People`s Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research
Echahid Cheikh Larbi Tebessi University –Tebessa-
Exact Sciences Natural and Life Sciences Faculty
Department of Computer Science

MASTER THESIS

Domain: Mathematics and Computer Sciences
Field: Computer Science
Option: Systems and Multimedia

AI-Based Universal Lesion Segmentation Application for Thoracic-Abdominal in Computed Tomography Scans.

Defended by:
MOHAMED LADJAL

Supervised by:
Dr. BOUALLEG Yaakoub

In front of the jury composed of:

Khediri Abderrazak	MCB	Echahid Cheikh Larbi Tebessi University	President
Khediri Samir	MAA	Echahid Cheikh Larbi Tebessi University	Examiner

Defended on ……

Note: ……. Mention:……..……

To my family and friends, whose constant support and belief in me have made this journey possible. To my mentors and colleagues, for their insightful guidance and encouragement. This thesis is a reflection of your trust and inspiration.

Acknowledgement

All praise and thanks are due to Allah alone, who guided me throughout this journey and blessed me with the opportunity to complete this thesis. I express my deepest gratitude to my supervisor, **Dr. Yaakoub Boualleg**, for his invaluable guidance, support, and encouragement. His insights and expertise have been instrumental in the completion of this thesis.

I am also grateful to the members of my thesis evaluation committee, **Dr. Khediri Abderrazak** and **Dr. Khediri Samir**, for their time, effort, and valuable feedback. Their constructive critiques and suggestions have greatly enhanced the quality of my work.

Furthermore, I extend my heartfelt thanks to **my family and friends** for their unwavering support and understanding during this journey. Their encouragement has been a constant source of motivation.

Lastly, I acknowledge with deep appreciation all those who have contributed, directly or indirectly, to the successful completion of this thesis. Your assistance and encouragement have been invaluable.

Abstract

Medical imaging plays a crucial role in the diagnosis and treatment of various diseases, particularly in identifying and evaluating lesions within the thoracic-abdominal region. Traditional methods of lesion segmentation, which rely heavily on manual delineation by radiologists, are time-consuming and prone to variability. Advances in deep learning and artificial intelligence offer promising solutions to these challenges by automating the segmentation process, thus enhancing accuracy and efficiency.

This master thesis presents an innovative framework for universal lesion segmentation in thoracic-abdominal computed tomography (CT) scans using advanced deep learning techniques. The proposed methodology leverages the U-Mamba model, which integrates Convolutional Neural Networks (CNNs) and State Space Models (SSMs), to enhance segmentation accuracy and efficiency.

The framework involves meticulous preprocessing steps, including lesion selection and Volume of Interest (VOI) cropping, to ensure precise targeting of lesions. The U-Mamba model, trained with the LION (EvoLved Sign Momentum) optimizer, demonstrates superior performance in capturing long-range dependencies and handling complex lesion structures across various datasets. The model's encoder-decoder architecture, coupled with residual blocks and skip connections, enables robust and detailed segmentation outputs. This study validates the efficacy of the proposed framework through comprehensive evaluations, underscoring its potential to improve diagnostic imaging and clinical workflows.

Keywords: Automated Universal Lesion Segmentation, Thoracic-Abdominal Lesions, Medical Imaging, Deep Learning, Convolutional Neural Networks, U-Mamba Network, LION Optimizer.

Table des matières

List of Figures

List of Tables

General Introduction

Medical imaging plays a crucial role in the diagnosis and treatment of various diseases. With the advanced imaging technologies, there has been a significant increase in the volume of medical scans generated worldwide. Accurate and efficient analysis of these scans is essential for effective clinical decision-making. Traditional methods of medical image analysis, which heavily rely on manual annotations by radiologists, are becoming increasingly impractical due to the volume of data and the inherent limitations of human capabilities.

Deep learning has emerged as a powerful tool in medical image analysis, offering the potential to automate and enhance the accuracy of lesion segmentation. Lesion segmentation is a critical task in medical imaging, as it helps in identifying and delineating pathological regions within the body. Accurate segmentation is particularly important for the diagnosis and treatment planning of various diseases, including cancer.

This thesis focuses on designing and evaluating new framework for universal lesion segmentation across different body regions. By leveraging state-of-the-art architectures such as the U-Mamba model and integrating innovative optimization strategies like the LION optimizer, this work aims to improve the performance of lesion segmentation models. The comprehensive evaluation of these models on diverse datasets demonstrates their robustness and potential for clinical application, ultimately aiming to enhance diagnostic accuracy and patient outcomes.

In summary, the primary contributions of this thesis include:

1. The development of a universal lesion segmentation model capable of segmenting several types of lesions in the abdominal-thoracic area.
2. The integration of advanced optimization techniques to enhance model performance.
3. A comprehensive evaluation framework to assess the effectiveness of the proposed models across multiple datasets.
4. Insights into the practical application of these models in clinical settings, highlighting their potential impact on diagnostic workflows and patient care.

The subsequent chapters of this thesis are organized as follows:

— Chapter 1 provides the medical background and introduces the different types of

lesions in the thoracic-abdominal area. It also discusses the clinical workflow in lesion segmentation and the limitations of traditional methods.

— Chapter 2 delves into the fundamental concepts of machine learning and segmentation techniques, focusing on deep learning approaches.

— Chapter 3 presents the state of the art in medical image segmentation, reviewing recent advancements and identifying gaps in the current research.

— Chapter 4 outlines the contributions and results of this thesis, detailing the proposed framework, experimental studies, and comparative analysis of the developed models.

— The final chapter concludes the thesis, summarizing the key findings and suggesting directions for future research.

Through this comprehensive investigation, the thesis aims to contribute significantly to the field of medical image analysis, advancing the development of automated and accurate lesion segmentation methods.

Chapter 1

Basic Concepts

1.1 Medical Background Concepts

In recent years, there has been a notable surge in the number of computed tomography (CT) examinations conducted annually [1]. This trend has contributed to increased workloads for radiologists [2], particularly in oncological radiology. The projected rise of 47% in the global cancer burden by 2040 compared to 2020 [3] indicates that oncological radiology will play a pivotal role in addressing this escalating demand. Notably, patients with cancer often undergo multiple imaging examinations over extended periods to monitor disease progression and treatment response.

Quantifying disease progression and treatment response in longitudinal CT scans typically involves manual measurements of lesions. These measurements, often interpreted using the Response Evaluation Criteria In Solid Tumors (RECIST) guidelines [4], serve to standardize and expedite the assessment process. However, RECIST guidelines limit the number of lesions to be measured to a maximum of five "target lesions" across multiple organs or structures, thereby constraining the overall assessment.

To alleviate the time-intensive nature of lesion annotation in oncological scans, automatic segmentation models have emerged as a promising solution. These models can extract lesion information with minimal guidance from radiologists, such as a single-click within the lesion or utilizing bounding box predictions from detection models. Moreover, segmenting lesion volumes in 3D offers additional insights that can facilitate the calculation of more informative lesion characteristics. Additionally, registration algorithms enable the propagation of segmented lesions, leading to significant time savings during follow-up examinations [5].

While significant progress has been made in AI-based automatic segmentation models for tumors, challenges persist in developing efficient and robust models able to accurately segment various lesion types in the thorax-abdomen area. A Universal Lesion Segmentation (ULS) model is sought after to address this diversity of lesion types effectively.

However, previous research on ULS [6] [7] [8], has primarily relied on single partially annotated datasets [9], lacking comprehensive 3D segmentation masks for evaluation. Hence, there is a pressing need for well-curated and varied datasets to drive further advancements in this critical area of medical imaging.

Workforce Shortage: The clinical radiology workforce often struggles to keep pace with the growing volume of imaging studies. This shortage increases the radiologists workload, leading to fatigue, burnout, and compromised patient care.

Rising Imaging Volumes: The demand for medical imaging is on the rise, driven by factors such as an ageing population, advances in technology, and the introduction of new screening programs. Coping with the ever-increasing number of imaging studies poses significant challenges for radiologists.

Time-Intensive Reporting: The process of interpreting and reporting medical images is time-consuming and requires meticulous attention to detail. Radiologists must review clinical histories, previous imaging studies, and current scans to generate accurate reports, adding to their workload.

Non-Reporting Tasks: Contrary to common perception, radiologists engage in various non-reporting tasks, such as attending multidisciplinary team meetings, communicating findings to patients and referring doctors, performing interventional procedures, and supervising trainees. Balancing these responsibilities alongside reporting duties can be demanding.

Technology Integration and Adaptation: Rapid advancements in imaging technology necessitate continuous learning and adaptation by radiologists. Staying abreast of new imaging modalities, software updates, and AI tools requires ongoing training and professional development.

Quality and Safety Concerns: Ensuring the accuracy and safety of medical imaging practices is paramount. Radiologists must adhere to stringent quality assurance protocols, minimize radiation exposure risks, and mitigate errors to uphold patient safety and quality of care.

Navigating these challenges requires collaboration, innovation, and strategic planning within radiology departments and healthcare systems to support radiologists in delivering timely and high-quality diagnostic services.

1.2 Types of Lesions in the Thoracic-Abdominal Area

Lesions in the thoracic-abdominal area encompass a wide spectrum of abnormalities, ranging from benign to malignant. Radiologists frequently encounter the following types of lesions during diagnostic imaging assessments:

Solid Masses: Solid masses can be either benign or malignant and originate from organs such as the liver, kidneys, adrenal glands, pancreas, spleen, or solid components

of the lungs. Benign solid masses include adenomas, hemangiomas, and focal nodular hyperplasia, while malignant masses may represent primary or metastatic neoplasms.

Cysts: Cysts are fluid-filled sacs that can be either benign or potentially concerning, depending on their characteristics and clinical context. Benign cysts, such as simple renal cysts or ovarian cysts, are common incidental findings. However, complex cystic lesions or those associated with solid components may raise suspicion for malignancy.

Calcifications: Calcifications can occur within various organs and vascular structures and may be indicative of benign or malignant processes. Benign calcifications, such as those seen in renal calyceal calculi or vascular calcifications, are typically asymptomatic. Conversely, calcifications within solid organs, such as the liver or pancreas, may raise concern for malignancy, especially when associated with other worrisome features.

Fluid Collections: Fluid collections, including ascites, pleural effusions, pericardial effusions, and abscesses, can result from a variety of etiologies, including inflammation, infection, trauma, or malignancy. Characterization of fluid collections is essential for guiding appropriate management, as malignant effusions may require further investigation and treatment.

Vascular Lesions: Vascular lesions, such as aneurysms, arteriovenous malformations, and thrombosis, can have significant clinical implications. While some vascular lesions are benign and incidental findings, others may pose risks of rupture, thrombosis, or embolization, necessitating careful evaluation and management.

Inflammatory Lesions: Inflammatory lesions encompass a broad spectrum of conditions, including infections, granulomatous diseases, autoimmune conditions, and radiation-induced changes. While many inflammatory lesions are benign and self-limiting, others may mimic malignant processes or contribute to chronic morbidity if left untreated.

Understanding the diverse spectrum of lesions encountered in the thoracic-abdominal area is essential for radiologists to provide accurate diagnoses and optimize patient care. Differential diagnosis considerations, clinical correlation, and multidisciplinary collaboration are integral components of lesion evaluation and management in this anatomical region.

1.3 Clinical Workflow in Lesion Segmentation and Limitations

1.3.1 Traditional Methods Used by Radiologists

Radiologists employ a variety of traditional methods to measure and evaluate lesions from imaging modalities such as CT or magnetic resonance imaging (MRI). These methods include:

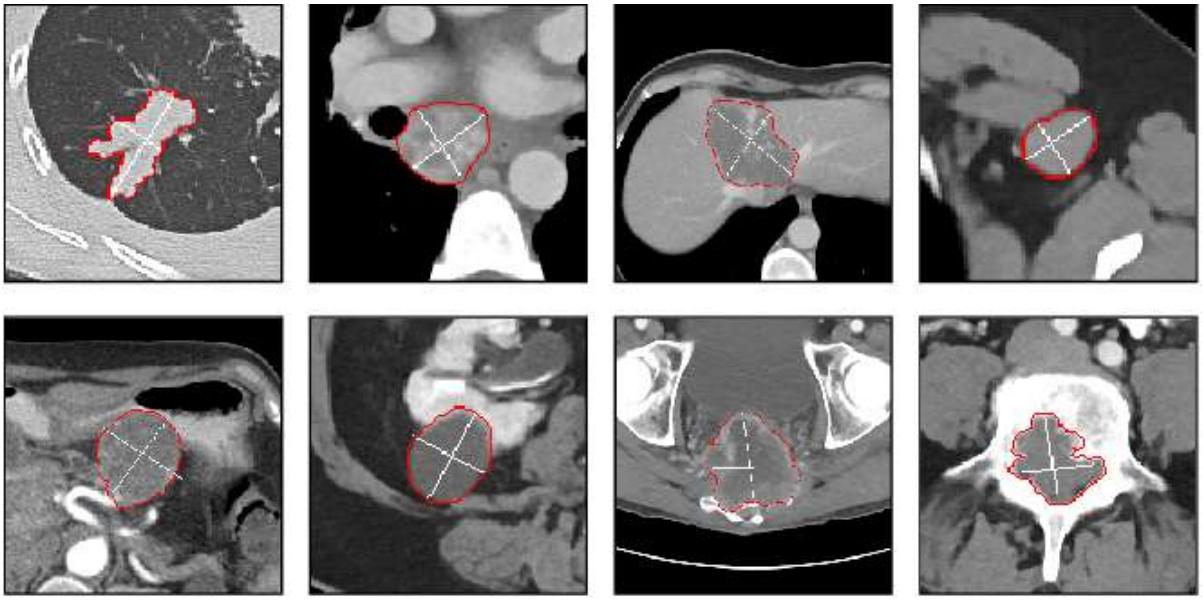

Figure 1.1: Examples from The DeepLesion dataset with 8 lesions. From left to right and top to bottom, lung, mediastinum, liver, soft-tissue, abdomen, kidney, pelvis, and bone, respectively. The RECIST slices are shown with manually delineated boundaries in red and bookmarked RECIST diameters in white.

Manual Measurements: Radiologists manually measure lesion dimensions using built-in measurement tools in imaging software or calipers on printed films. This method allows for precise measurement of lesion size, aiding in treatment planning and monitoring.

Region of Interest (ROI) Analysis: Radiologists define specific regions within the lesion or surrounding tissues to analyze pixel intensities, textures, or other quantitative features. ROI analysis provides insights into lesion characteristics and helps in differentiation from normal tissues.

Volume Measurements: Software tools calculate lesion volumes based on segmented regions on imaging studies. This information is valuable for monitoring lesion growth over time and assessing treatment response.

Standardized Reporting Systems: Radiologists utilize standardized reporting systems such as Response Evaluation Criteria in Solid Tumors (RECIST) or World Health Organization (WHO) criteria for lesion measurement and response assessment, particularly in oncology imaging. These systems ensure consistency in reporting and facilitate comparison of results across different studies. Figure 1.1 illustrates different lesion categories within RECIST slices that are manually bookmarked RECIST diameters.

Computer-Aided Detection (CAD): CAD systems assist radiologists in lesion detection and measurement by highlighting suspicious areas and providing automated measurements. CAD enhances efficiency and accuracy in lesion evaluation, especially in complex cases.

