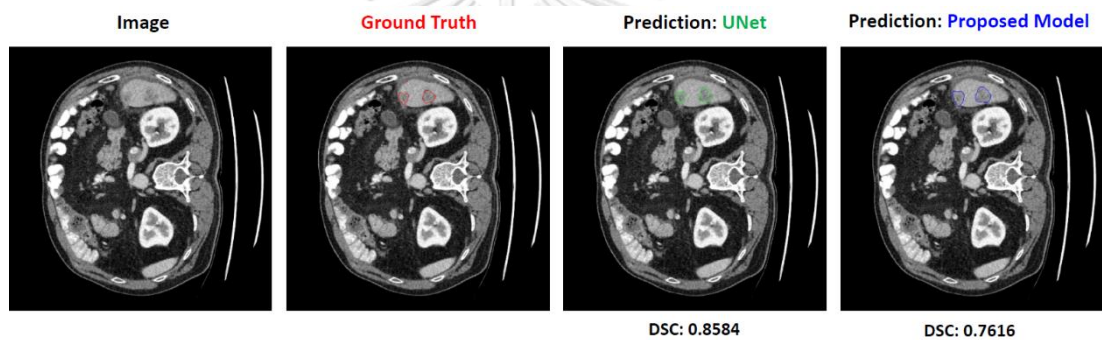
Figure 63: Liver tumor segmentation sample 18.

The comment of the radiologist for sample 18 (Figure 63): The segmentations from the UNet and the proposed model are acceptable. However, the proposed model has included the fuzzy boundary to the segmentation mask and is comparatively better than UNet.
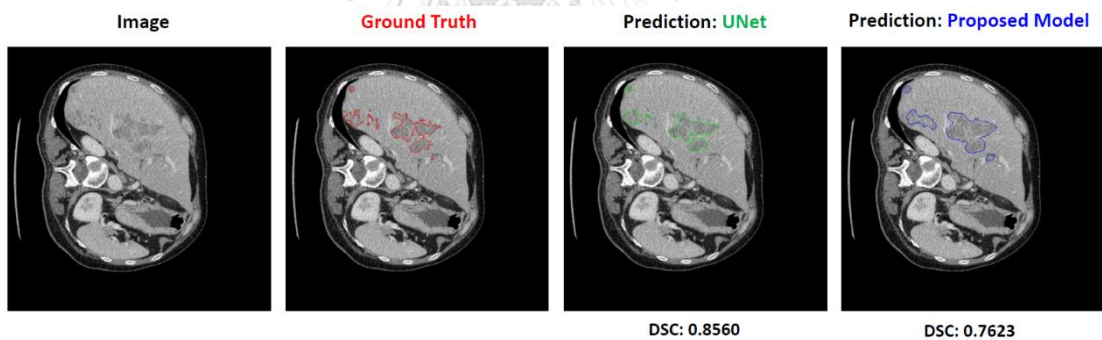
Validation sample	Remark of the Radiologist		
	Poor than UNet	Same as UNet	Better than UNet
Sample 1			✓
Sample 2			✓
Sample 3			✓
Sample 4			✓
Sample 5		✓	
Sample 6			✓
Sample 7	✓		
Sample 8		✓	
Sample 9			✓
Sample 10			✓
Sample 11		✓	
Sample 12			✓
Sample 13			✓
Sample 14			✓
Sample 15			✓
Sample 16		✓	
Sample 17		✓	
Sample 18			✓

Table 18: The remark of the radiologist for the liver tumor segmentation results validation.

The liver segmentation results validation proves that the proposed model has demonstrated better performance according to real clinical evaluation by an experienced radiologist. In particular, the proposed model can segment the tumor region with accurate edge precision. The recognition of tumors with fuzzy tumor boundaries can be challenging even for a radiologist. However, the proposed model has proven superior in recognizing tumors with fuzzy boundaries.

In some cases, the radiologist mentioned that the proposed model can recognize tumor regions even better than the ground truth. It further indicates that some ground truths have errors in tumor boundary annotation. Due to the issues in ground truth, accurate segmentation can be misevaluated by the evaluation metrics used in quantitative analysis. Furthermore, the validation of the results proves the superiority of the proposed model in liver tumor segmentation. The performance of the model can be maximized by training and evaluating the model with high-quality data and tumor region annotations.

5.7. Qualitative Analysis of Segmentation Mask

The qualitative analysis additionally proves essential for evaluating the performance and effectiveness of the proposed model for the tumor segmentation task. We performed qualitative analysis by classifying segmentations into four categories: large tumors, minor tumors, poor segmentations, and over/non-segmentations. The liver tumor segmentation samples of the proposed MANet and other comparison networks from the volume-based and slice-based liver tumor segmentation experiments are illustrated in Figure 64, Figure 65, and Figure 66. The liver tumor segmentation samples of the proposed MANet and state-of-the-art networks from slice-based liver tumor segmentation experiments are illustrated in Figure 67.

5.7.1. Qualitative Analysis of Comparison Networks

Qualitative Analysis is conducted to evaluate the correctness of the tumor boundary identification and the accurate localization of the tumor segmentation. It is easily noticeable that all the models demonstrated acceptable segmentation performance for large tumor segmentation. However, in the first large tumor sample in both slice-based and volume-based cases, a partial tumor segmentation has been shown by the comparison models while the proposed MANet performed accurate segmentation that is almost identical to its ground truth (see Figure 64(1-1) and Figure 65(1-1)). All the networked has predicted two large tumors as one tumor in the second large tumor sample in Figure 64(1-2) and Figure 65(1-2), this segmentation is performed with slight false positive in accordance with the ground truth. The fuzzy boundary of the tumor can result in comparatively minor over-segmentation or false positives in the segmentation. Minor over-segmentation blobs appeared in comparison models except the UNet+Resunet18 model in the experiment based on slices. A slight over-segmentation has commonly presented in all the models in volume-based experiment results (see Figure 64(1-2) and Figure 65(1-2)). In general, the proposed network has predicted almost similar segmentation masks to the ground truth in large tumor cases in both slice-based and volume-based segmentation experiments.

Computer-aided diagnosis (CAD) systems for liver tumor segmentation commonly endure poor performance in small tumor segmentation. Achieving greater performance in small tumor segmentation is still challenging not only for CAD systems but also for experienced radiologists. In particular, small tumor recognition or segmentation is paramount to treat the disease in an earlier stage in order to achieve better survival. It is easily noticeable that the proposed MANet model has proven its small tumor segmentation capability with better accuracy compared to the baseline methods (see Figure 64(2-1, 2-2) and Figure 65(2-1, 2-2)). In the segmentation performance visualization in Figure 64 and Figure 65, we have selected two tumor segmentation samples under large tumor and small tumor categories in both slice-based and volume-based experiment test sets (i.e., Figure 64(1-2, 2-2) and Figure 65(1-2, 2-2)). Those two common samples in both experiments have demonstrated the robustness of the proposed MANet model's liver tumor segmentation in both experiments. Furthermore, UNet+CBAM, which used attention at each level of the network in the same way as the proposed MANet, presented nearly identical segmentation performance in most of the cases in Figure 64 and Figure 65. A poor segmentation prediction showed the failure to accurately imitate ground truth and partial segmentations (see Figure 64(3-1) and Figure 65(3-1). In this instance, the proposed model could segment all tumor regions with comparatively lower edge precision, whereas baseline methods failed to predict all the tumor regions.