These traditional methods play a crucial role in accurately measuring and evaluating lesions from imaging studies, contributing to precise diagnosis, treatment planning, and monitoring of disease progression.

1.3.2 Limitations of Traditional Methods

The traditional methods employed in radiology practice for measuring and evaluating lesions from imaging studies have certain limitations, including:

Subjectivity: Manual measurements and ROI analysis can be subjective and prone to inter-observer variability, leading to inconsistencies in lesion measurements.

Inaccuracy: Manual measurements may be inaccurate due to variations in measurement techniques, slice thickness, and image quality, potentially affecting the reliability of lesion size assessments.

Time-Consuming: Manual measurements and ROI analysis can be time-consuming, especially in cases with multiple lesions or complex anatomical structures, impacting workflow efficiency.

Limited Reproducibility: Volume measurements may lack reproducibility if the segmentation of lesions is challenging or if there are discrepancies in defining lesion boundaries, affecting the reliability of volume calculations.

Dependency on Operator Skills: The accuracy of measurements and analyses heavily relies on the operator's skills and experience, leading to variability in results based on individual expertise.

Limited Standardization: While standardized reporting systems like RECIST provide guidelines for lesion measurement, there may still be variations in interpretation and implementation across different institutions or radiologists, affecting consistency in reporting.

CAD System Limitations: Computer-Aided Detection (CAD) systems may have limitations in accurately detecting and measuring lesions, especially in cases of atypical or complex lesions that may not conform to standard detection algorithms.

Addressing these limitations through advancements in technology, automation, standardization of protocols, and ongoing training and education can help improve the accuracy and reliability of lesion measurements and evaluations in radiology practice.

1.4 Significance of Automated Segmentation

1.4.1 Importance of Automated Models

Based on all of these limitations associated with traditional methods for lesion measurement and evaluation in radiology practice, the importance of automated models be-

comes evident. Automated models offer several advantages over traditional methods, including increased objectivity, efficiency, and accuracy in lesion assessment.

By leveraging advanced algorithms and machine learning techniques, automated models can mitigate the subjectivity and interobserver variability inherent in manual measurements and ROI analysis. Moreover, automated models have the potential to streamline the lesion evaluation process, reducing the time required for analysis and reporting. Additionally, these models can provide more comprehensive quantitative information about lesions, beyond simple size measurements, enabling a deeper understanding of lesion characteristics and behavior.

Furthermore, automated models have the capability to enhance lesion detection sensitivity and specificity, particularly in cases with subtle or complex lesions, thus improving diagnostic accuracy. Overall, the adoption of automated models in radiology practice represents a significant opportunity to overcome the limitations of traditional methods and enhance the efficiency and effectiveness of lesion measurement and evaluation.

1.5 Diagnostic Imaging Technologies and Data Representation

In this section, we will delve into diagnostic imaging technologies and how their data is represented. We'll start by providing an overview of various imaging modalities, including X-ray radiography, computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, and nuclear medicine imaging. Then, we'll discuss the image acquisition process in CT scans and explore the different data formats commonly used in CT imaging. This groundwork will lay the foundation for a deeper understanding of the intricacies of diagnostic imaging technology.

1.5.1 Medical Imaging Modalities

Imaging modalities play a pivotal role in diagnostic medicine, enabling the visualization and characterization of anatomical structures, physiological processes, and pathological findings. Key imaging modalities utilized in clinical practice can be summarized as follow:

X-Ray Radiography: utilizes ionizing radiation to produce two-dimensional images of internal structures, making it suitable for detecting skeletal abnormalities, pulmonary pathologies, and foreign bodies. It is widely employed for initial screening and evaluation of trauma and orthopedic conditions.

Computed Tomography (CT): involves the acquisition of multiple cross-sectional images through the use of X-rays and advanced computer processing techniques. This modality provides detailed anatomical information with high spatial resolution, making

it valuable for the evaluation of soft tissue structures, organ morphology, and vascular abnormalities. CT imaging is particularly adept at detecting and characterizing lesions in the thoracic-abdominal area, making it the focus of our thesis and the primary modality on which our segmentation model was trained.

Magnetic Resonance Imaging (MRI): utilizes strong magnetic fields and radio-frequency pulses to generate detailed cross-sectional images of the body. Unlike CT, MRI does not involve ionizing radiation, making it suitable for imaging soft tissues and organs without radiation exposure. MRI is widely utilized for neurological, musculoskeletal, and abdominal imaging, offering excellent soft tissue contrast and functional information.

Ultrasound Imaging: employs high-frequency sound waves to produce real-time images of internal structures. It is non-invasive, radiation-free, and particularly useful for imaging the abdomen, pelvis, and obstetric applications. Ultrasound is valuable for guiding intervention procedures and assessing vascular flow dynamics.

Nuclear Medicine Imaging: involves the administration of radio-pharmaceuticals that emit gamma rays, which are detected by specialized cameras to generate functional images. Techniques such as single-photon emission computed tomography (SPECT) and positron emission tomography (PET) provide insights into metabolic processes, organ function, and disease localization.

Each imaging modality offers unique strengths and limitations, enabling clinicians to select the most appropriate modality based on clinical indications, patient characteristics, and diagnostic objectives. In the context of our thesis, we focus on CT imaging due to its significance in lesion evaluation.

1.5.2 Image Acquisition in CT Scans

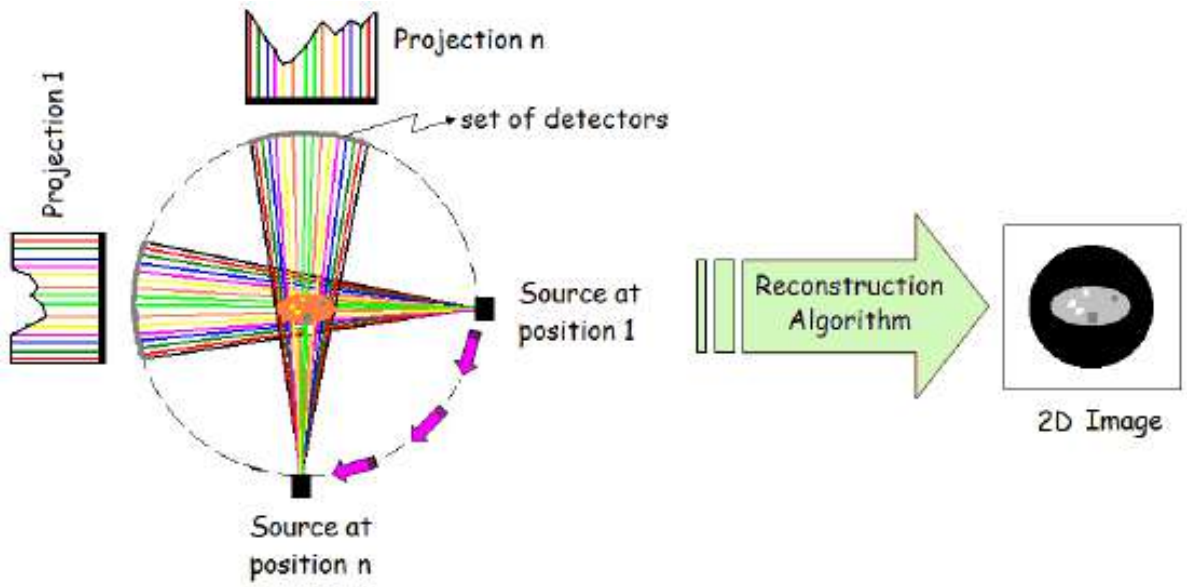

Figure 1.2: Computed Tomography Scan Acquisition Process.

As illustrated in Figure 1.2, the acquisition process in CT scans involves an X-ray source and a set of detectors positioned opposite each other. The source emits X-rays that pass through the patient's body and are detected on the other side. The source rotates around the patient, taking multiple X-ray images (projections) from different angles. Each position generates a projection that captures a cross-sectional view of the body.

During this process, the X-ray source moves in a circular (or helical) trajectory around the patient. This motion allows the acquisition of numerous projections from various angles, which collectively form a comprehensive dataset. The detectors measure the X-ray intensity after it passes through the body, capturing information about the tissue's attenuation properties.

The collected data from the detectors consist of a series of projections, each corresponding to a specific angle of the X-ray source. These projections reflect the average linear attenuation coefficient, which indicates how much the X-rays are reduced in intensity by the tissues they traverse.

To transform these projections into meaningful images, a reconstruction algorithm is employed. This algorithm processes the raw projection data and reconstructs detailed 2D images of the internal structures. The resulting images provide a cross-sectional view of the patient's anatomy, allowing radiologists to analyze the body's internal features comprehensively.

By stacking these 2D cross-sectional images, or slices, in the order they were acquired, a three-dimensional (3D) representation of the patient's anatomy can be obtained. This 3D reconstruction enables more comprehensive visualization and analysis, allowing for enhanced diagnostic capabilities and improved surgical planning.

This advanced imaging technique is essential for diagnosing and monitoring various medical conditions, offering high-resolution insights into the body's internal structures.

1.5.3 Data Formats in CT Scans

CT scans generate volumetric datasets containing detailed anatomical information in digital form. These datasets represent the spatial distribution of X-ray attenuation coefficients within the imaged anatomy. Depending on the acquisition protocol, CT scan data is typically represented in three-dimensional (3D) formats, capturing spatial information in three dimensions (length, width, and depth).

Digital Imaging and Communications in Medicine (DICOM): is the standard format for storing and transmitting medical images, including CT scans. DICOM files encapsulate volumetric data as a series of two-dimensional slices arranged along the axial, sagittal, or coronal planes. This facilitates 3D reconstruction and visualization of anatomical structures, enabling clinicians to navigate through the dataset in multiple

dimensions.

Neuro-imaging Informatics Technology Initiative (NIfTI): NIfTI is a standardized format commonly used for neuroimaging data, including CT scans of the brain. NIfTI files encode volumetric data in a standardized manner, facilitating interoperability and compatibility with neuro-imaging software packages. While primarily designed for static 3D imaging, NIfTI format extensions allow for the representation of dynamic data, such as time-resolved CT angiography studies.

1.6 Conclusion

In summary, a marked rise in CT exams especially within oncological radiology has led to increased work load for the Radiologist. Pairing this trend with the forecast increase in cancer cases worldwide, it reaffirms the critical demand for cost-effective imaging solutions. Despite contributing to subjective results that are somewhat time-consuming and challenging for any one observer to reproduce, traditional lesion measurement assessment methods, such as manual annotation and the RECIST guidelines.

Thus, automated segmentation models provide an interesting solution to these problems. In this way, missegmentation problems will be eliminated or minimized and the reproducibility issue related to inter-observer variability will also be settled while considerable time savings can be achieved with these models. Although improved, well-globally annotated benchmark datasets are still necessary to make progress towards universal whole lesion segmentation models on thoracic-abdominal data.

The next chapter will delve into machine learning concepts and segmentation techniques, exploring the theoretical foundations and practical applications necessary for developing effective automated segmentation models .

Chapter 2

Machine Learning Concepts and Segmentation Techniques

2.1 Introduction to Machine Learning

2.1.1 Introduction

Artificial intelligence (AI) finds its roots in the mid-20th century, where early visionaries laid the groundwork for intelligent machines. Machine learning, a subset of AI, emerged as a critical approach, empowering computers to learn from data and make predictions autonomously, marking a significant leap forward in technological advancement.

The history of **deep learning,** as illustrated in Figure 2.1, showcases the evolution of AI from its early days with the development of the perceptron in the 1950s to the breakthroughs in the 2010s. This timeline highlights significant milestones such as the

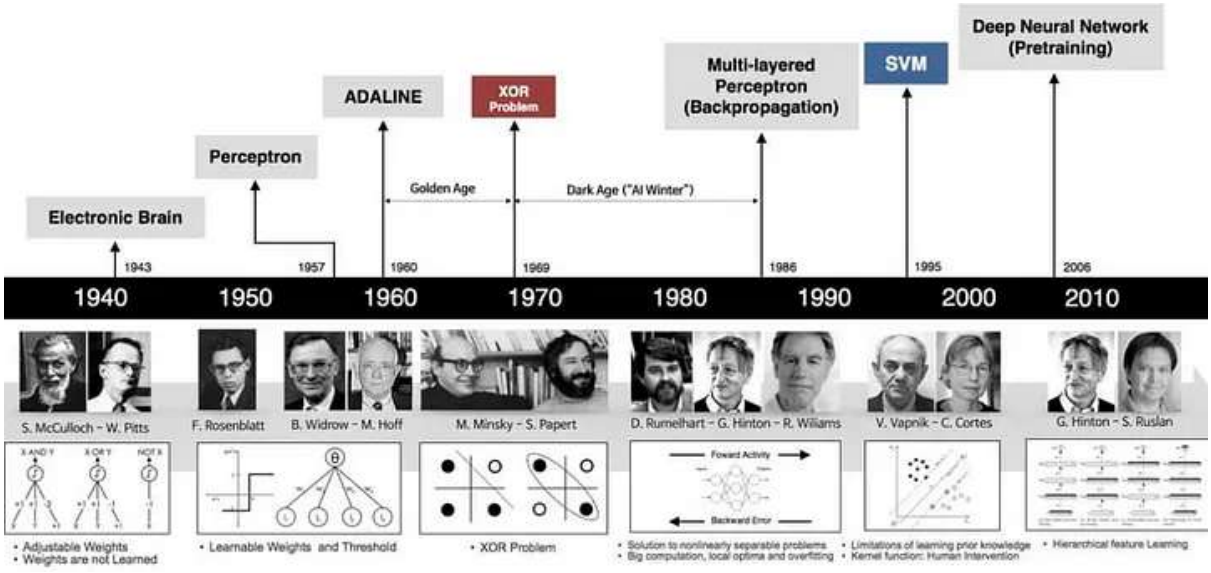

Figure 2.1: The history of deep learning.

introduction of the perceptron, ADALINE, the multi-layered perceptron with backpropagation, and the resurgence of deep neural networks.

Deep learning experienced a significant breakthrough in the mid-2010s, particularly around 2012, when deep neural networks demonstrated remarkable performance in various tasks, surpassing traditional machine learning methods by a large margin. One of the key milestones was the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012, where a deep learning model known as AlexNet, developed by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton [10], achieved a dramatic reduction in error rates for image classification tasks. AlexNet's success marked the beginning of the deep learning revolution, showcasing the power of convolutional neural networks (CNNs) in computer vision tasks.

In 2024, we're seeing the fruits of decades of AI research come to fruition. It's incredible how AI has evolved from simple automation to something that can create artwork that rivals a human artist's and even produce documentaries like a seasoned professional. And when it comes to healthcare, AI's impact has been immense. It's not just about reshaping industries anymore; it's about saving lives through more accurate and efficient medical diagnostics. This is truly a game-changer in human-machine interaction.

As we stand on the brink of a new era driven by intelligent machines, the potential for AI to augment human capabilities and solve complex challenges is boundless. With continued innovation and collaboration, the future holds even greater promise for realizing the full potential of artificial intelligence.

2.1.2 Artificial Neural Networks

Artificial Neural Networks (ANNs), have gained widespread recognition for their versatile applications across various fields, including computer vision, natural language processing, speech recognition, financial forecasting, bioinformatics, and robotics. These networks simulate the behavior of the human brain's neurons and enable computers to recognize patterns and address complex tasks.

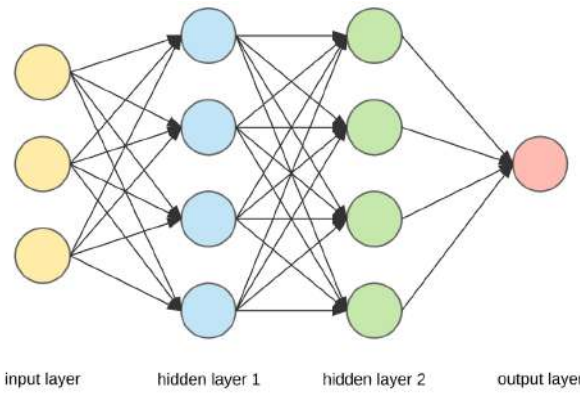

Figure 2.2: Structure of an artificial neural network.

The structure of the ANN models consists of three main layers (see Figure 2.2); the input layer, one or more hidden layers, and the output layer. The input layer receives the initial data, the hidden layers apply mathematical functions to transform the data through weighted connections, and the output layer produces the final prediction or classification. Hidden layers are crucial as they enable the network to learn complex patterns and representations.

The base unit of the ANN models is the artificial perceptron, developed in 1969 by Marvin Minsky and Seymour Papert [11]. As shown in figure 2.3, this model introduced numerical weights and a learning mechanism to facilitating scientific evaluation of different tasks.

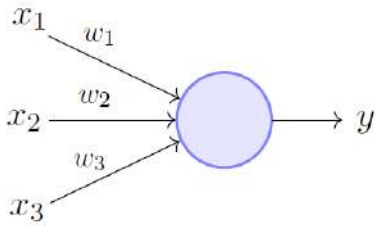

Figure 2.3: Illustration of an artificial perceptron.

The Perceptron operates as a binary classification algorithm in four steps: initialization, function calculation, error assessment, and weight adjustment. The learning process involves optimizing weights to minimize the error between actual and predicted values, typically using stochastic gradient descent and backpropagation algorithms.

Activation functions are crucial in determining ANN accuracy and performance. Non-linear activation functions, such as ReLU, Tanh, and Sigmoid, are preferred over linear functions due to their effectiveness in complex situations with multivariable inputs.

2.1.3 Deep Learning

Deep Learning (DL), is a subset of machine learning, that is deeply intertwined with the principles and structures of ANNs. In the past decade, DL has garnered significant attention for its ability to tackle complex problems across various domains, owing much of its success to the foundational concepts established by ANNs.

Similar to ANNs, Deep Neural Networks (DNNs) consist multiple layers of artificial neurons that receive inputs, perform computations, and generate outputs. However, unlike traditional ANNs, which typically have one or two hidden layers, DNNs comprise multiple hidden layers, enabling them to learn hierarchical representations of data.

The depth of DNNs allows them to automatically extract intricate features from raw data, without the need for manual feature engineering. This capability is particularly

beneficial in tasks involving high-dimensional data, such as image and speech recognition, where traditional machine learning approaches may struggle.

In Figure 2.4, the process of in-model feature generation (feature extraction) in a DL actitecture is illustrated by analyzing an input image. Initially, the model identifies edges, crucial for outlining structural boundaries. Subsequently, it combines these edges to form complex features, capturing spatial relationships. Next, it detects facial components like eyes and nose, vital for person identification. Finally, after successive iterations, the model produces an output identifying who is the person, showcasing deep learning's prowess in deciphering complex visual data.

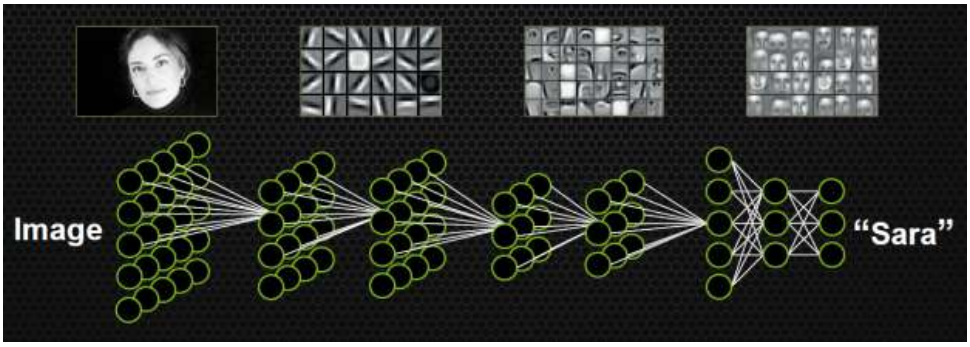

Figure 2.4: Illustration of deep learning feature extraction process, showcasing the model's ability to identify edges and facial features.

2.1.4 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a type of DNNs that are commonly used for analyzing visual imagery. CNNs are specifically designed to recognize patterns and features in images through the application of convolutions, which are mathematical operations that scan across the image to extract relevant information.

One of the key advantages of CNNs is their ability to automatically learn hierarchical representations of visual data, capturing increasingly complex features as information flows through the network. During training, CNNs learn to recognize discriminative features from labeled images and optimize their parameters (weights and biases) through techniques like backpropagation and gradient descent.This hierarchical feature learning enables CNNs to achieve superior performance in tasks like object recognition, image classification and image segmentation.

Building blocks of CNNs:

Convolutional Neural Networks (CNNs) are composed of several key layers that work together to analyze and interpret visual data (see Figure 2.5). These layers include

convolutional layers, pooling layers, and fully connected layers. Each layer type plays a distinct role in the network's ability to learn and recognize patterns.

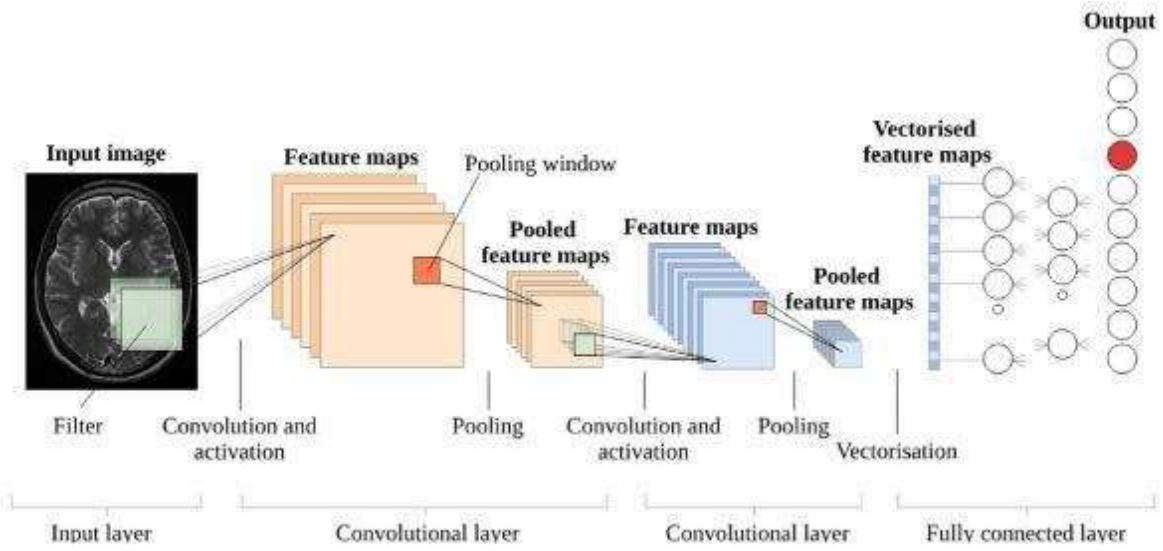

Figure 2.5: Architecture of Convolutional Neural Networks.

Convolutional Layers

A convolutional layer is designed to detect features such as edges, textures, and patterns within the input image. The leftmost grid in Figure 2.6 represents the input image, a 5x5 matrix of pixel values. The blue square outlines a 3x3 section (patch) of the input image processed by the convolutional layer. This patch is referred to as the local receptive field. The middle grid represents the kernel (or filter), a 3x3 matrix with values typically learned during training. The convolution operation involves element-wise multiplication of the image patch and the kernel, followed by summing the results.

The resulting sum is placed in the corresponding position in the output matrix. This process repeats as the kernel slides over the entire input image, producing the full output matrix.

Pooling Layers

Following the convolutional layers, pooling layers are applied to reduce the spatial dimensions of the feature maps while retaining the most significant information. Pooling layers achieve this by summarizing the presence of features in sub-regions of the feature map, as shown in Figure 2.7. The 2×2 max-pooling operation. The input feature map (left grid) is divided into non-overlapping 2×2 regions. For each region, the maximum value is selected.

Pooling layers serve several purposes in CNNs:

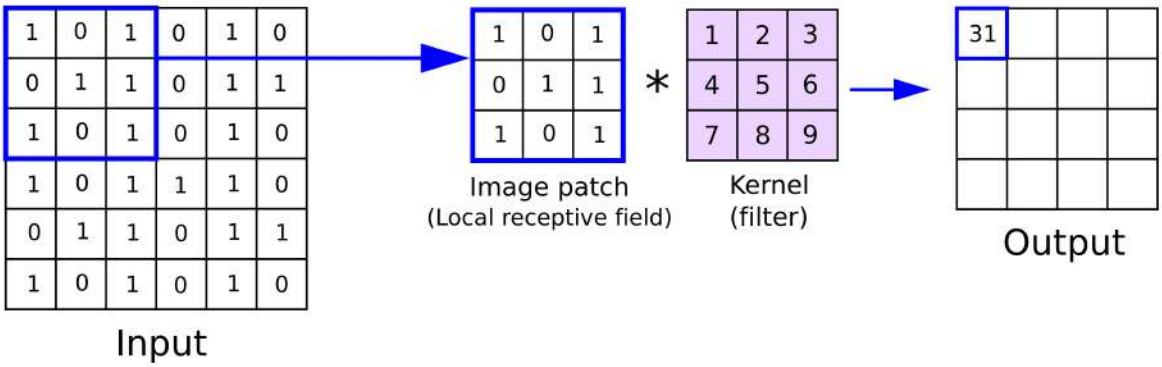

Figure 2.6: Convolution operation in a CNN.

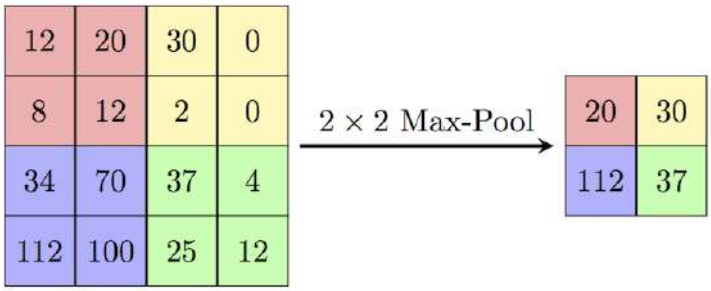

Figure 2.7: Max-pooling operation in a CNN.

— **Dimensionality Reduction**: Reduces the spatial dimensions of feature maps, decreasing computational load and the number of parameters, which helps to mitigate overfitting.
— **Translation Invariance**: Helps the network to recognize features regardless of their position, enhancing robustness to small translations.
— **Noise Reduction**: Filters out noise and retains the most salient features.

Fully Connected Layers

After the convolutional and pooling layers have extracted the features of the input image, the fully connected layer (FCL) is used to interpret these features and perform the final prediction. In a fully connected layer, every neuron from the previous layer (typically a flattened version of the feature maps) is connected to every neuron in the current layer, as illustrated in Figure 2.8.

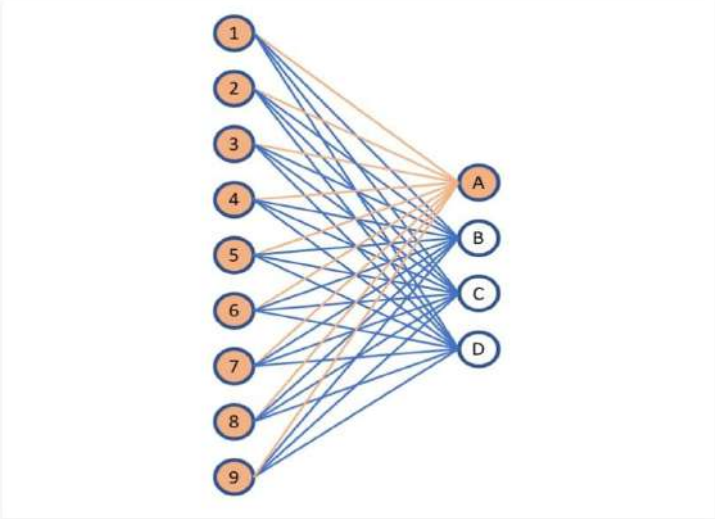

Figure 2.8: Fully connected layer in a CNN.

The left side represents the input layer, each corresponding to a feature. The right side represents the output layer N neurons which could represent different classes in a classification task. Each connection between neurons has an associated weight that is learned during training. The FCL integrates all the extracted features to make the final prediction.

The purpose of the FCL is to combine the features learned by the convolutional and pooling layers to predict the correct output. This layer is essentially a traditional neural network layer, where the high-level reasoning about the input data is performed. It translates the spatially distributed features into a decision or classification.

Moreover, CNNs have been instrumental in advancing computer vision applications, including medical image analysis [12], autonomous driving [13], and image-based diagnostics. Their robustness, scalability, and ability to handle large-scale datasets make CNNs indispensable tools for researchers and practitioners in various domains. In the realm of medical imaging, CNN-based segmentation techniques have facilitated the identification and delineation of anatomical structures and abnormalities, aiding in diagnosis and treatment planning.

2.2 Training

Training deep learning models involves iterative processes that adjust the model parameters to minimize the error between predictions and actual values. This section covers the essential concepts of backpropagation and optimization, which are foundational to the training process.

2.2.1 Backpropagation

Backpropagation is the key algorithm used for training DL models, particularly neural networks [14]. It involves computing the gradient of the loss function with respect to each weight by the chain rule, iterating backwards from the output layer to the input layer. This process consists of two main steps:

- **Forward Pass:** The input data passes through the network to generate the output predictions.
- **Backward Pass:** The error between the predicted output and the actual output is propagated back through the network. Gradients of the error with respect to each weight are calculated and used to update the weights.

Mathematically, for a neural network with weight parameters $\mathbf{w}$, the loss L, and a learning rate η, the update rule for the weights is:

$$\mathbf{w} \leftarrow \mathbf{w} - \eta \nabla_{\mathbf{w}} L$$

Where $\nabla_{\mathbf{w}} L$ represents the gradient of the loss with respect to the weights. This process is repeated iteratively for multiple epochs until the model converges [15].

2.2.2 Optimization

Optimization techniques are crucial in training neural networks as they determine how the model's weights are updated based on the computed gradients. Several optimization algorithms have been developed to improve convergence speed and performance. Key optimization techniques include:

Stochastic Gradient Descent (SGD): Stochastic Gradient Descent (SGD) is the basic and most commonly used optimization algorithm. It updates the weights incrementally, using a single or a few training examples. While SGD is simple and computationally efficient, it can be slow to converge and may get stuck in local minima [16].

Momentum: Momentum is an extension of SGD that helps accelerate convergence by adding a fraction of the previous update to the current update, smoothing out oscillations. The update rule with momentum is:

$$v_t = \gamma v_{t-1} + \eta \nabla_{\mathbf{w}} L$$

$$\mathbf{w} \leftarrow \mathbf{w} - v_t$$

Momentum helps to smooth out the oscillations and speeds up convergence [17].