For making accurate clinical management decisions, recognizing all the tumors present in the CT image is necessary. Complete survival or recovery may not be achieved in case of missing small tumors in the radiological scan due to poor

clinical management decision-making. To assess the capability of the proposed design to segment tumors in multiple tumor cases, randomly selected multiple tumor samples were visualized in Figure 66. In this illustration, we recognize that all of the models are capable of segmenting large tumors with enough accuracy and edge precision. Nonetheless, in the sample CT slices, the majority of the baseline models were struggling to segment small tumors. Among the comparison models, UNet and UNet+Resnet18 models commonly illustrated partial or missing segmentation for small tumors in multiple cases. However, the baseline models that were designed with attention mechanisms (i.e., Attention UNet and UNet+CBAM) demonstrated their capabilities in capturing all the tumors with or without better edge precision, as almost similar to the segmentation of the proposed MANet model (see Figure 66(2, 3)). Furthermore, Figure 66(2) has shown a poor segmentation sample from the proposed model compared to other randomly selected samples illustrated in Figure 66. However, the proposed model depicts comparatively better segmentation than other comparison models. In short, we can see that the formation of the proposed MANet architecture with attention mechanisms is significantly more effective than the baseline models in performing accurate and stable liver tumor segmentation.

5.7.2. Qualitative Analysis of state-of-the-art Networks

We visualized the same samples that were shown in slice-based segmentation in Figure 64 to make a fair comparison with state-of-the-art approaches. Figure 67 depicts the segmentation performance of state-of-the-art models. In general, all the state-of-the-art approaches show almost identical segmentation performance in small tumor samples. However, TA-Net and UNet 3+ models performed the segmentation with slight over-segmentation in large tumor samples which can be seen in Figure 67(1,2). ResUNet++ and SmaAt-Net were not able to capture the tumor boundaries with greater edge precision in large tumor segmentation. Minor under-segmentation appeared in Figure 67(1), which is classified as false negatives. It is noticeable that the models could not perform accurate segmentation in multiple tumor cases in common. However, the proposed MANet model could recognize and segment all the tumor regions compared to the state-of-the-art approaches with acceptable accuracy.
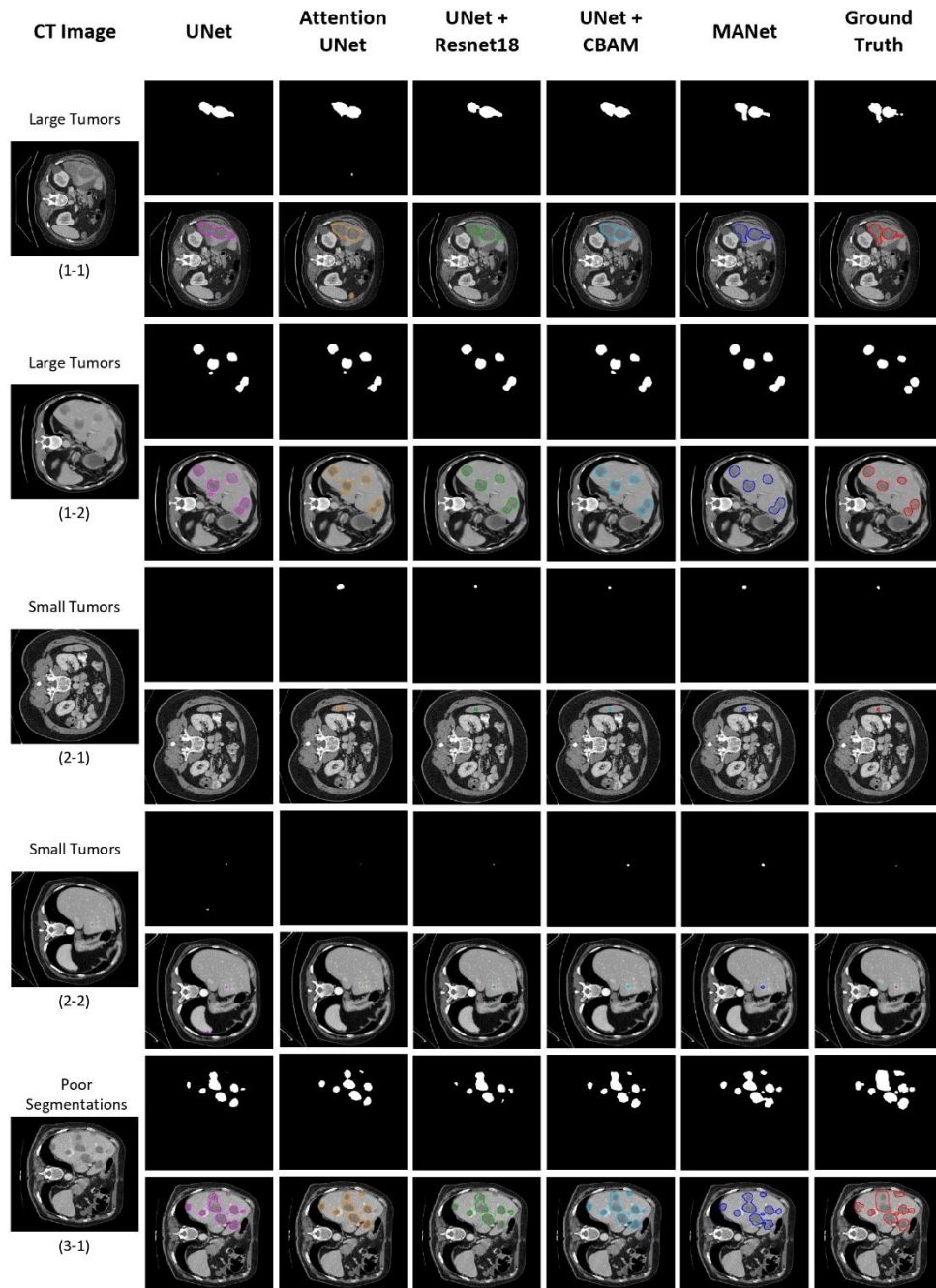
Figure 64: The qualitative analysis of liver tumor segmentation performance of the proposed MANet model and comparison models from the slice-based segmentation experiment. The contour image of the liver tumor segmentation is generated and illustrated right below the liver tumor segmentation mask to visualize the liver tumor boundary. From left to right: the original CT image, results obtained by UNet (pink), Attention UNet (orange), UNet+Resnet18 (green), UNet+CBAM (cyan), MANet (blue), and the corresponding ground truth mask (red). Large tumors, Small tumors, and Poor Segmentation are the three different perspectives that are illustrated.
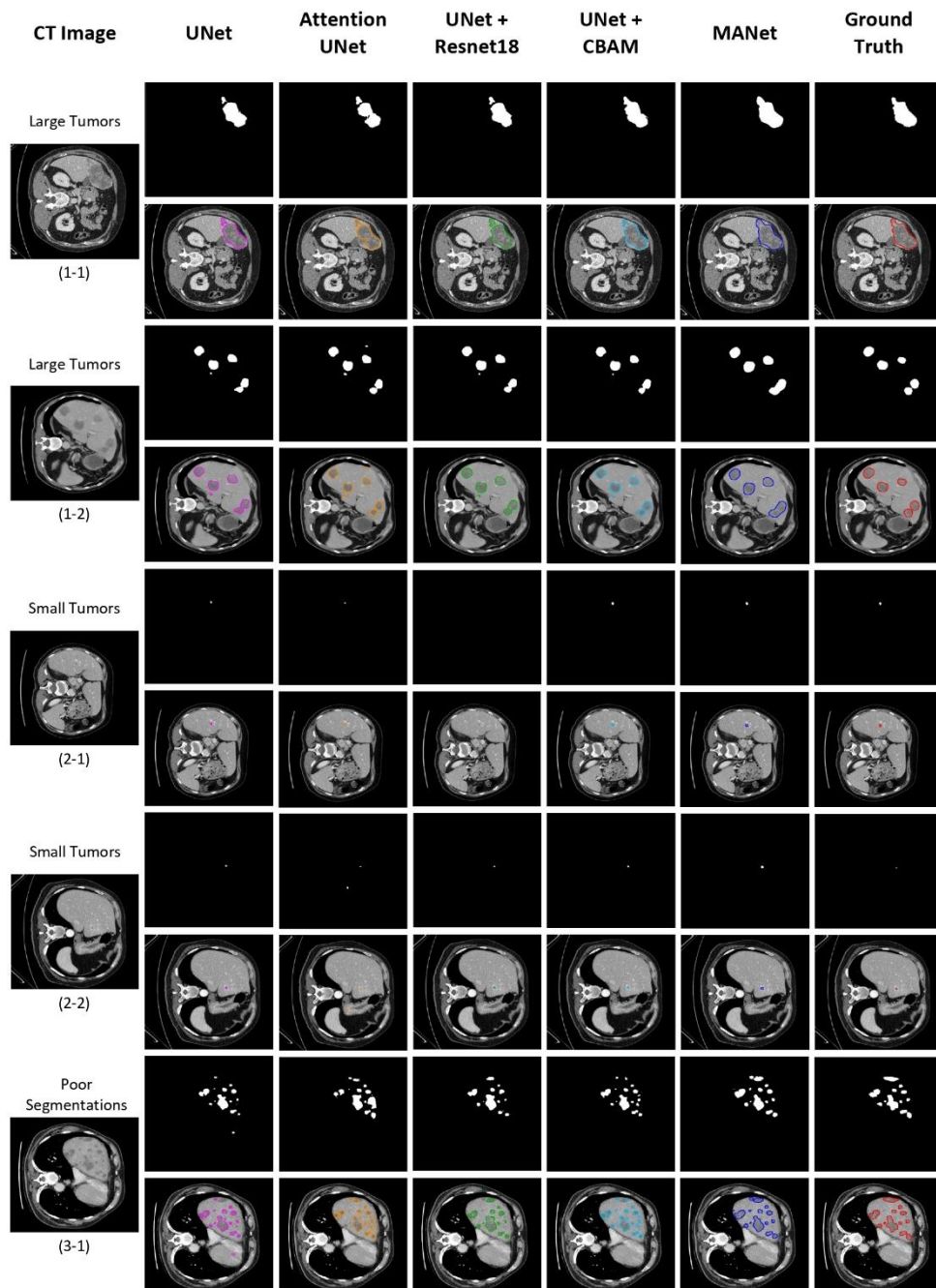
Figure 65: The qualitative analysis of liver tumor segmentation performance of the proposed MANet model and comparison models from the volume-based segmentation experiment. The contour image of the liver tumor segmentation is generated and illustrated right below the liver tumor segmentation mask to visualize the liver tumor boundary. From left to right: the original CT image, results obtained by UNet (pink), Attention UNet (orange), UNet+Resnet18 (green), UNet+CBAM (cyan), MANet (blue), and the corresponding ground truth mask (red). Large tumors, Small tumors, and Poor Segmentation are the three different perspectives that are illustrated.
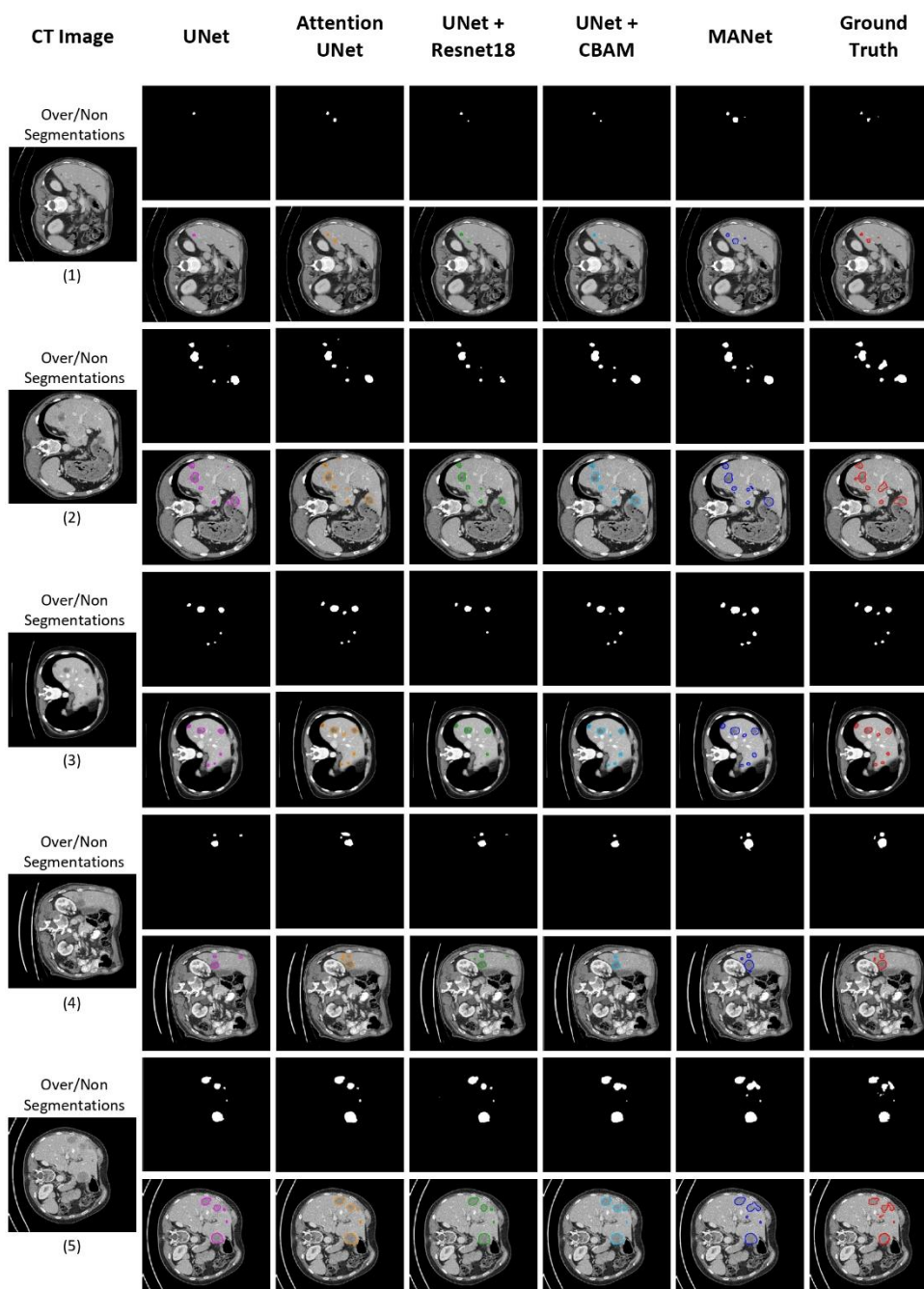
Figure 66: The qualitative analysis of liver tumor segmentation performance in over/non-segmentations in multiple tumor cases of the proposed MANet model and comparison models from the slice-based segmentation experiment. The contour image of the liver tumor segmentation is generated and illustrated right below the liver tumor segmentation mask to visualize the liver tumor boundary. From left to right: the original CT image, results obtained by UNet (pink), Attention UNet (orange), UNet+Resnet18 (green), UNet+CBAM (cyan), MANet (blue), and the corresponding ground truth mask (red). Multiple tumor cases of five different samples are illustrated.
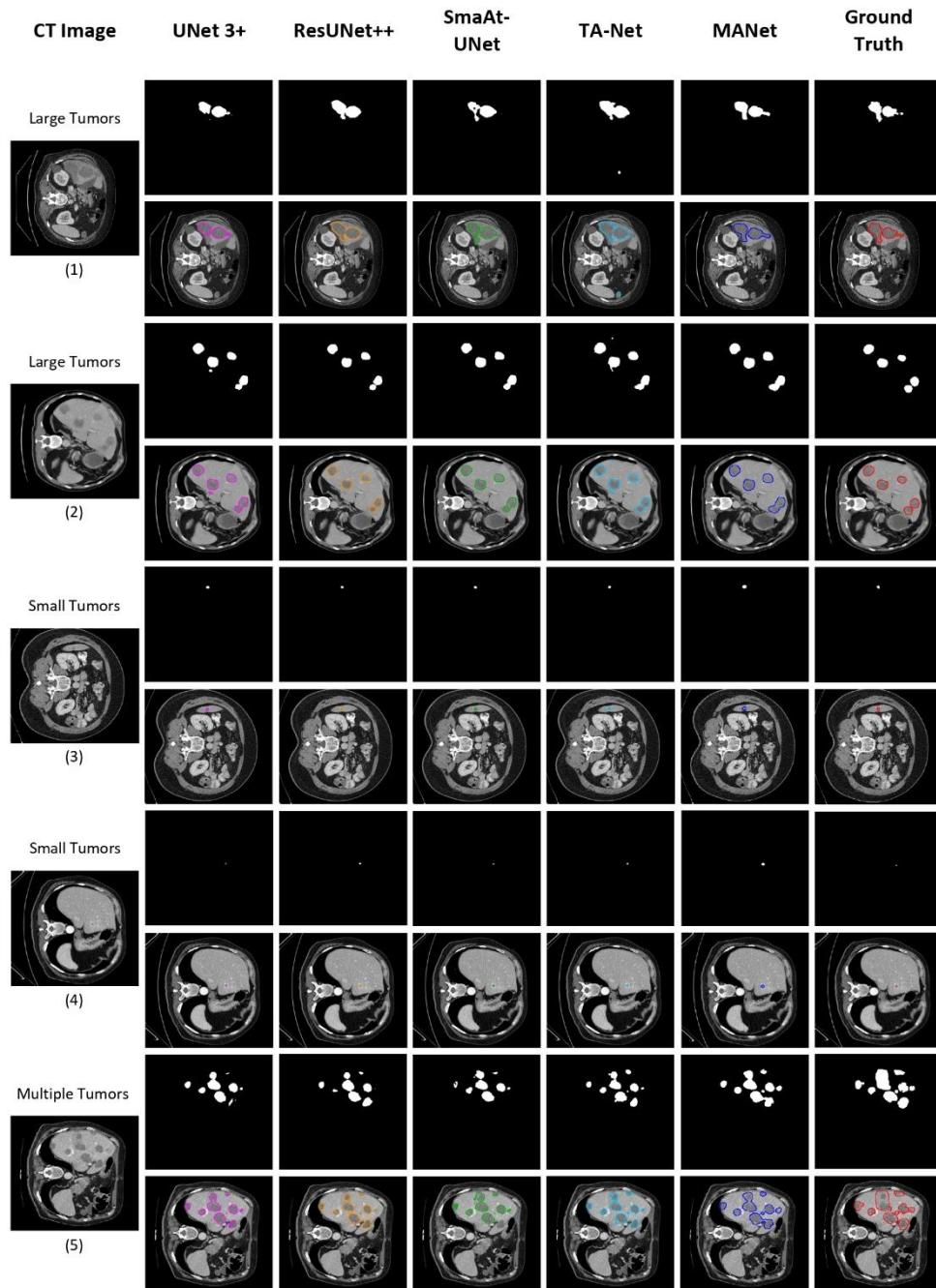
Figure 67: The qualitative analysis of liver tumor segmentation performance of the proposed MANet model and state-of-the-art models from the slice-based segmentation experiment. The contour image of the liver tumor segmentation is generated and illustrated right below the liver tumor segmentation mask to visualize the liver tumor boundary. From left to right: the original CT image, results obtained by UNet 3+ (H. Huang et al., 2020) (pink), ResUNet++ (Jha et al., 2019) (orange), SmaAt-UNet (Kevin et al., 2021) (green), TA-Net (Shuchao et al., 2021) (cyan), MANet (blue), and the corresponding ground truth mask (red). Large tumors, Small tumors, and Poor Segmentation are the three different perspectives that are illustrated.