Adaptive Learning Rate Methods: Several adaptive learning rate methods have been developed to improve upon SGD:

— **AdaGrad:** Adjusts the learning rate for each parameter individually based on the historical gradients [18].

— **RMSProp:** An improvement over AdaGrad that deals with its radically decreasing learning rates by using a moving average of squared gradients [19].

— **Adam:** Adam, an adaptive optimization algorithm, amalgamates the benefits of RMSProp and Momentum. It dynamically adjusts learning rates for each parameter while maintaining an exponentially decaying average of past gradients and squared gradients [20]. Its update formulas are as follows:

$$
\begin{aligned}
m_t &= \beta_1 m_{t-1} + (1-\beta_1)\nabla_{\mathbf{w}} L \\
v_t &= \beta_2 v_{t-1} + (1-\beta_2)(\nabla_{\mathbf{w}} L)^2 \\
\hat{m}_t &= \frac{m_t}{1-\beta_1^t} \\
\hat{v}_t &= \frac{v_t}{1-\beta_2^t} \\
\mathbf{w} &\leftarrow \mathbf{w} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}
\end{aligned}
$$

2.3 Deep Learning for Image Segmentation

Image segmentation is pivotal in computer vision, entailing the partitioning of an image into distinct regions or objects. This process underpins numerous applications, from medical imaging to autonomous vehicles [21], by enabling machines to interpret visual data for tasks like object recognition and scene understanding.

2.4 Deep Learning for Medical Image Segmentation

2.4.1 Original U-Net Architecture

We can't talk about medical image segmentation without mentioning the U-Net architecture, introduced by Olaf Ronneberger, Philipp Fischer, and Thomas Brox. in their seminal paper "U-Net: Convolutional Networks for Biomedical Image Segmentation" [22] presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) in 2015. The U-Net architecture has become one of the most influential models in medical image segmentation due to its ability to produce precise and reliable segmentations with limited training data.

Architecture Overview: The U-Net architecture consists of a symmetric encoder-decoder structure with skip connections, forming a "U" shape. This structure allows for

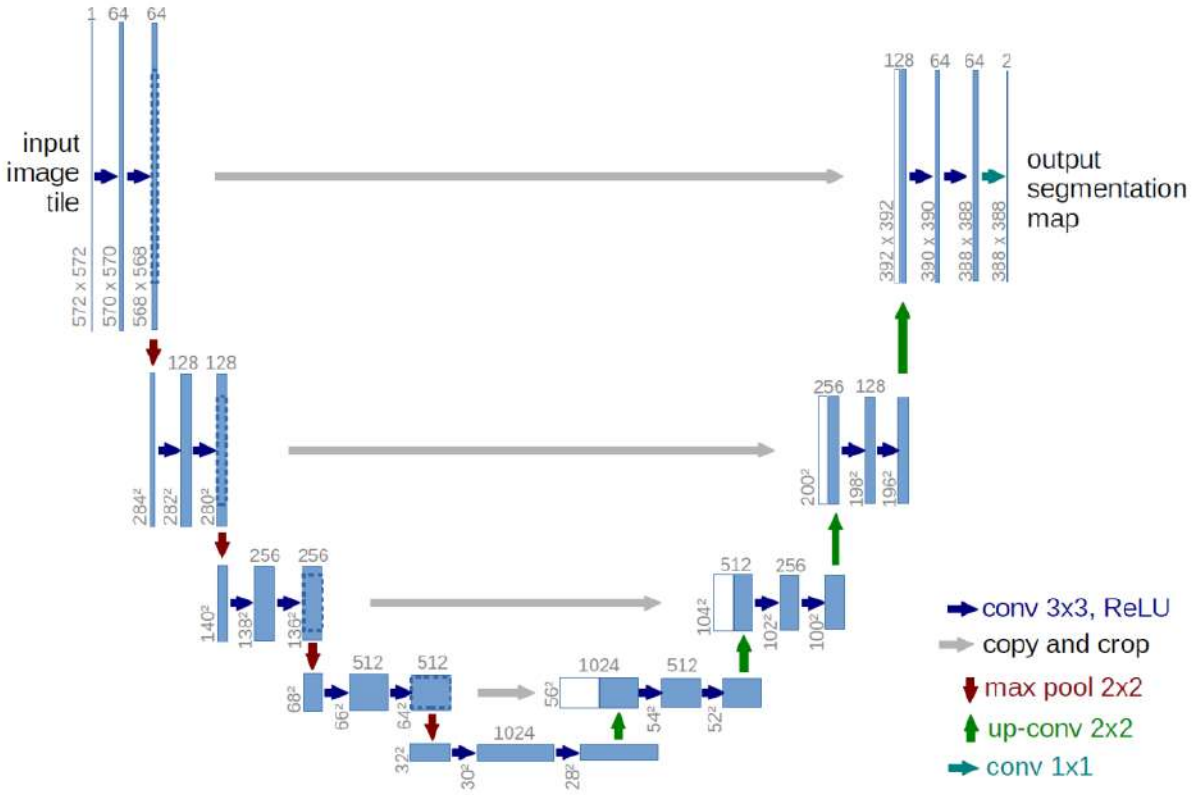

Figure 2.9: standard U-Net architecture

efficient and precise segmentation by combining local and global contextual information.

Encoder (Contracting Path):

— The encoder path on the left side of the U-Net consists of repeated application of two 3x3 convolutional layers, each followed by a ReLU activation function (blue arrows).

— After each convolutional block, a 2x2 max pooling operation (red arrows) is applied, reducing the spatial dimensions while doubling the number of feature channels. This progressive downsampling captures increasingly abstract features.

— The spatial dimensions reduce from 572x572 to 32x32, while the number of channels increases from 1 to 1024.

Bottleneck:

— The bottleneck part at the bottom consists of two 3x3 convolutional layers followed by a ReLU activation, capturing the most abstract features before upsampling begins.

Decoder (Expanding Path):

— The decoder path on the right side mirrors the encoder but replaces pooling layers with 2x2 up-convolution layers (green arrows), which increase the spatial resolution.

— At each upsampling step, feature maps from the corresponding encoder layer are

copied and cropped (gray arrows) to match the dimensions of the upsampled features. This is done through skip connections, which concatenate these copied feature maps with the upsampled output.

— These skip connections ensure that fine-grained spatial information is retained, facilitating precise localization.

Output Layer:

— The final layer consists of a 1x1 convolution (green arrows) that reduces the number of channels to the desired number of classes for the segmentation task, producing the output segmentation map.

Dimensions and Operations: The provided image illustrates the process from input to output. Starting with an input image tile of 572x572 pixels, the image undergoes several stages of convolution (conv 3x3, ReLU), max pooling (2x2), and up-convolution (2x2), resulting in an output segmentation map with dimensions of 388x388 pixels. The architecture effectively handles the trade-off between localization accuracy and the use of context, achieving high performance even with limited annotated data.

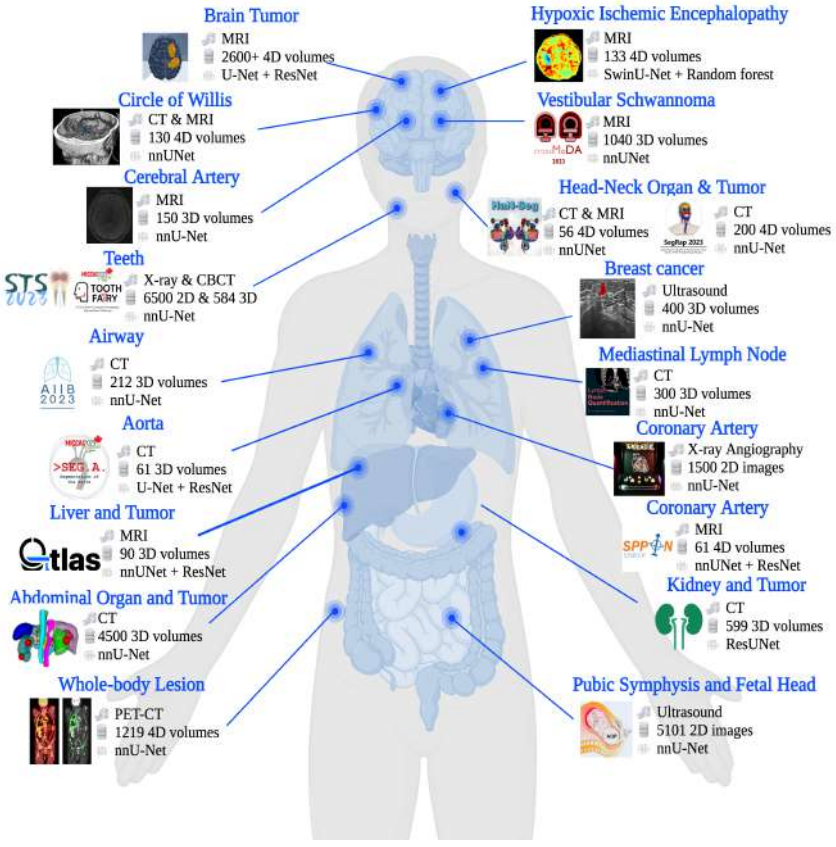

Figure 2.10: Overview of the winning solutions in the MICCAI 2023 segmentation challenges. For each competition, the segmentation target, imaging modality, dataset size, and the neural network architecture used in the winning solution are presented. The challenges encompass a variety of modalities and segmentation targets, demonstrating the diverse and complex nature of medical imaging. U-Net and its variants remain the predominant architectures among the winning solutions.

As depicted in Figure 2.10, the U-Net's innovative design, particularly its use of skip connections, allows it to perform exceptionally well in medical image segmentation tasks. This makes it a cornerstone in the field and has inspired numerous subsequent variants and improvements. Moreover, its versatility in accepting different types of input data, such as grayscale, color, and multi-channel images, enhances its applicability across a wide range of medical imaging modalities and datasets.

2.4.2 Improved U-Net Variants

2.4.2.1 3D U-Net

The 3D U-Net [23] is an extension of the original U-Net architecture designed to handle volumetric data. This adaptation is particularly useful for medical imaging tasks where the data is inherently three-dimensional, such as in MRI or CT scans. The key difference between the original 2D U-Net and the 3D U-Net lies in their ability to process and learn from volumetric data, capturing spatial context in three dimensions rather than two.

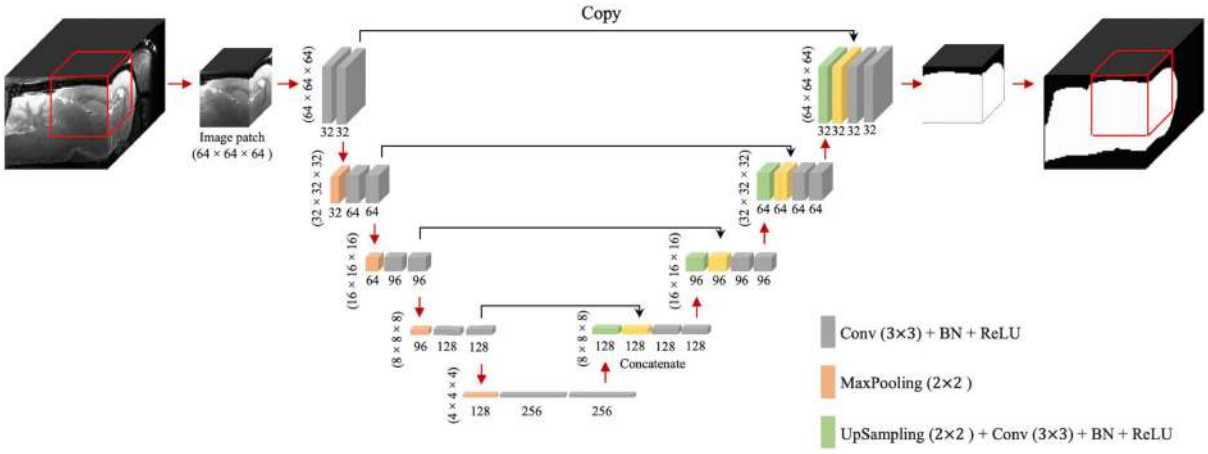

Figure 2.11: Architecture of the 3D U-Net.

Differences and Enhancements in 3D U-Net

1. Input and Output:

— *3D Input Data*: Unlike the original U-Net which processes 2D image slices, the 3D U-Net takes 3D image patches as input. In the provided diagram, the input is a volumetric patch of size $64 \times 64 \times 64$.

— *3D Output*: The output is also a 3D volume, with the same spatial dimensions as the input, but segmented to highlight the regions of interest.

2. Convolutional Layers:

— *3D Convolutions*: The 3D U-Net employs 3D convolutional layers (Conv $3 \times 3 \times 3$) to extract features across three dimensions. This allows the network to learn

spatial hierarchies within the volume, enhancing the ability to capture complex structures.

— *Batch Normalization*: Each convolutional layer is followed by batch normalization (BN) and ReLU activation, aiding in the stabilization and acceleration of the training process.

3. Downsampling and Upsampling:

— *3D Max Pooling*: Similar to the 2D U-Net, the 3D U-Net uses max pooling layers for downsampling, but these operations are now performed in three dimensions (MaxPooling $2 \times 2 \times 2$).

— *3D Up-Convolution*: The upsampling path uses 3D transposed convolutions (Up-Sampling $2 \times 2 \times 2$) to expand the feature maps back to their original dimensions. This is crucial for maintaining the volumetric context of the data.

4. Skip Connections:

— *Copy and Crop*: The skip connections in the 3D U-Net are similar to those in the original U-Net, involving copying and cropping feature maps from the contracting path and concatenating them with the corresponding upsampled feature maps in the expansive path. This helps in preserving fine-grained spatial information throughout the network.

5. Feature Channel Expansion:

— *Channel Dimensions*: As with the 2D U-Net, the number of feature channels increases as the spatial dimensions decrease in the contracting path (e.g., from 32 to 128 channels). This allows the network to capture more complex features at deeper layers.

Overall Structure:

— *Symmetric U-Shape*: The overall structure retains the symmetric U-shape, ensuring a balanced approach to downsampling and upsampling while effectively utilizing the skip connections to merge low-level and high-level features.

— *Final Segmentation Map*: The final segmentation map has the same 3D dimensions as the input, ensuring a direct and accurate representation of the segmented regions within the original volume.

The 3D U-Net architecture significantly improves upon the original U-Net by enabling the processing of volumetric data, making it highly suitable for medical imaging tasks that require a detailed understanding of spatial context in three dimensions. This enhancement allows for more accurate and context-aware segmentation of lesions and other structures within volumetric medical images.

2.4.2.2 nnU-Net

The nnU-Net [24] represents a transformative approach in the field of medical image segmentation by introducing a self-configuring method that adapts to different datasets without manual tuning. The nnU-Net framework, proposed by Fabian Isensee and colleagues, automates the configuration of the network architecture, training, and inference pipelines, making it a robust and versatile solution for various segmentation tasks.

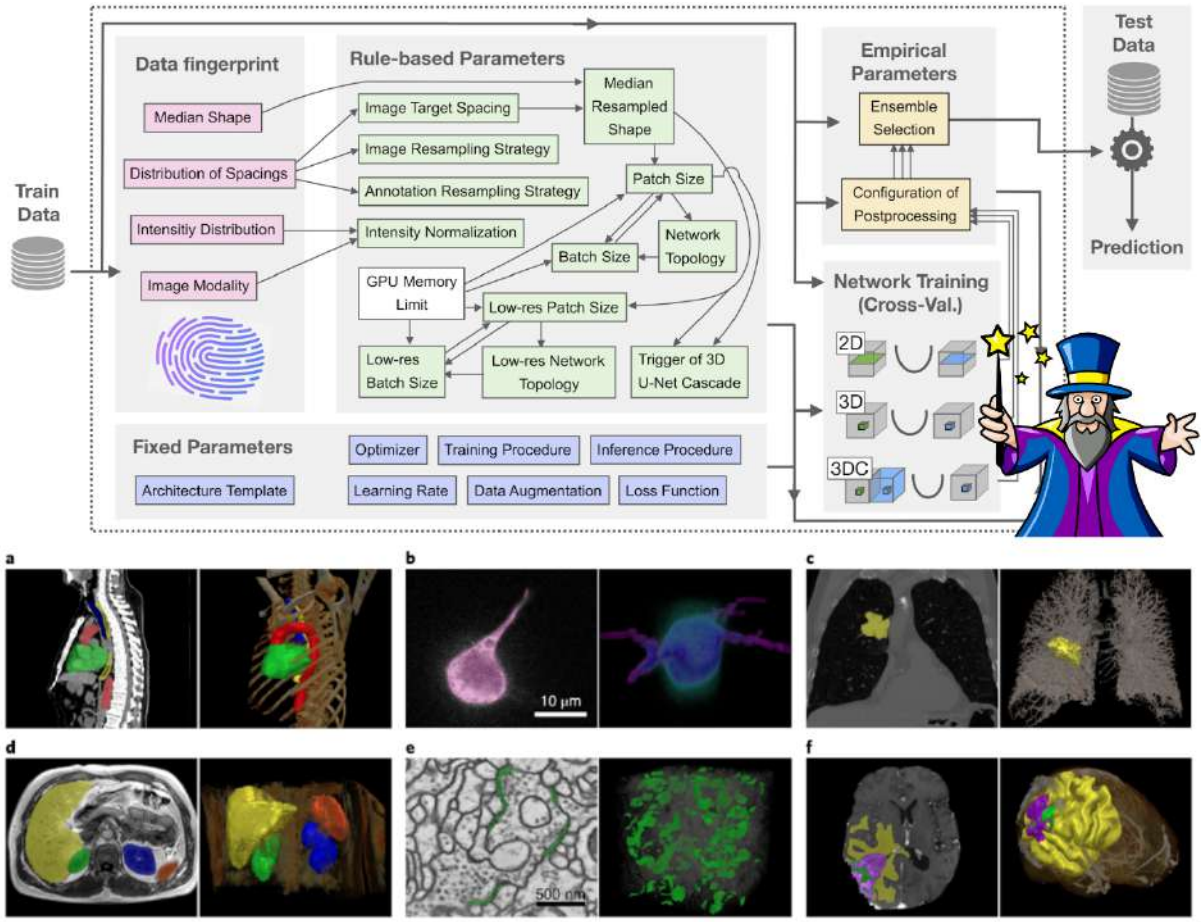

Figure 2.12: Workflow and examples of nnU-Net segmentation results.

Key Features of nnU-Net:

— **Automated Configuration:** nnU-Net leverages a set of rule-based heuristics to automatically adapt the network architecture and training pipelines based on the characteristics of the input data. This eliminates the need for extensive manual tuning and expert knowledge.

— **Data Fingerprint:** The framework starts by analyzing the input data to create a 'data fingerprint,' which includes information such as median shape, spacing distribution, intensity distribution, and image modality. This fingerprint guides the configuration process.

— **Rule-based Parameters:** Based on the data fingerprint, nnU-Net adjusts pa-

rameters like image resampling strategy, intensity normalization, batch size, patch size, and network topology. These adjustments ensure optimal performance for the specific dataset.

— **Fixed Parameters:** Despite its adaptability, nnU-Net maintains certain fixed parameters, such as the optimizer, learning rate, data augmentation techniques, and loss function. These fixed elements provide a stable foundation for training.

— **Empirical Parameters:** The method also includes empirical parameters for ensemble selection and postprocessing configuration, enhancing the robustness and accuracy of the final predictions.

— **Cross-Validation and Ensemble Learning:** nnU-Net employs cross-validation to assess the performance of different configurations and uses ensemble learning to combine the strengths of multiple models, leading to more reliable segmentation results.

Workflow Overview: The nnU-Net workflow can be summarized as follows:

— *Data Fingerprinting:* Analyze the training data to extract key characteristics.

— *Rule-based Configuration:* Automatically configure the network architecture and training pipeline based on the data fingerprint.

— *Training:* Train the network using cross-validation to ensure robust performance.

— *Ensemble Selection and Postprocessing:* Select the best-performing models and configure postprocessing steps for optimal predictions.

— *Inference:* Apply the trained model to test data to generate segmentation results.

Results and Impact: The bottom part of Figure 2.12 showcases example results from nnU-Net, demonstrating its capability to accurately segment various medical images, such as MRI, CT scans, and microscopic images. The nnU-Net has been recognized for its state-of-the-art performance across numerous medical image segmentation challenges, highlighting its versatility and effectiveness.

2.5 Evaluation Metrics

Evaluation is an essential aspect of assessing the performance of segmentation algorithms. It involves quantifying the accuracy, precision, and other metrics to determine how well the segmented regions align with ground truth annotations or expected outcomes. In this section, we'll explore common evaluation techniques.

2.5.1 Confusion Matrix

The confusion matrix is a tabular representation used to visualize the performance of a segmentation algorithm by comparing predicted labels with ground truth annotations. As shown in Figure 2.13, the confusion matrix consists of four quadrants:

— **True Positives (TP)**: Pixels correctly classified as part of the segmented region.
— **True Negatives (TN)**: Pixels correctly classified as not part of the segmented region.
— **False Positives (FP)**: Pixels incorrectly classified as part of the segmented region.
— **False Negatives (FN)**: Pixels incorrectly classified as not part of the segmented region.

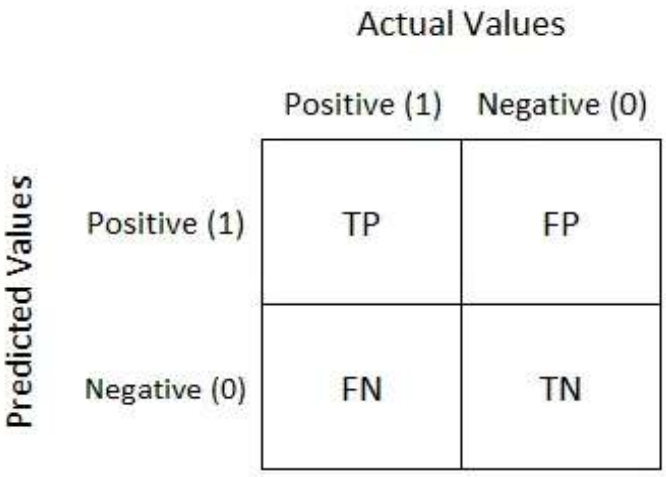

Figure 2.13: Confusion Matrix.

Each quadrant represents the number of pixels or regions correctly or incorrectly classified by the algorithm. The confusion matrix serves as the foundation for calculating various performance metrics.

2.5.2 Metrics

Various metrics can be derived from the confusion matrix to quantify the performance of segmentation algorithms:

— **Accuracy**: The ratio of correctly classified pixels to the total number of pixels in the image.

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

— **Precision**: The ratio of true positive pixels to the total number of pixels classified as positive (true positives plus false positives). It measures the algorithm's ability to correctly identify relevant regions.

$$\text{Precision} = \frac{TP}{TP + FP}$$

— **Recall (Sensitivity)**: The ratio of true positive pixels to the total number of ground truth positive pixels (true positives plus false negatives). It measures the algorithm's ability to correctly detect relevant regions.

$$\text{Recall} = \frac{TP}{TP + FN}$$

— **F1 Score**: The harmonic mean of precision and recall, providing a balance between the two metrics.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

— **Intersection over Union (IoU)**: Also known as the Jaccard index, it measures the overlap between predicted and ground truth regions.

$$\text{IoU} = \frac{TP}{TP + FP + FN}$$

— **Dice Coefficient**: Similar to IoU, it measures the overlap between predicted and ground truth regions but provides a slightly different formulation.

$$\text{Dice Coefficient} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

These metrics provide insights into different aspects of segmentation performance, including accuracy, completeness, and robustness. Depending on the specific application and requirements, different metrics may be prioritized. Evaluation metrics play a crucial role in comparing different segmentation algorithms, optimizing parameters, and assessing algorithmic improvements.

Chapter 3

State of the Art

Introduction

Accurate and efficient analysis of medical images is essential for early detection, treatment planning and monitoring of many diseases at different stages related to lesions or tumors. The most important step in this analysis is lesion segmentation.

Deep learning, especially Convolutional Neural Networks (CNNs) has transformed the medical image analysis domain with automatic and precise segmentation tools. Models from this family have shown superb performance across several tasks related to medical image analysis that far exceeds traditional methods.

This chapter provides a comprehensive overview of the state-of-the-art deep learning approaches for lesion segmentation, given the focus of this thesis on universal lesion segmentation across different types of lesions. The chapter is structured to address the specific challenges and advancements in segmentation for various lesion types present in our thesis. These include bone lesions, lung nodules, kidney lesions, liver lesions, lymph node lesions, pancreas lesions, and colon lesions.

3.1 Kidney and Kidney Tumor Segmentation

Kidney cancer ranks among the most prevalent adult malignancies globally, with an increasing incidence [25]. Most renal tumors are detected at an early, localized, and operable stage. However, significant questions regarding the management of localized kidney tumors remain unresolved [26], and metastatic renal cancer continues to be almost invariably fatal [27]. Renal tumors are well-known for their distinct appearance in computed tomography (CT) imaging, facilitating significant research by radiologists and surgeons into the correlation between tumor size, shape, appearance, and treatment outcomes [28] [29] [30].

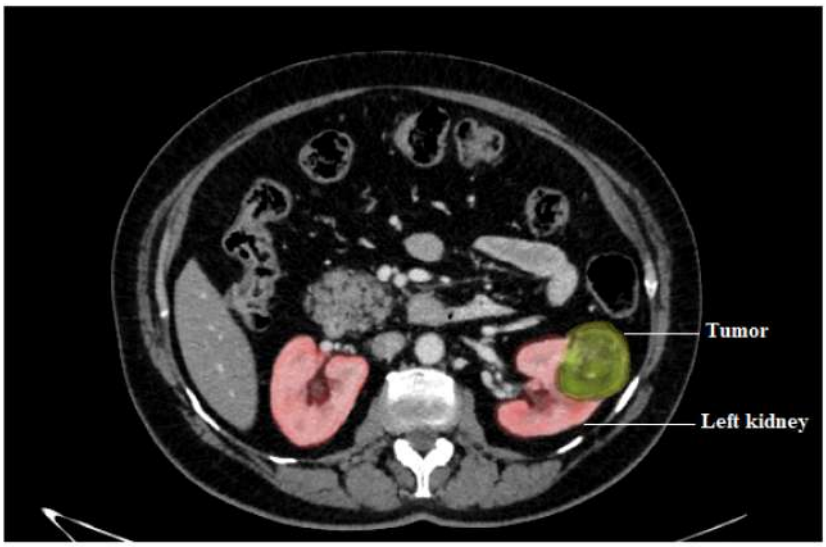

Figure 3.1: Illustration of Kidney and Kidney tumor segmentation result.

The automatic segmentation of renal tumors and surrounding anatomy offers a promising solution to these challenges. Segmentation-based evaluations are objective and inherently well-defined, and automation reduces the required effort to a single action.

In order to advance the development of robust tools to meet this need while providing a high-quality benchmark for various segmentation methodologies, the 2019 Kidney Tumor Segmentation Challenge [31] where firstly created for this task. Next, the 2021 Kidney and Kidney Tumor Segmentation Challenge (KiTS21) is proposed to address the need for efficient and reliable tools for automated renal tumors and adjacent anatomical structures segment. Several methods have been proposed to compete in this challenge to come up with the most robust solution for kidney and kidney tumor segmentation.

Transfer learning-based methods have been proposed in [32] for kidney segmentation. In which the authors have used a trained 3D-Unet model on LITs datasets (liver tumor segmentation). the model weights have been transferred for kidney segmentation. This proposed method has shown promising results for this task obtaining second place in the Kits challenge.

3D-Unet model is also used in [33] for kidney and kidney tumor segmentation .The model was trained from scratch on the kits21 CT images dataset. The experimental results proved the effectiveness of the 3D-Unet model for kidney tumor segmentation.

Authors in [34] have proposed a coarse-to-fine tumor segmentation framework based on the nnUNet model. The proposed framework has consisted on two main phases the coarse phase focuses on training the nnUNet model for segmenting the kidney organ whereas the fine-grained phase aims to train the nnUNet model on the kidney organ result to obtain the final kidney tumor segmentation mask. the framework has been proposed for the kits21 challenge to obtain the first place.

A two-stage deep learning architecture, named convolutional computer tomography

network (CCTNet) [35], was proposed for kidney and kidney lesions segmentation. The first stage uses a pre-trained 3D low-resolution nnU-Net to locate the volume bounding box containing both kidneys. The second stage performs the segmentation of the kidney, kidney tumors, and cysts using mixup augmentation to enhance the performance of a second 3D full-resolution nnU-Net. This method demonstrated improved segmentation accuracy for kidneys, kidney tumors, and cysts.

A cascaded method based on the 3D U-Net architecture was proposed to accurately segment kidneys and their lesions [36]. The approach involves a two-stage process: the first stage uses a lower-resolution network to obtain a rough segmentation of the kidney region, while the second stage refines the segmentation of tumors and cysts within the kidney region. This method showed high segmentation performance.

3.2 Lung Nodules Segmentation

Lung nodules are common abnormal growths in the lungs, often discovered through imaging tests like CT scans or X-rays (see Figure 3.2). While these nodules are usually benign and not cancerous, they can occasionally indicate lung cancer. They appear as shadows or spots on the scans and can result from respiratory infections, scarring, or inflammation. Although most lung nodules do not cause symptoms, larger ones may lead to coughing, wheezing, or shortness of breath. Diagnosing lung nodules often involves follow-up imaging and, if necessary, biopsies to rule out malignancy.

Figure 3.2: CT scan image showing a lung nodule (circled in red) in the right lung.

Lung cancer is a leading cause of death globally, surpassing other cancers in incidence and mortality rates [37], posing a significant public health challenge. Early signs of lung cancer include the formation of lung nodules, categorized into seven types: isolated, juxta pleural, juxta vascular, cavitary, calcific, and small nodules. Most lung cancer diagnoses occur too late for optimal treatment, making early detection crucial. Precise segmentation of lung nodules is essential for effective treatment but challenging for clinicians, leading to misdiagnosis and missed diagnoses. While deep learning has made progress in lung nodule segmentation, significant challenges remain.

A new system combining SegNet and Generative Adversarial Networks (GANs) was developed in [38], to improve the segmentation of pulmonary nodules, which are critical in lung cancer diagnosis. The method generates a large set of synthetic realistic images using GANs. These GAN-based generated images along with real images from the LUNA16 dataset, are used to enhance SegNet through transfer learning, boosting segmentation accuracy. The research results were validated with various combinations of methods showing that incorporating GAN-generated images significantly improved the efficiency and accuracy of the segmentation process.