5.8. Model Feature Visualization

The proposed MANet architecture is designed by utilizing the strengths of the attention mechanisms. To assess the effectiveness of the proposed architecture that formed implementing attention mechanisms in all the levels of the network, features of the network are visualized in Figure 68. To compare the feature propagation with comparative networks, the corresponding initial encoder block feature maps and final decoder block feature maps were visualized in Figure 69. All feature maps are constructed by mapping features between their maximum and minimum values.
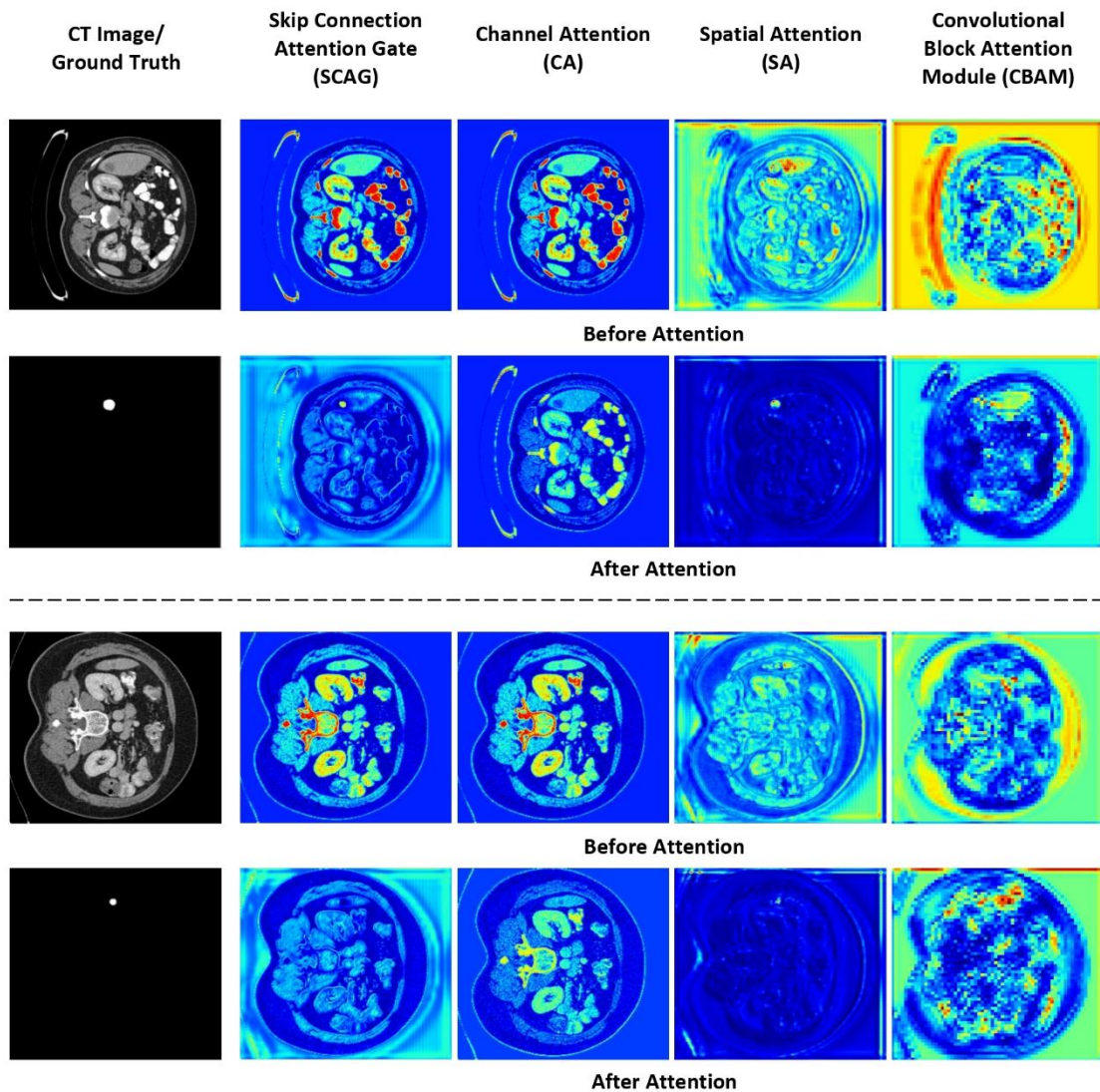
Figure 68: Feature visualization before and after the MANet architecture's Skip Connection Attention Gate (SCAG), Channel Attention (CA), Spatial Attention (SA), and Convolutional Block Attention Module (CBAM).