To address the diverse shapes and visual features of nodules for lung nodules segmentation in CT images, a new deep learning architecture was introduced in [39], This model integrates a bidirectional feature network (Bi-FPN) between the encoder and decoder, and employs the Mish [40] activation function along with class weights for masks to enhance segmentation performance. The model was trained and tested on the LUNA16 dataset, and further evaluated on the QIN Lung CT dataset. The use of weighted binary cross-entropy loss during training improved the segmentation accuracy. The proposed architecture outperformed existing models like U-Net.

The multi-encoder-based self-adaptive hard attention network (MESAHA-Net) was proposed in [41]addressing challenges such as nodule heterogeneity and size diversity. MESAHA-Net features three encoding paths, an attention block, and a decoder block, enabling it to integrate CT slice patches, forward and backward maximum intensity projection (MIP) images, and region of interest (ROI) masks. Using an adaptive hard attention mechanism, the network iteratively performs slice-by-slice 2D segmentation, focusing on nodule regions to produce 3D volumetric segmentations. Evaluated on the LIDC-IDRI dataset, MESAHA-Net demonstrated high robustness across various lung nodule types, outperforming previous state-of-the-art techniques in segmentation accuracy and computational efficiency, making it suitable for real-time clinical use.

An improved U-Net method was proposed [42]to enhance the segmentation accuracy of lung nodule CT images. This approach integrates dense network connections and sawtooth-expanded convolution designs in the feature extraction phase, along with a local residual design during upsampling. The effectiveness of this algorithm was validated using the LIDC-IDRI lung nodule dataset, demonstrating significantly higher performance

compared to the standard U-Net. This method showcased a more effective network structure for lung nodule segmentation tasks

3.3 Liver Tumor Segmentation

.

The liver, the largest solid organ in the human body, is crucial for metabolism and digestion. Globally, primary liver cancer ranks as the second deadliest cancer [43]. Computed tomography (CT) is extensively used to evaluate liver structure, texture, and focal lesions [44]. Liver abnormalities are critical biomarkers for diagnosing and assessing both primary and secondary hepatic tumors [45]. The liver often serves as the origin of primary tumors. Moreover, cancers originating from other abdominal organs like the colon, rectum, and pancreas, as well as from distant organs such as the breast and lung, frequently metastasize to the liver. Consequently, the liver and its lesions are regularly analyzed for comprehensive tumor staging.

The standard Response Evaluation Criteria in Solid Tumor (RECIST) or modified RECIST guidelines necessitate measuring the diameter of the largest target lesion [46]. Therefore, precise segmentation of focal lesions is vital for cancer diagnosis, treatment planning, and monitoring treatment response. Specifically, identifying tumor lesions in a given scan is essential for many treatment options, including thermal percutaneous ablation [47], radiotherapy, surgical resection [48], and arterial embolization [49]. Similar to other medical imaging tasks, manually delineating target lesions in 3D CT scans is time-intensive, inconsistent, and operator-dependent [50].

Achieving fully automated segmentation of the liver and its lesions presents several challenges. Firstly, variations in lesion-to-background contrast [51] arise from different contrast agents and variations in contrast enhancement due to differing injection timings and acquisition parameters. Secondly, the presence of various types of focal lesions (benign vs. malignant and different tumor subtypes) with diverse image appearances further complicates automated segmentation. Thirdly, chronic liver disease, a common precursor to liver cancer, can significantly alter liver tissue background signals. Many algorithms struggle with disease-specific variability, including differences in lesion size, shape, and number, as well as changes in the liver's shape and appearance due to treatment [51]. Figure 3.3 depicts a CT scan image of the liver with multiple tumors.

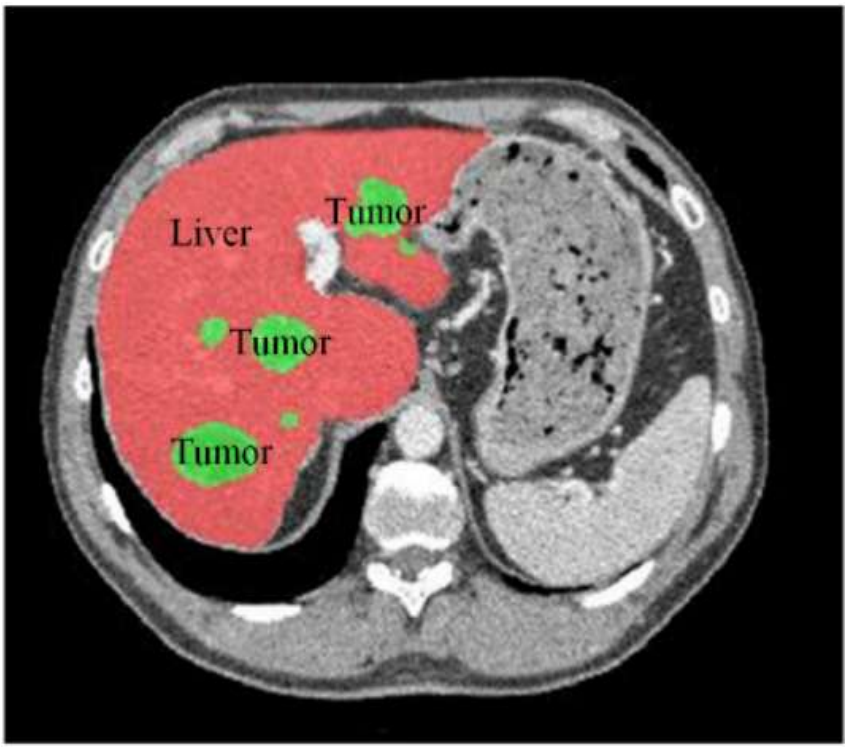

Figure 3.3: CT scan highlighting the liver and tumors in the abdominal region.

The study in [52] introduces LightM-UNet, a model designed to overcome the computational challenges of traditional U-Net and Transformer-based architectures in medical image segmentation. By leveraging the lightweight Mamba framework, LightM-UNet drastically reduces parameter and computation costs while maintaining high performance. Evaluations on the LiTs (3D CT images) and Montgomery and Shenzhen (2D X-ray images) datasets show that LightM-UNet surpasses existing models like nnU-Net, delivering superior results with significantly lower resource requirements.

A new method was introduced for liver tumor segmentation using multi-phase CT images, addressing limitations in current techniques that often overlook spatial relationships between different phases [53]. The proposed approach incorporates a Spatial Aggregation Module (SAM) to enhance pixel-wise interactions across phases, ensuring more effective integration of multi-phase information. Additionally, an Uncertain Region Inpainting Module (URIM) is utilized to refine uncertain regions by leveraging neighboring features. This method was tested on a multi-phase CT image dataset of focal liver lesions, demonstrating superior segmentation performance compared to existing state-of-the-art techniques. The network employs a ResNet-50 backbone with Siamese streams for phase-specific feature extraction, followed by bi-directional aggregation and multi-level feature fusion, culminating in precise tumor segmentation.

Researchers developed a method for segmenting liver tumors in CT scans using synthetic tumors instead of manual annotations [54]. These synthetic tumors offer two main advantages: they closely resemble real tumors in shape and texture, making them al-

most indistinguishable to medical professionals, and they are highly effective for training AI models. Remarkably, AI models trained with these synthetic tumors achieved performance on par with those trained on real tumors. This innovation could significantly reduce the manual effort required for tumor annotation, which is time-consuming. Additionally, the method can generate numerous examples of small tumors, improving early cancer detection rates and enabling robust assessment of AI models' reliability. The process involves a pre-trained nnUNet to localize the liver, followed by steps to generate tumors: location selection, texture and shape generation, and post-processing.

The study in [55], introduces an advanced Deep Convolutional Neural Network (DCNN) for liver lesion segmentation, specifically designed for the LiTS challenge. The model processes 2.5D CT slices, leveraging a 32-layer architecture that combines U-Net's long-range concatenation connections with ResNet's short-range residual connections. This hybrid design enhances segmentation accuracy by utilizing rich contextual information and high-resolution features, resulting in superior performance in the LiTS challenge. This approach demonstrates the efficacy of integrating advanced network architectures for precise medical image analysis.

The work in [56] presents an automatic liver tumor segmentation method using a fully convolutional neural network (FCNN) complemented by object-based postprocessing. Applied to CT images, the method employs a 2D FCNN followed by a cascading model design that significantly reduces false positives by 85%. Evaluated on the LiTS challenge training dataset, the method demonstrates segmentation quality comparable to human performance for detected tumors but lags in detection performance. This approach underscores the potential of combining neural networks with object-level postprocessing to enhance accuracy in medical image segmentation.

3.4 Bone Lesion Segmentaiton

Bone metastases frequently occur in various cancers, particularly lung, breast, and prostate [57]. These lesions often cause significant pain and are associated with high mortality rates. Patients with metastases from breast, prostate, and renal cancers typically survive between 12 to 33 months, while those with primary lung cancer and bone metastases face significantly lower survival rates, ranging from 9.5% to 12% within one year [57]. The extent of bone metastasis is strongly correlated with reduced survival outcomes [58]. Current clinical approaches primarily involve palliative chemotherapy and radiotherapy [59]. Recent advancements in image-guided radiotherapy techniques, such as stereotactic body radiotherapy (SBRT), offer the potential to deliver curative radiation doses while minimizing damage to surrounding healthy tissues [60–62]. Clinical trials, like the SABR-COMET trial, have demonstrated SBRT's efficacy in managing metastatic disease [63]. Effective treatment strategies show promise in enhancing overall

survival and long-term progression-free survival rates [64]. Using radiomic biomarkers to quantify metastatic lesion numbers, their spatial distribution, and predict treatment responses can further refine treatment planning [65].

In medical image analysis, various imaging modalities such as positron emission tomography (PET) [66], whole-body Magnetic Resonance Imaging (MRI) [67], and hybrid nuclear medicine combining CT with PET play crucial roles in diagnostics and clinical monitoring. PET imaging, known for its functional insights, is particularly valuable in cancer evaluation. Hybrid nuclear medicine equipment offers benefits like attenuation correction and improved anatomical-functional image correlation. Recent studies highlight the use of CT and PET for volume of interest (VOI) segmentation based on anatomical contours . Medical literature extensively covers segmentation methods, their efficacy, and challenges, emphasizing the integration of clinical expertise to optimize algorithms beyond conventional CNN models. While machine learning aids radiologists, it doesn't replace clinical judgment due to limitations in supervised learning and inter-observer variability among radiologists. Further research is essential to validate algorithmic use in clinical decision-making and ensure their integration into medical practice enhances patient outcomes.

The paper in [68] introduces a deep-learning segmentation method for bone scan images to aid in monitoring bone metastases. It utilizes a Swin Transformer for feature extraction and step-by-step feature fusion to emphasize lesion details while minimizing distractions. The method demonstrates effective performance on a constructed dataset.

In the paper [69] Researchers introduced a convolutional neural network (ConvNet) for segmenting CT images, specifically for lesions, bone, and background. Inspired by the Fuzzy C-means algorithm, this method employs unsupervised, semi-supervised, and supervised loss functions. It significantly outperforms traditional clustering techniques, achieving higher accuracy and faster computation. This flexible ConvNet offers efficient and reliable segmentation for CT scans.

In the study presented by [70], a semi-supervised segmentation model is proposed for automated identification and delineation of skeletal metastasis lesions in bone scan images. The model utilizes dilated residual convolution, inception connection, and feature aggregation for feature extraction. It applies pixel classification and boundary delineation through closed curves, aiming to assist nuclear medicine physicians by focusing on segmented lesions and reducing background interference in bone metastasis diagnosis from low-resolution images.

The nnU-Net model was utilized for segmenting bones and bone lesions on PET/CT images in metastatic breast cancer [71]. Including both bone and lesion data in the training process improved lesion detection precision and segmentation accuracy. An automatic PET bone index was also computed, aligning with manually calculated indices. This approach offers an automated solution for monitoring metastatic breast cancer.

The paper [72] proposes SEAGNET, an intelligent framework for segmenting bone malignant tumor lesions in medical images. SEAGNET uses a supervised edge-attention guidance approach to accurately locate tumors with complex backgrounds and unclear boundaries. It incorporates a module for selecting key points on boundaries to capture detailed edge features, enhancing segmentation precision. By employing mixed attention mechanisms, SEAGNET addresses uncertainties in tumor boundaries. Experimental results on real medical data validate SEAGNET's effectiveness, demonstrating its potential to improve diagnostic accuracy and clinical efficiency.

3.5 Colon Lesion Segmentation

Colon or colorectal cancer is one of the most prevalent cancers globally and ranks as the second leading cause of cancer-related mortality in men. The development of cancer in the lower digestive tract often originates from polyps. Early detection of these polyps can significantly increase survival rates from 10-15% to 60-80% [73].

According to the American Cancer Society, survival rates for colorectal cancer survivors are lower at 65.4% compared to breast cancer (90.35%) and prostate cancer (99.6%) [74]. Two-dimensional imaging technologies like computed tomography (CT) play a crucial role in detecting colorectal cancer. CT scans have been instrumental in reducing mortality rates by 20%. While MRI offers higher accuracy than CT scans, its higher cost makes CT scans more accessible in developing countries [75] [76].

Advances in CT scanning technology have led to a vast amount of electronic health records, posing challenges for radiologists in diagnosing colon cancer. Moreover, manually detecting and diagnosing polyps from extensive datasets is time-consuming and labor-intensive. To address these challenges and improve detection efficiency, computer-aided diagnosis (CAD) systems are being integrated into radiology practices [77].

DeepCRC in the paper [78] introduces a deep learning method for segmenting the colorectum and colorectal cancer (CRC) in abdominal CT scans. It employs a 1D colorectal coordinate system to improve segmentation accuracy and continuity by encoding positional information. The method includes an auxiliary regression task for predicting colorectal coordinates, enhancing global topological integration. Self-attention layers are added to the model to further improve global context modeling. Tested on 107 cases, DeepCRC surpasses nnUNet in segmentation performance and achieves human-like reproducibility in tumor segmentation.

The proposed 3D RoI-aware U-Net (3D RU-Net) [79] enhances colorectal tumor segmentation from 3D MR images by integrating RoI localization and in-region segmentation within a single encoder network. This approach improves segmentation accuracy and efficiency by enlarging the receptive field and preserving details. The model is trained using a Dice-formulated loss function and demonstrates superior performance on 64 cancer-

ous cases with a four-fold cross-validation. The framework shows promise for other 3D medical image segmentation tasks, with code available publicly.

In the paper [80] A method using fully convolutional networks (FCNs) has been developed for segmenting colorectal tumors on T2-weighted MRI. This approach reduces the tedium and variability of manual segmentation. It uses VGG-16 for feature extraction and combines predictions from multiple layers to determine tumor boundaries. A quantitative comparison of 2772 manual segmentations showed this method outperforms U-net in accuracy.

The paper [81] showcasing that the Adaptive Lesion-Aware Attention Network (ALA-Net) enhances 3D colorectal tumor segmentation on MRI by integrating contextual information and spatial details. It uses two parallel encoding paths to capture global features and sharp object boundaries, while an attention module focuses on discriminative features. This approach reduces errors and improves accuracy, outperforming other methods and generalizing well to various 3D medical image segmentation tasks.

Segmenting colorectal tumors in CT scans is difficult due to irregular tumor contours. To address this, a new algorithm using a convolutional neural network with an attention mechanism is proposed [82]. This network has three main modules: an encoder for feature mapping, a dual attention module for enhancing contextual information, and a decoder to resize the feature map to the input image's original size. Tested on 1131 CT slices, this method outperformed U-Net and CE-Net, showing improved effectiveness in colorectal tumor segmentation.

3.6 Pancrease Lesion Segmentation

Pancreatic lesions, often indicative of pancreatic cancer or cystic neoplasms, are critical areas of focus due to the aggressive nature and poor prognosis associated with pancreatic malignancies. These lesions can be detected using imaging techniques such as CT scans, MRI, and endoscopic ultrasound (EUS). However, the accurate segmentation of pancreatic lesions in these imaging modalities presents significant challenges due to the pancreas's deep anatomical location and the complex surrounding structures.

Segmenting pancreatic tumors in multi-parametric MRI is essential for MRI-guided Online Adaptive Radiation Therapy (MRgOART). A CNN-based model was developed [83] and trained on DCE MRI sets from 27 patients. It was tested on additional MRI sets, showing performance comparable to expert oncologists, offering a reliable method for auto-segmentation in MRgOART.

Deep LOGISMOS proposed in [84] is a method for 3D tumor segmentation in CT scans, combining deep learning and graph-based techniques. First, a UNet is trained to provide initial segmentation. This segmentation is refined using Gaussian Mixture Models (GMM) and morphological operations. The refined boundaries are used to build

a segmentation graph, and a max-flow algorithm finds the optimal solution. Tested on 51 CT scans, this approach showed significant improvement in accuracy compared to using UNet or LOGISMOS alone.

The study in [85] evaluates deep learning for segmenting pancreatic tumors in contrast-enhanced endoscopic ultrasound (CE-EUS) video images. Using U-Net and manual segmentation as the ground truth, the study tested 100 patients' CE-EUS video segments. Results showed a median Intersection over Union (IoU) of 0.77, with clearer tumor boundaries yielding higher IoU. The segmentation was not significantly affected by respiratory movement.

The study in [86] addresses the segmentation of pancreatic tumors using a deep learning architecture called Multi-Scale Channel Attention Unet (MSCA-Unet). The tumors are often small, irregular, and have unclear boundaries. The model embeds a multi-scale network into the encoder to extract semantic information at different scales, and into the decoder to mitigate information loss during upsampling. Channel attention units emphasize informative channels, improving the accuracy of tumor localization and reducing false positives. Tested on both private and public datasets, the network achieved superior performance in tumor segmentation.

The work in [87] proposes an Antlion Optimization-Convolutional Neural Network-Gated Recurrent Unit (ALO-CNN-GRU) model for the segmentation and classification of pancreatic cancer in CT scans. The model consists of preprocessing to reduce noise using a hybrid Gaussian and median filter, segmentation of the affected pancreatic area using the Antlion optimization method, and classification of the cancer as benign or malignant using CNN and GRU networks.

3.7 Lymph Node Segmenation

Lymph node lesions are critical in the diagnosis and staging of various cancers, including colorectal, breast, and lung cancers. These lesions can indicate the presence of metastatic disease, making their accurate detection and segmentation essential for effective treatment planning. Lymph nodes can be challenging to segment due to their small size, low contrast in imaging, and variability in shape and appearance.

In the paper [88] the authors introduces a method for detecting and segmenting lymph node gross tumor volumes (GTVLN) in 3D CT/PET imaging used in radiotherapy. By dividing GTVLN into "tumor-proximal" and "tumor-distal" subgroups, the approach mimics clinical reasoning to streamline the process. A specialized multi-branch detection-by-segmentation network handles each subgroup, with outputs fused during inference. Tested on a dataset of esophageal cancer patients, the method showed significant improvements, demonstrating its clinical relevance and value.

The authors in [89] addresses the detection and segmentation of cancerous lymph

nodes in PET/CT imaging for breast cancer staging. It explores the use of region-based features to enhance segmentation accuracy. Initially, various features are evaluated using a Random Forest framework to determine their relevance. The top-ranked features are then integrated into a U-Net segmentation architecture for further validation. The hierarchical modelling of PET images aids in defining these features, demonstrating their potential to significantly improve lymph node segmentation in PET/CT.

Histology assessment of biopsy tissue remains the gold standard for cancer diagnosis, but manual segmentation of nuclei poses challenges due to overlapping. The study in [90] aims to automate nuclei segmentation using MATLAB to accelerate histological assessment. A dataset profiling normal and cancerous tissues was compiled for reference. MATLAB was employed for image analysis, including 8-bit grayscale conversion, adaptive filtering with Wiener filter for noise reduction, and Otsu thresholding for segmentation. Results were comparable to ImageJ and manual counting, showing lower cellularity in cancerous lymph nodes compared to normal ones.

The study in [91] focuses on automating the segmentation of primary tumors (GTVp) and pathological lymph nodes (GTVn) in oropharyngeal cancer using deep learning. The Swin UNETR model, initialized with pre-trained weights from a self-supervised Swin UNETR model, was employed for 3D PET/CT imaging. Data from the HECKTOR 2022 challenge, comprising CT and PET scans with ground-truth segmentations, were used to validate the approach.

An algorithm for lymph node image segmentation was proposed in [92], the authors used improved FCM clustering and multi-threshold techniques. The method aims to enhance segmentation accuracy in CT images with blurred edges typical of lymph nodes. Improved FCM peak clustering sharpens fuzzy boundaries, followed by a multi-threshold approach based on image entropy changes to refine segmentation. Experimental results demonstrate superior segmentation outcomes compared to traditional FCM methods, particularly effective in cases with fuzzy lymph node edges.

3.8 Gaps and Contributions

After a careful review of our related works, we noticed that there is no study that perform a universal lesion segmentation. Additionally, there is no attempt that investigate the role of recent optimizers for lesion segmentation. Moreover, the performance of the current lesion segmentation system has not yet reached its optimal potential. Our proposition goes in contrast to that of many related works that require a huge amount of labeled data to achieve acceptable performance.

Conclusion

In this chapter, we outlined our related works concerning Kidney and Kidney Tumor Segmentation Lung Nodules, Segmentation Liver Tumor Segmentation, Bone Lesion Segmentation, Colon Lesion Segmentation, Pancrease Lesion Segmentation and Lymph Node Segmentation. We conclude with a synthesis. The next chapter will introduce our proposed framework for universal lesion segmentation.

Chapter 4

Contribution and Results

Introduction

This chapter outlines the key contributions and results of the research presented in this thesis. Building on the foundations laid in the previous chapter, which reviewed the state-of-the-art techniques in medical image segmentation.

This chapter presents the proposed universal lesion segmentation framework leverages advanced deep learning techniques to improve the accuracy and efficiency of identifying and segmenting lesions in thoracic-abdominal computed tomography (CT) scans. Through meticulous experimental studies and comprehensive evaluations, this chapter demonstrates the efficacy and robustness of the designed framework. The results highlight significant advancements over existing methods, underscoring the potential of these techniques to enhance clinical workflows and diagnostic outcomes. The detailed analysis provided in this chapter offers insights into the performance metrics, comparative analyses of our framework with other relevant related works.

4.1 Proposed Framework

Our universal lesion segmentation framework is designed to enhance the accuracy and efficiency of identifying and segmenting universal lesions for thoracic-abdominal. As illustrated in Figure 4.1, the proposed framework consists of the following steps:

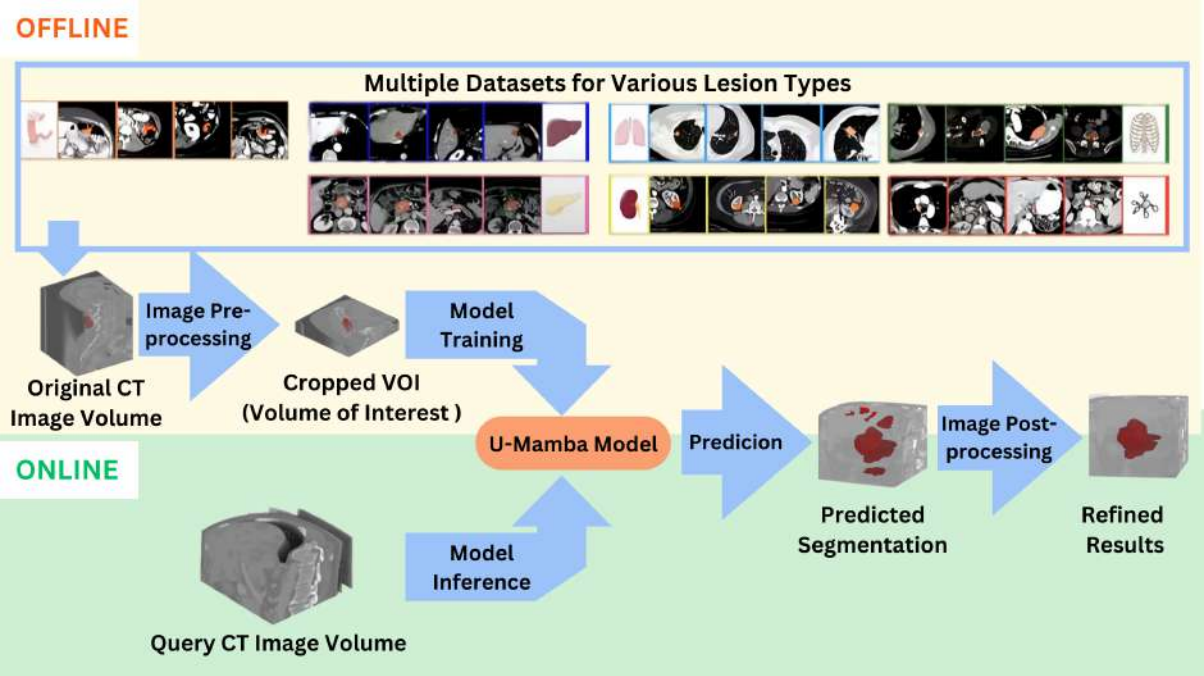

Figure 4.1: The Proposed Framework for Thoracic-Abdominal Universal Lesion Segmentation.

4.1.1 Image Preprocessing

To prepare the volume for inference, we preprocessed the volume through the following steps:

1. Lesion Selection: For an input 3D CT image volume, a specific slice is firstly targeted. Then, the region of interest is selected. This initial selection is crucial for accurately targeting the 2D lesion area for the next phase.

2. The volume of Interest (VOI) Cropping: Once the 2D region of interest is initialised from the target slice, a VOI is cropped around the selected area including all the depth slices from the input 3D CT image volume. This specific cropping ensures that even the largest lesions are entirely encapsulated within the VOI, providing a focused and detailed view of the area for subsequent analysis.

4.1.2 Lesion Segmentation Model

The cropped VOI is then processed using our segmentation model. The model runs inference on this volume to produce segmentation results, which highlight the boundaries and extent of the lesion. For this task, we used the U-Mamba model leveraging the synergy between Convolutional Neural Networks (CNNs) and State Space Models SSMs [93].

The U-Mamba model is trained from scratch using the LION (EvoLved Sign Momentum) optimizer [94] to accurately segment multiple lesions in the thoracic-abdominal

regions. To overcome the challenge of limited annotated data, our methodology integrates various CT image datasets, selected from recent research, each designed for segmenting specific organ lesions. These encompass bone, liver, pancreas, lung, lymph node, and colon lesions (detailed in Chapter 3, Sections 1 to 7).

Following, we detail the selected U-Mamba model architecture as well as the used optimizer in our proposed universal lesion segmentation

U-Mamba Model Architecture

U-Mamba architecture is designed specifically for biomedical image segmentation [93], U-Mamba represents a significant advancement in the field, offering superior performance and adaptability across various datasets.

Selecting U-Mamba for universal lesion segmentation in CT scans offers multiple advantages over traditional U-Net-based methods. Its ability to capture long-range dependencies, self-configuring mechanism, superior performance across diverse tasks, computational efficiency, and advanced optimization techniques collectively make it a powerful and reliable tool for biomedical image segmentation. These attributes ensure that U-Mamba can handle the complexities of lesion segmentation in CT scans more effectively, providing more accurate and robust segmentation results [93].

The overall network structure of U-Mamba is illustrated in Figure 4.2 (b). the model network is based on an encoder-decoder architecture as follows:

- **Encoder**: The encoder consists of multiple layers of U-Mamba blocks that capture long-range dependencies and hierarchical features from the input images.
- **Bottleneck**: The deepest layer employs U-Mamba blocks to consolidate the most abstract features.
- **Decoder**: The decoder reconstructs the high-resolution segmentation maps using Residual blocks. These blocks recover the spatial details through transposed convolutions.
- **Skip Connections**: Skip connections are utilized to transfer features directly from the encoder to the decoder, preserving spatial information and facilitating better gradient flow.

The overall network structure consists of several U-Mamba blocks (Figure 4.2 (a)). U-Mamba block is the core component of the architecture. It efficiently captures both local and long-range dependencies through the following steps:

- **Convolution**: Two convolutional layers with Instance Normalization and Leaky ReLU activation are applied to the input tensor of shape (B, C, H, W, D). Where B is the Batch size, representing the number of images processed together in one forward/backward pass, C is the Number of channels. H is the Height of the image, referring to the number of pixels in the vertical dimension and W is the

Width of the image, referring to the number of pixels in the horizontal dimension. D is the Depth, which is relevant in 3D medical imaging like CT or MRI scans, representing the number of slices or layers in the volume.

— **Normalization and Flattening**: The output from the convolutional layers is normalized using layer normalization, then flattened, changing its shape from (B, C, H, W, D) to (B, L, C), where $L = H \times W \times D$.

— **Dual Branch Processing**:
  — **Linear Branch**: The input tensor is processed through a linear layer followed by SiLU activation.
  — **SSM Branch**: Simultaneously, the input is processed through a 1D convolution layer, then a Structured State Space Model (SSM) layer, followed by a linear transformation and the SiLU activation function.

— **Merging**: The outputs from both branches are merged using the Hadamard product.

— **Reshape**: The merged output is reshaped back to the original tensor shape (B, C, H, W, D).