| CT Image/ Ground Truth | UNet | Attention UNet | UNet + Resnet18 | UNet + CBAM | MANet |

Initial Encoder Block Features

Final Decoder Block Features

Initial Encoder Block Features

Final Decoder Block Features
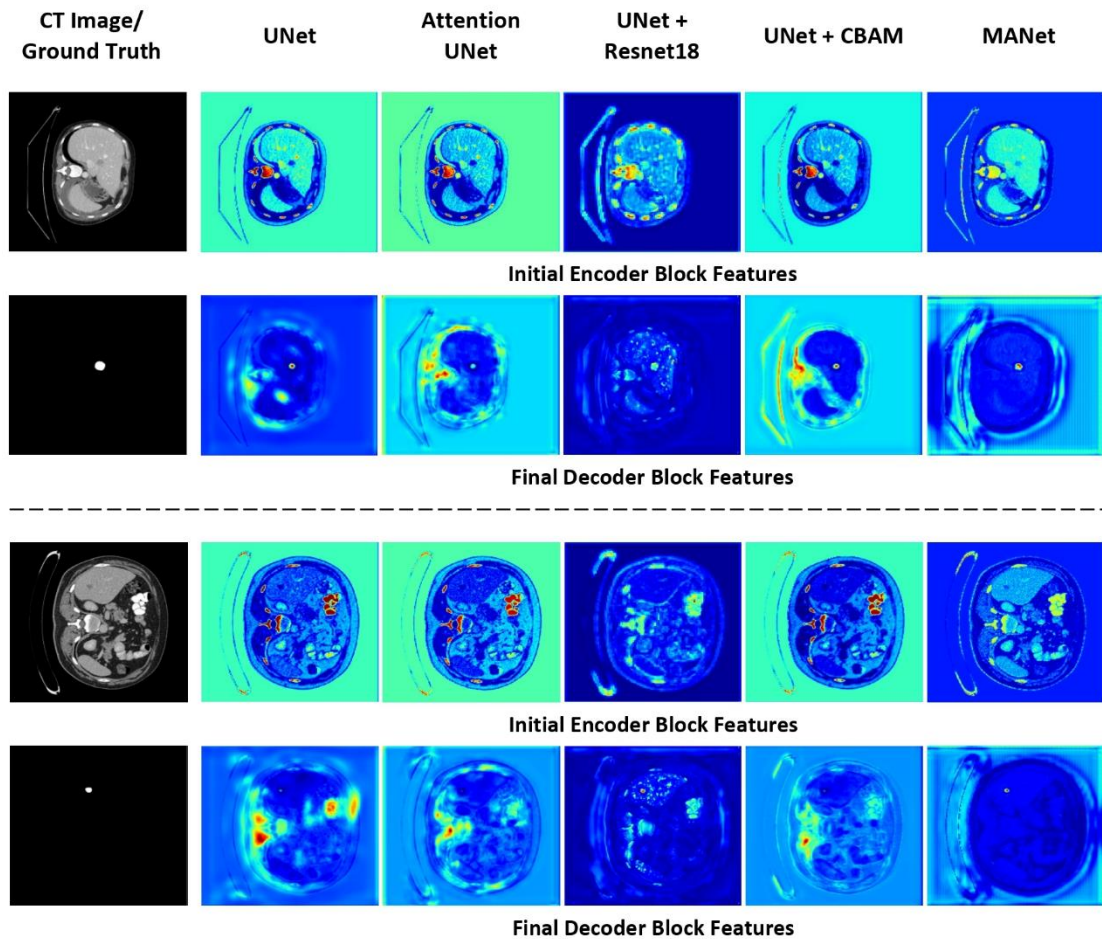
Figure 69: Feature visualization of corresponding initial encoder block and final decoder block feature maps of comparison networks.

The proposed architecture is comprised mainly of two attention mechanisms, those are channel-wise attention mechanisms and spatial-wise attention mechanisms. The channel attention mechanism highlights important features while suppressing irrelevant features to liver tumor segmentation in the channel dimension. The convolutional block attention module (CBAM) initiates focusing on the region of interest (ROI) for tumor segmentation at the deepest stage of the network. The network continued to retain contextual features in the decoder path by the spatial attention mechanism that was implemented in all the levels of the decoder path. The effectiveness of all the attention mechanisms in the network is illustrated by calculating the feature maps of the network before and after the attention mechanisms (see Figure 68). As illustrated in Figure 68, spatial attention (SA) significantly suppressed irrelevant features in the spatial dimension. The skip connection attention gate (SCAG) that is implemented in skip connection is designed with a spatial attention mechanism. It creates the focus on ROI by extracting important features from the encoder path to concatenate with corresponding deep features in the decoder.

The feature map visualization is further continued with comparison models to compare the effectiveness of the proposed architecture among comparison models (see Figure 69). The initial encoder block feature and the final decoder block features are visualized to compare the networks. The proposed MANet model shows a greater focus on the ROI through

the feature propagation process while suppressing irrelevant regions for the segmentation task. Initial encoder block features depict the effectiveness of the channel attention mechanism to enhance the segmentation performance compared to the comparison models. The retention of important features in spatial dimension is visualized in final decoder block features and it proves the superiority of the proposed architectural design with spatial attention. In summary, the proposed MANet architecture depicts a substantially better interpretation of liver tumor segmentation.

5.9. Computational Cost Analysis

To contrast the proposed model among other comparison models in terms of computational cost, computational complexity, total parameter counts, and the inference time per slice are calculated in Table 19. The UNet model demonstrated minimum computational complexity. However, it utilizes a significantly higher amount of memory due to a comparatively large number of parameters which is slightly similar to double of parameters in the proposed MANet architecture. The Attention UNet is lighter in terms of parameter count (i.e., 6.34 M) with minor inference time. However, the computational complexity is slightly higher than the base model UNet. The UNet+Resnet18 model has the highest total parameter count (i.e., 17.85 M) while archiving the best inference time of 36.2 ms.

The highest computational complexity is 166.8 GMac, which is the model comprised of CBAM (i.e., UNet+CBAM model). Furthermore, the proposed MANet model computational complexity is almost closer to the UNet+CBAM model. It is clear that the proposed model has comparatively higher computational complexity due to the implementation of channel attention, spatial attention, and CBAM in the network in all stages. However, we could reduce the computational complexity of the proposed model compared to UNet+CBAM, using channel attention and spatial attention separately in the encoder and decoder path respectively, rather than applying CBAM to all the stages of the network. In summary, the proposed MANet architecture manifested significantly better segmentation performance by utilizing approximately half of the parameter count of the base model of UNet.

Network	Computational complexity (MACs(G))	Total training parameters (M)	Inference time (ms)
UNet	**94.45**	13.37	41.60
Attention UNet	97.07	**6.34**	38.80
UNet + Resnet18	119.14	17.85	**36.20**
UNet + CBAM	166.80	8.39	82.60
MANet (Proposed model)	132.37	7.83	81.80

Table 19: Analysis of computational costs based on computational complexity, total training parameters, and inference time. The best values are in bold.

5.10. Ablation Analysis

The effectiveness of the proposed MANet architecture design is evaluated with an ablation analysis under 8 steps (Table 20). UNet, the base model for the proposed design evaluated in the first step of the ablation analysis. The convolutional blocks in UNet were redesigned by integrating the residual learning function to create the UNet+RB model for the second experiment. After that, each attention mechanism implemented in the proposed architecture was separately implemented to the UNet+RB model to conduct three experiments (i.e., No.3, 4,5, and 6) by selecting UNet+RB as the backbone of the proposed network. Moreover, the UNet model was the base model that used to evaluate the effectiveness of all the attention mechanisms and techniques in the experiments. The effectiveness of all the attention mechanisms shown in experiment No. 7, further indicates that the residual learning has demonstrated a significant impact on the performance of the proposed MANet architecture.

No	Method	Dice score	ASSD	Jaccard index (IoU)	VOE	Accuracy	Sensitivity (Recall)	Specificity
1	UNet	0.7522 ± 0.178	1.4342 ± 1.320	0.6310 ± 0.190	0.3606 ± 0.190	0.9928 ± 0.006	0.8425 ± 0.204	0.9956 ± 0.003
2	UNet + RB	0.7533 ± 0.182	1.5172 ± 1.395	0.6359 ± 0.192	0.3640 ± 0.192	0.9925 ± 0.006	0.8512 ± 0.202	0.9951 ± 0.004
3	UNet + RB + SCAG	0.7532 ± 0.195	1.4247 ± 1.298	0.6353 ± 0.202	0.3646 ± 0.202	0.9927 ± 0.006	0.8329 ± 0.224	0.9956 ± 0.003
4	UNet + RB + CA	0.8010 ± 0.155	1.0137 ± 1.000	0.6901 ± 0.177	0.3027 ± 0.177	0.9940 ± **0.004**	0.8708 ± **0.173**	0.9965 ± **0.002**
5	UNet + RB + SA	0.7550 ± 0.201	1.1610 ± 1.069	0.6389 ± 0.205	0.3610 ± 0.205	0.9929 ± 0.006	0.8292 ± 0.233	0.9958 ± 0.003
6	UNet + RB + CBAM	0.8006 ± 0.157	0.8842 ± 0.814	0.6897 ± 0.178	0.3038 ± 0.178	0.9938 ± **0.004**	0.8712 ± 0.180	0.9962 ± 0.003
7	UNet + SCAG + CA + SA + CBAM	0.8056 ± 0.153	0.8376 ± 0.733	0.6992 ± 0.174	0.3007 ± 0.174	0.9941 ± **0.004**	0.8715 ± 0.177	0.9967 ± 0.003
8	**MANet: UNet + RB + SCAG + CA + SA + CBAM**	**0.8145 ± 0.150**	**0.7084 ± 0.701**	**0.7084 ± 0.171**	**0.2915 ± 0.171**	**0.9947 ± 0.004**	**0.8723 ± 0.173**	**0.9970 ± 0.002**

Table 20: Ablation analysis for the proposed MANet architecture. The result from MANet and the best values are in bold.

The results of the ablation analysis show that the developments of the proposed MANet architecture are substantially beneficial to the liver tumor segmentation performance of the network. In experiment No.2, the residual structure showed slight improvements. It has further proved the crucial importance of residual blocks in experiment No.7. Because the integration of all the attention mechanisms in UNet failed to outperform the proposed MANet that comprised of residual structure. Among the attention mechanisms, channel attention (CA) delivered a significant performance boost to the proposed MANet architecture. It has further validated in experiment No. 6, the model designed with convolutional block attention module (CBAM) using the backbone of UNet+RB. It can be demonstrated that the combination of UNet, Residual block, and attention mechanism outperforms the integration of a single mechanism to the base model UNet. Since the fusion of all the mechanisms achieves better feature extraction compared to the integration of a single mechanism to the base model UNet. We can conclude that combining deep learning techniques in MANet could significantly improve the performance of liver tumor segmentation.

6. Conclusion and Future Direction

We have presented a novel architectural design called a multi-attention network (MANet) comprised of channel and spatial attention mechanisms and residual learning. The purpose of the development is to automatically segment liver tumors by Computed tomography (CT) image of the hepatobiliary phase. This system is highly beneficial for radiologists to diagnose liver tumors with better accuracy in order to perform tumor burden analysis and treatment planning. The proposed architecture is formed with the basis of UNet and strengthened with channel and spatial attention mechanisms and residual learning. Moreover, channel attention is employed to recalibrate features in the channel dimension while spatial attention to extract important features in the spatial dimension to achieve better focus on the location of the tumors. The spatial attention mechanism is implemented in skip connection to extract important features from low-level encoder features to concatenate with high-level semantic features in the decoder path. To address the degradation and learning errors while improving the gradient flow, residual learning is employed in the encoder of the network.

We have evaluated the effectiveness of the proposed design on the LiTS17 and 3DIRCADb datasets using slice-based and volume-based segmentation examinations, which indicated the superiority of our network over baseline approaches. Furthermore, we conducted an empirical study to compare the quantitative and qualitative studies in order to evaluate performance, which could be used to prove the efficiency and robustness of the attention techniques applied in the proposed network design. The effectiveness of the proposed architecture is assessed further in the ablation study under 8 steps. However, there was a considerable performance difference between slice-based segmentation and volume-based segmentation. Because of the substantial heterogeneity in data sources in terms of the shape and appearance of liver tumors and intensity variations, this should be one of the difficulties. Consequently, significant challenges must be addressed before the model can be generalized to perform volume-based segmentation in a real-world clinical environment.

We intend to analyze the model with other datasets in the future to validate its generalizability. Furthermore, we do experiments to assess the model's ability to segment the liver and other organs with tumors (i.e., kidney, renal tumors) using various medical imaging modalities such as MRI, PET, and US. Our goal is to create this architecture using state-of-the-art deep learning approaches to reduce computational complexity while boosting segmentation performance with a greater level of stability.

REFERENCES

Alirr, O. I. (2020). Deep learning and level set approach for liver and tumor segmentation from CT scans. *Journal of Applied Clinical Medical Physics*, *21*(10), 200-209. https://doi.org/https://doi.org/10.1002/acm2.13003

Alksas, A., Shehata, M., Saleh, G. A., Shaffie, A., Soliman, A., Ghazal, M., Khelifi, A., Khalifeh, H. A., Razek, A. A., Giridharan, G. A., & El-Baz, A. (2021). A novel computer-aided diagnostic system for accurate detection and grading of liver tumors. *Scientific Reports*, *11*(1), 13148. https://doi.org/10.1038/s41598-021-91634-0

Ayalew, Y. A., Fante, K. A., & Mohammed, M. A. (2021). Modified U-Net for liver cancer segmentation from computed tomography images with a new class balancing method. *BMC Biomedical Engineering, 3*(1), 4. https://doi.org/10.1186/s42490-021-00050-y

Bilic, P., Christ, P., Li, H. B., Vorontsov, E., Ben-Cohen, A., Kaissis, G., Szeskin, A., Jacobs, C., Mamani, G. E. H., Chartrand, G., Lohöfer, F., Holch, J. W., Sommer, W., Hofmann, F., Hostettler, A., Lev-Cohain, N., Drozdzal, M., Amitai, M. M., Vivanti, R., . . . Menze, B. (2023). The Liver Tumor Segmentation Benchmark (LiTS). *Medical Image Analysis*, *84*, 102680. https://doi.org/https://doi.org/10.1016/j.media.2022.102680

Bioengineering, N. I. o. B. I. a. (2022). *Magnetic Resonance Imaging (MRI)*. https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: Fast and Flexible Image Augmentations. *Information*, *11*(2).

Cancer, W. H. O. I. A. f. R. o. (2022). *Liver cancer statistics*. https://www.wcrf.org/cancer-trends/liver-cancer-statistics/

Chaoqun, W., Binbin, L., & Bin, J. (2021). Fault Diagnosis of Rolling Bearing Based on Convolutional Neural Network of Convolutional Block Attention Module.

Journal of Physics: Conference Series ,, 1732(1), 012045.
https://doi.org/10.1088/1742-6596/1732/1/012045

Chen, B., Zhang, Z., Liu, N., Tan, Y., Liu, X., & Chen, T. (2020). Spatiotemporal
Convolutional Neural Network with Convolutional Block Attention Module for
Micro-Expression Recognition. *Information*, *11*(8), 380.
https://www.mdpi.com/2078-2489/11/8/380

Devidas, T. K., & Sanjay, N. T. (2021). MS-UNet: A multi-scale UNet with feature
recalibration approach for automatic liver and tumor segmentation in CT
images. *Computerized Medical Imaging and Graphics*, *89*, 101885.
https://doi.org/https://doi.org/10.1016/j.compmedimag.2021.101885

Devidas, T. K., & Sanjay, N. T. (2022). HFRU-Net: High-Level Feature Fusion and
Recalibration UNet for Automatic Liver and Tumor Segmentation in CT Images.
Computer Methods and Programs in Biomedicine, *213*, 106501.
https://doi.org/https://doi.org/10.1016/j.cmpb.2021.106501

Devidas, T. K., Shweta, T., & Sanjay, N. T. (2023). LiM-Net: Lightweight multi-level
multiscale network with deep residual learning for automatic liver segmentation
in CT images. *Biomedical Signal Processing and Control*, *80*, 104305.
https://doi.org/https://doi.org/10.1016/j.bspc.2022.104305

Fan, T., Wang, G., Li, Y., & Wang, H. (2020). MA-Net: A Multi-Scale Attention Network for
Liver and Tumor Segmentation. *IEEE Access*, *8*, 179656-179665.
https://doi.org/10.1109/ACCESS.2020.3025372

Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., & Liu, J. (2019).
CE-Net: Context Encoder Network for 2D Medical Image Segmentation. *IEEE
Transactions on Medical Imaging*, *38*(10), 2281-2292.
https://doi.org/10.1109/TMI.2019.2903562

Hamm, C. A., Wang, C. J., Savic, L. J., Ferrante, M., Schobert, I., Schlachter, T., Lin, M.,
Duncan, J. S., Weinreb, J. C., Chapiro, J., & Letzen, B. (2019). Deep learning for
liver tumor diagnosis part I: development of a convolutional neural network
classifier for multi-phasic MRI. *European Radiology*, *29*(7), 3338-3347.
https://doi.org/10.1007/s00330-019-06205-9

Harriet, R., Melina, A., Jacques, F., Olufunmilayo, L., Citadel, J. C., Jérôme, V., Mathieu, L., Katherine, A. M., & Isabelle, S. (2022). Global burden of primary liver cancer in 2020 and predictions to 2040. *Journal of Hepatology, 77*(6), 1598-1606. https://doi.org/https://doi.org/10.1016/j.jhep.2022.08.021

He, K., Zhang, X., Ren, S., & Sun, J. (2016, 2016). Deep Residual Learning for Image Recognition.

HEALTH, S. M. C. S. (2023). *Anatomy and Function of the Liver*. https://www.stanfordchildrens.org/en/topic/default?id=anatomy-and-function-of-the-liver-90-P03069

Hennedige, T., & Venkatesh, S. K. (2013). Imaging of hepatocellular carcinoma: diagnosis, staging and treatment monitoring. *Cancer Imaging, 12*(3), 530-547. https://doi.org/10.1102/1470-7330.2012.0044

Hope, C. o. (2023). *Liver cancer type*. https://www.cancer.org/cancer/types/liver-cancer/about/what-is-liver-cancer.html

Hu, J., Shen, L., & Sun, G. (2018, 2018). Squeeze-and-Excitation Networks.

Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.-W., & Wu, J. (2020). *UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation*

Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y. W., & Wu, J. (2020, 4-8 May 2020). UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),

Huaxiang, L., Youyao, F., Shiqing, Z., Jun, L., Yong, W., Guoyu, W., & Jiangxiong, F. (2023). GCHA-Net: Global context and hybrid attention network for automatic liver segmentation. *Computers in Biology and Medicine, 152*, 106352. https://doi.org/https://doi.org/10.1016/j.compbiomed.2022.106352

Jetley, S. a. L., Nicholas A. and Lee, Namhoon and Torr, Philip H. S. (2018). *Learn To Pay Attention* ICLR 2018, https://arxiv.org/abs/1804.02391

Jha, D., Smedsrud, P. H., Riegler, M. A., Johansen, D., Lange, T. D., Halvorsen, P., & D. Johansen, H. (2019, 2019). ResUNet++: An Advanced Architecture for Medical Image Segmentation.

Jiang, H., Shi, T., Bai, Z., & Huang, L. (2019). AHCNet: An Application of Attention Mechanism and Hybrid Connection for Liver Tumor Segmentation in CT Volumes. *IEEE Access*, *7*, 24898-24909. https://doi.org/10.1109/ACCESS.2019.2899608

Jin, Q., Meng, Z., Sun, C., Cui, H., & Su, R. (2020). RA-UNet: A Hybrid Deep Attention-Aware Network to Extract Liver and Tumor in CT Scans [Original Research]. *Frontiers in Bioengineering and Biotechnology*, *8*. https://doi.org/10.3389/fbioe.2020.605132

Kevin, T., Tomasz, S., & Siamak, M. (2021). SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture. *Pattern Recognition Letters*, *145*, 178-186. https://doi.org/https://doi.org/10.1016/j.patrec.2021.01.036

Kim, J., Min, J. H., Kim, S. K., Shin, S.-Y., & Lee, M. W. (2020). Detection of Hepatocellular Carcinoma in Contrast-Enhanced Magnetic Resonance Imaging Using Deep Learning Classifier: A Multi-Center Retrospective Study. *Scientific Reports*, *10*(1), 9458. https://doi.org/10.1038/s41598-020-65875-4

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM*, *60*(6), 84–90 , numpages = 87. https://doi.org/10.1145/3065386

Lee, H., Park, J., & Hwang, J. Y. (2020). Channel Attention Module With Multiscale Grid Average Pooling for Breast Cancer Segmentation in an Ultrasound Image. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, *67*(7), 1344-1353. https://doi.org/10.1109/TUFFC.2020.2972573

Lei, T., Wang, R., Zhang, Y., Wan, Y., Liu, C., & Nandi, A. K. (2022). DefED-Net: Deformable Encoder-Decoder Network for Liver and Liver Tumor Segmentation. *IEEE Transactions on Radiation and Plasma Medical Sciences*, *6*(1), 68-78. https://doi.org/10.1109/TRPMS.2021.3059780

Lencioni, R., Cioni, D., Crocetti, L., Pina, C. D., & Bartolozzi, C. (2004). Magnetic resonance imaging of liver tumors. *Journal of Hepatology*, *40*(1), 162-171. https://doi.org/https://doi.org/10.1016/S0168-8278(03)00455-0

Li, B., Wu, F., Liu, S., Tang, J., Li, G., Zhong, M., & Guan, X. (2022). CA-Unet++: An improved structure for medical CT scanning based on the Unet++ Architecture.

International Journal of Intelligent Systems, *37*(11), 8814-8832.
https://doi.org/https://doi.org/10.1002/int.22969

Li, C., Tan, Y., Chen, W., Luo, X., Gao, Y., Jia, X., & Wang, Z. (2020). *Attention Unet++: A Nested Attention-Aware U-Net for Liver CT Image Segmentation*

Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.-W., & Heng, P.-A. (2018). H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes. *IEEE Transactions on Medical Imaging*, *37*(12), 2663-2674.
https://doi.org/10.1109/TMI.2018.2845918

Li, Z., Zhang, H., Li, Z., & Ren, Z. (2022). Residual-Attention UNet++: A Nested Residual-Attention U-Net for Medical Image Segmentation. *Applied Sciences*, *12*(14).

Long, J., Shelhamer, E., & Darrell, T. (2015, 2015). Fully convolutional networks for semantic segmentation.

Luan, S., Xue, X., Ding, Y., Wei, W., & Zhu, B. (2021). Adaptive Attention Convolutional Neural Network for Liver Tumor Segmentation [Original Research]. *Frontiers in Oncology*, *11*. https://doi.org/10.3389/fonc.2021.680807

Luc Soler, A. H., Vincent Agnus, Arnaud Charnoz, Jean-Baptiste Fasquel, Johan Moreau, Anne-Blandine, Osswald, Mourad Bouhadjar, Jacques Marescaux. (2010). 3D image reconstruction for comparison of algorithm database: A patient specific anatomical and medical image database. *IRCAD, Strasbourg, France, Tech. Rep.*
https://www.ircad.fr/research/data-sets/liver-segmentation-3d-ircadb-01/

Medicine, J. H. (2023). *Computed Tomography (CT) Scan*.
https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/computed-tomography-ct-scan

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M. J., Heinrich, M. P., Misawa, K., Mori, K., McDonagh, S. G., Hammerla, N. Y., Kainz, B., Glocker, B., & Rueckert, D. (2018). Attention U-Net: Learning Where to Look for the Pancreas. *ArXiv*, *abs/1804.03999*.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In (pp. 234-241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

Sherif, R. Z. A.-M., & Mark, B. (2010). Liver Anatomy. *Surgical Clinics of North America*, *90*(4), 643-653. https://doi.org/https://doi.org/10.1016/j.suc.2010.04.017

Shuchao, P., Anan, D., Mehmet, A. O., Yunyun, W., & Zhenmei, Y. (2021). Tumor attention networks: Better feature selection, better tumor segmentation. *Neural Networks*, *140*, 203-222. https://doi.org/https://doi.org/10.1016/j.neunet.2021.03.006

Siegel, R. L., Miller, K. D., & Jemal, A. (2019). Cancer statistics, 2019. *CA: A Cancer Journal for Clinicians*, *69*(1), 7-34. https://doi.org/https://doi.org/10.3322/caac.21551

Siegel, R. L., Miller, K. D., & Jemal, A. (2020). Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, *70*(1), 7-30. https://doi.org/https://doi.org/10.3322/caac.21590

Society, A. C. (2019). *What Is Liver Cancer?* https://www.cancer.org/cancer/types/liver-cancer/about/what-is-liver-cancer.html

Thomas, E., Pawan, S. J., Kumar, S., Horo, A., Niyas, S., Vinayagamani, S., Kesavadas, C., & Rajan, J. (2021). Multi-Res-Attention UNet: A CNN Model for the Segmentation of Focal Cortical Dysplasia Lesions from Magnetic Resonance Images. *IEEE Journal of Biomedical and Health Informatics*, *25*(5), 1724-1734. https://doi.org/10.1109/JBHI.2020.3024188

Today, M. N. (2022). *What is a liver ultrasound?* https://www.medicalnewstoday.com/articles/liver-ultrasound

Wang, J., Zhang, X., Lv, P., Wang, H., & Cheng, Y. (2022). Automatic Liver Segmentation Using EfficientNet and Attention-Based Residual U-Net in CT. *Journal of Digital Imaging*, *35*(6), 1479-1493. https://doi.org/10.1007/s10278-022-00668-x

Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). CBAM: Convolutional Block Attention Module. In (pp. 3-19). Springer International Publishing. https://doi.org/10.1007/978-3-030-01234-2_1

Xiwang, X., Xipeng, P., Feng, S., Weidong, Z., & Jubai, A. (2022). MCI-Net: Multi-scale context integrated network for liver CT image segmentation. *Computers and Electrical Engineering*, *101*, 108085. https://doi.org/https://doi.org/10.1016/j.compeleceng.2022.108085

Yeghiazaryan, V., & Voiculescu, I. (2018). Family of boundary overlap metrics for the evaluation of medical image segmentation. *Journal of Medical Imaging, 5*(01), 1. https://doi.org/10.1117/1.jmi.5.1.015006

Zhang, Z., Liu, Q., & Wang, Y. (2018). Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters, 15*(5), 749-753. https://doi.org/10.1109/LGRS.2018.2802944

Zhao, P., Zhang, J., Fang, W., & Deng, S. (2020). SCAU-Net: Spatial-Channel Attention U-Net for Gland Segmentation [Original Research]. *Frontiers in Bioengineering and Biotechnology, 8.* https://doi.org/10.3389/fbioe.2020.00670

Zhen, S.-h., Cheng, M., Tao, Y.-b., Wang, Y.-f., Juengpanich, S., Jiang, Z.-y., Jiang, Y.-k., Yan, Y.-y., Lu, W., Lue, J.-m., Qian, J.-h., Wu, Z.-y., Sun, J.-h., Lin, H., & Cai, X.-j. (2020). Deep Learning for Accurate Diagnosis of Liver Tumor Based on Magnetic Resonance Imaging and Clinical Data [Original Research]. *Frontiers in Oncology, 10.* https://doi.org/10.3389/fonc.2020.00680

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In (pp. 3-11). Springer International Publishing. https://doi.org/10.1007/978-3-030-00889-5_1

Zisserman, K. S. a. A. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition* ICLR-2015,

VITA

NAME	Kasun Gayashan Hettihewa
DATE OF BIRTH	12 April 1994
PLACE OF BIRTH	Matara, Sri Lanka
INSTITUTIONS ATTENDED	Asian Institute of Technology (AIT)
HOME ADDRESS	361/1T, Kumarathunga Road, Matara, Sri Lanka