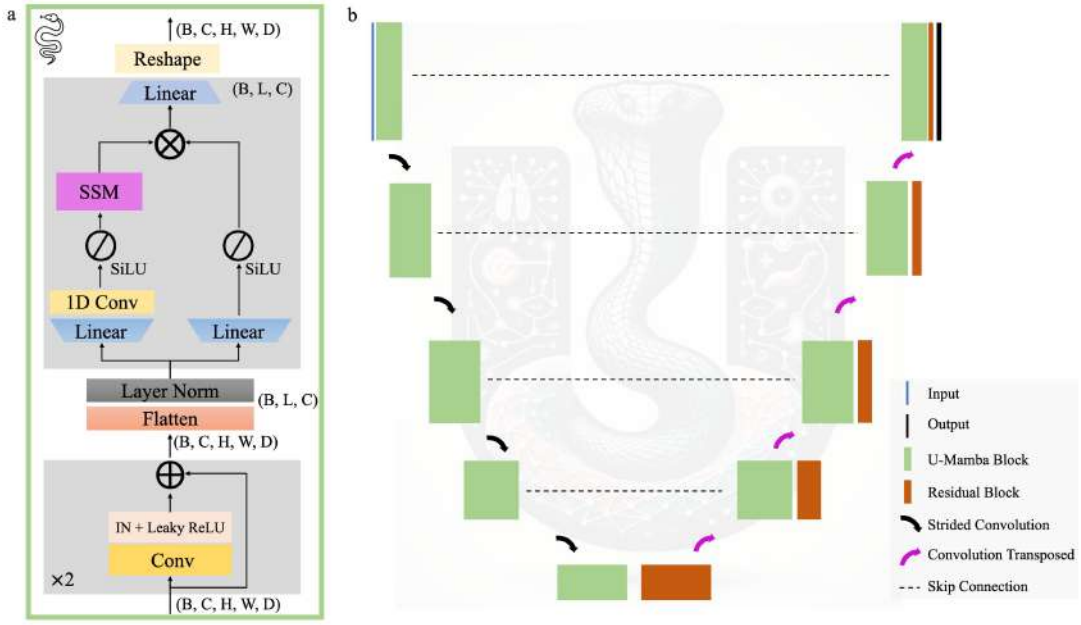

Figure 4.2: (a) U-Mamba Block architecture. (b) U-Mamba network structure with encoder-decoder design, incorporating U-Mamba blocks and Residual blocks.

LION Optimizer

The LION (EvoLved Sign Momentum) optimizer [94], has recently achieved first place ImagNet challenge 2022. This optimizer is based optimization algorithm discovered through symbolic program search, pioneered by Google researchers. It differentiates itself from traditional adaptive optimizers by focusing solely on momentum tracking and using

the sign operation to compute updates. This design choice leads to several key strengths:

— **Memory Efficiency**: LION is more memory-efficient compared to optimizers like Adam, as it only tracks momentum without maintaining second moments.

— **Uniform Update Magnitude**: By applying the sign function, LION ensures that update magnitudes are uniform across all parameters, which can lead to more stable training dynamics.

— **Performance Gains**: Empirical results demonstrate that LION significantly boosts the performance of various models. For instance, it improves the accuracy of ViT by up to 2% on ImageNet [95] and reduces pre-training compute requirements by up to 5x.

— **Adaptability**: LION shows superior performance in different training scenarios, including vision-language contrastive learning and diffusion models, where it achieves better results with reduced computational cost.

In this research study, we aim to leverage LION optimizer strengths to enhance training performance and model accuracy for volumetric image segmentation tasks. This novel approach not only addresses the limitations of existing optimizers but also sets a new benchmark in optimization techniques for deep network architecture in image segmentation.

4.1.3 Segmentation Post-Processing

The segmentation results obtained from the U-Mamba model are further post-processed to enhance clarity and usability. This step involves keeping only the central lesion and its connected parts while discarding the rest of the segmented regions. This approach is based on the clinical assumption that the largest and most central lesion is likely to be the primary region of interest, whereas smaller, disconnected segments are often less relevant lesions. The proposed post-processing phase consists of the following steps:

1. **Identify the Central Lesion:** The central lesion is identified based on its size and connectivity. This involves selecting the largest connected component from the initial prediction.
2. **Connected Component Analysis:** Using connected component analysis, we ensure that only the segments connected to the central lesion are retained.
3. **Discard Irrelevant Parts:** Smaller, disconnected regions that do not connect to the central lesion are discarded, reducing the number of false positives.

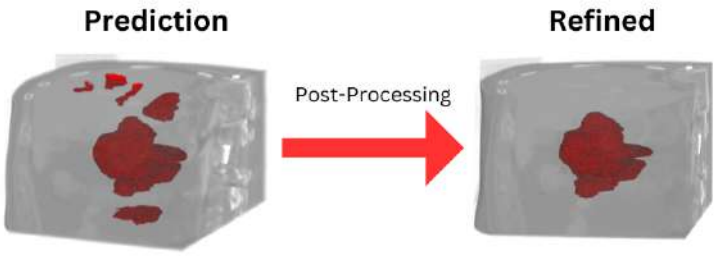

Figure 4.3: The effect of post-processing on the predicted lesion segmentation. The left image shows the initial prediction with multiple disconnected regions. The right image shows the refined segmentation, where only the central lesion and its connected parts are retained.

The refined segmentation is more accurate and clinically relevant, ensuring that the most significant lesions are correctly identified and analyzed. As we can clearly observe in Figure 4.3, The proposed post-processing phase significantly improve the quality of the final segmentation result. This post-processing step is a crucial part of our lesion segmentation pipeline, enhancing the overall performance and reliability of the model's predictions.

This comprehensive pipeline, as illustrated in Figure 4.1, ensures that lesions are accurately identified, segmented, and presented for clinical use, thereby enhancing diagnostic capabilities and patient outcomes.

4.2 Experimental Study

In this section we present the experimental study for evaluating our proposed universal lesion segmentation framework. Starting by the used datasets for training and validation, the used evaluation metrics and we discuss the obtained experimental results.

4.2.1 Dataset Description

To evaluate our proposed framework for universal lesion segmentation in the thoracic-abdominal area, we train our framework on seven CT image volume datasets.

ULS23 Dataset: Universal Lesion Segmentation Challenge (ULS23) dataset is a novel realised dataset that comprises a diverse collection of annotated lesion volumes-of-interest (VOIs) sourced from multiple datasets including 750 lesions from the DeepLesion dataset, 744 bone lesions and 124 pancreas lesions from the Radboudumc dataset, curated by Natália Alves, Megan Schuurmans, and Henkjan Huisman.

These annotations are available through the Challenge repository on GitHub [1], offering researchers a comprehensive resource for training and evaluating segmentation algorithms

1. https://github.com/MJJdG/ULS23

across a variety of lesion types. This dataset is pivotal for advancing lesion segmentation research, supporting the development of robust models capable of accurately identifying and delineating lesions in medical imaging data [96].

KiTS21 Dataset: The Kidney Tumor Segmentation Challenge 2021 (KiTS21) dataset comprises 300 high-resolution abdominal CT scans with detailed manual segmentation of kidneys and kidney tumors by expert radiologists. This dataset aims to advance automated segmentation algorithms, providing a critical resource for the development and evaluation of methods to accurately identify and delineate renal structures and tumors, which are essential for diagnosis, treatment planning, and monitoring in clinical settings. [97].

LIDC-IDRI Dataset: The Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) dataset is a comprehensive collection of 1,018 thoracic CT scans annotated with detailed manual annotations of lung nodules by up to four expert radiologists. These annotations include the nodules' location, characteristics, and likelihood of malignancy, supporting the development of computer-aided detection and diagnosis systems. The dataset serves as a benchmark for evaluating lung nodule detection algorithms, contributing significantly to early lung cancer detection and improving clinical outcomes [98].

LiTS Dataset: The Liver Tumor Segmentation (LiTS) dataset includes 888 contrast-enhanced abdominal CT scans with precise manual segmentation of the liver and liver tumors. Created for the LiTS Challenge, this dataset supports the development of automated segmentation algorithms, which are vital for accurate diagnosis, treatment planning, and monitoring of liver diseases. The high-quality annotations and diverse imaging conditions make the LiTS dataset a key resource for bench-marking and advancing liver and tumor segmentation methods in medical imaging research [99].

MDSC Datasets (Task 6/7/10): The Medical Decathlon Segmentation Challenge (MDSC) datasets, specifically Task 6 (Lung), Task 7 (Pancreas), and Task 10 (Colon), provide a comprehensive collection of annotated CT scans to support the development and bench marking of automated segmentation algorithms for various organs and tumors. Task 6 focuses on lung tumor segmentation from thoracic CT scans, offering detailed annotations to advance early detection and treatment of lung cancer. Task 7 is dedicated to the segmentation of the pancreas and pancreatic tumors from abdominal CT scans, addressing the critical need for precise diagnosis and intervention in pancreatic diseases. Task 10 targets the segmentation of colon tumors from CT scans, contributing to improved diagnosis and treatment planning in colorectal cancer. Together, these datasets play a crucial role in enhancing the accuracy and reliability of medical image analysis algorithms, ultimately improving clinical outcomes across different types of cancer (Lung, Pancreas, and Colon) [100].

NIH-LN Dataset: The National Institutes of Health-Lymph Node (NIH-LN) dataset

includes comprehensive annotations for 558 abdominal lymph nodes and 379 mediastinal lymph nodes, significantly enriching the training data with precise and detailed lymph node lesion annotations. Developed by the National Institutes of Health (NIH), a leading medical research agency of the United States government, this dataset benefits from the expertise of experienced radiologists who meticulously performed the annotations, ensuring high accuracy and reliability. By providing detailed information on the exact locations and boundaries of these lymph nodes, the NIH-LN dataset supports the development and validation of advanced detection and segmentation algorithms. This enhanced training data is crucial for improving the accuracy of automated systems in identifying lymph node lesions, which is vital for the diagnosis, staging, and treatment planning of various cancers and other diseases affecting the lymphatic system [101].

The intensity statistics of the datasets used in this study are summarized in Table 4.1. These statistics include the maximum, minimum, mean, standard deviation of the image intensity, as well as the median intensity value. This information is crucial for understanding the normalization requirements and the dynamic range of the data, which can significantly affect the performance of the segmentation model.

Tableau 4.1: Dataset Intensity Statistics.

Dataset	**Max**	**Min**	**Mean**	**Std**	**Min%**	**Max%**	**Median**
ULS23_DeepLesion3D	2403.0	-1177.0	101.11	219.39	-816.0	1219.0	67.0
ULS23_Radboudumc_Bone	2782.0	-935.0	1256.61	552.61	31.0	2399.0	1276.0
ULS23_Radboudumc_Pancreas	1169.0	-1008.0	66.96	40.90	-52.0	182.0	68.0
kits21-master	3071.0	-1022.0	65.42	60.91	-52.0	288.0	58.0
LIDC-IDRI	2498.0	-1380.0	-226.75	358.50	-956.0	1327.24	-168.0
LiTS	4016.0	-1016.0	60.15	47.68	-82.0	167.0	60.0
LNDb	2746.0	-1411.0	-253.69	382.03	-1037.97	1318.0	-144.0
MDSC/Task06_Lung	2671.0	-1024.0	-158.39	324.82	-1024.0	327.0	-2.0
MDSC/Task07_Pancreas	3071.0	-520.0	71.30	57.57	-43.0	195.0	72.0
MDSC/Task10_Colon	445.0	-848.0	65.17	32.65	-31.0	165.0	65.0
NIH_LN/ABD	1926.0	-743.0	54.15	36.60	-63.0	148.0	55.0
NIH_LN/MED	3071.0	-1024.0	48.06	63.73	-325.0	190.0	53.0

4.2.2 Evaluation Metrics

To comprehensively assess the performance of our proposed universal lesion segmentation framework for thoracic-abdominal computed tomography scans, we employ a diverse set of evaluation metrics. These metrics provide a multifaceted analysis of the model's effectiveness across various lesion types and sizes. The selected metrics are as follows:

— **Overlap-based metrics:** including Dice Coefficient (DSC) and Intersection over Union (IoU), offering insights into the spatial accuracy of the segmentation.

1. Source: https://uls23.grand-challenge.org/evaluation/test-phase-leaderboard/leaderboard/

— **Pixel-wise metrics:** (Recall, Precision, and Accuracy) provide a detailed understanding of the model's performance at the pixel level.
— **Clinical relevance:** is addressed through the Long Axis Error (LAE), which directly relates to measurements used by radiologists in clinical practice.

By employing this diverse set of metrics, we aim to provide a thorough and clinically relevant evaluation of our Universal lesion segmentation framework. This approach allows for a nuanced understanding of the model's strengths and limitations across various lesion types and sizes, facilitating meaningful comparisons with existing methods and guiding future improvements. For more information and a detailed description of the used evaluation metrics, we refer readers to Chapter 2, Section 5.

4.2.3 Experimental Setup

Throughout the development of our experimental study, we utilized several key tools and frameworks to facilitate efficient data processing and model training. Using the Hold-out evaluation protocol, each dataset is split on 90% for training where the remain 10% is used for test.

The training environment features Intel Core i7-12700F processor, offering 12 cores up to 4.8 GHz. Graphics processing is handled by an NVIDIA GeForce RTX 3080 Ti with NVIDIA 550.67 driver and CUDA support, along with OpenGL 4.6.0. Memory resources include 64 GiB RAM with 500 GB NVMe drive of storage.

The training of the U-Mamba model was conducted over 1000 epochs using leave-one-out cross-validation. Due to the size of our dataset, data augmentation techniques were not needed. After several experiments, to balance memory constraints and computational efficiency while ensuring the model's capacity to learn complex features from the volumetric data, the model parameters were chosen as follows:

— **Batch Size:** 6
— **Patch Size:** [10, 256, 256]
— **Initial Learning Rate:** 0.01
— **Weight Decay:** 3e-5

4.2.4 Experimental Results

Several experiments are conducted to evaluate our proposed universal lesion segmentation framework. Firstly, we start with the U-Mamba network training evaluation, Then, we compare our trained U-Mamba model using Lion optimizer with the training results using standard SGD optimizer. Next, extensive evaluations are connected based on several evaluation metrics. Finally, the final segmentation results are compared with the most relevant segmentation architectures and related works.

4.2.5 Model Training and Inference

During the training process, we store both the final and best checkpoints. In five-fold cross-validation, each data point serves as a separate validation set, allowing for a rigorous assessment of model performance. As training progresses, checkpoints are saved periodically to capture the model's state at different iterations. The best checkpoint is determined by evaluating the validation loss across all folds, ensuring that the selected model represents the optimal performance achieved during training. This meticulous selection process guarantees that the model chosen for inference demonstrates robustness across diverse validation scenarios. Additionally, the final checkpoint captures the model's state after training, providing insight into its overall convergence and performance.

In addition to five-fold cross-validation, we reserve 10% of each dataset for independent testing purposes. This subset, comprising approximately 1020 volumes, is carefully selected to ensure an equitable representation of various lesion types from the different used datasets. Despite the inherent imbalances in lesion distribution within the dataset, such as some types being more prevalent than others, we ensure each lesion type is proportionally represented within the testing subset. This deliberate allocation allows us to evaluate the performance of our best checkpoint on previously unseen instances, providing a comprehensive assessment of its generalization capabilities across different lesion types. By withholding this portion from the training data, we mitigate the risk of overfitting and obtain a more accurate measure of the model's effectiveness in real-world scenarios. Figure 4.4 displays the duration of each epoch and the corresponding learning rate schedule throughout the training process.

Furthermore, upon completion of model training, we conduct inference on the test dataset. The inference process involves applying the trained model to the test data to make predictions on lesion segmentation. Utilizing our hardware infrastructure, the inference time per volume is measured at 3 seconds. This information offers insights into the computational efficiency of our model in processing volumetric data, crucial for scaling up to real-world applications where rapid and accurate inference is paramount.

4.2.6 Optimizer Comparison

In this experiment, to prove the efficacity of the LION optimizer in medical lesion segmentation, we compare the performance of two different optimizers: the well-used standard SGD optimizer with the LION optimizer. Figure 4.5 illustrates the comparison results of the U-Mamba model training and validation in terms of losses and pseudo-dice metrics over 1000 epochs.

In terms of training and validation loss, the U-Mamba model with LION optimizer achieves lower results compared to the U-Mamba model with SGD optimizer, indicating better convergence and generalization. Also, from Figure 4.5, we can clearly observe that

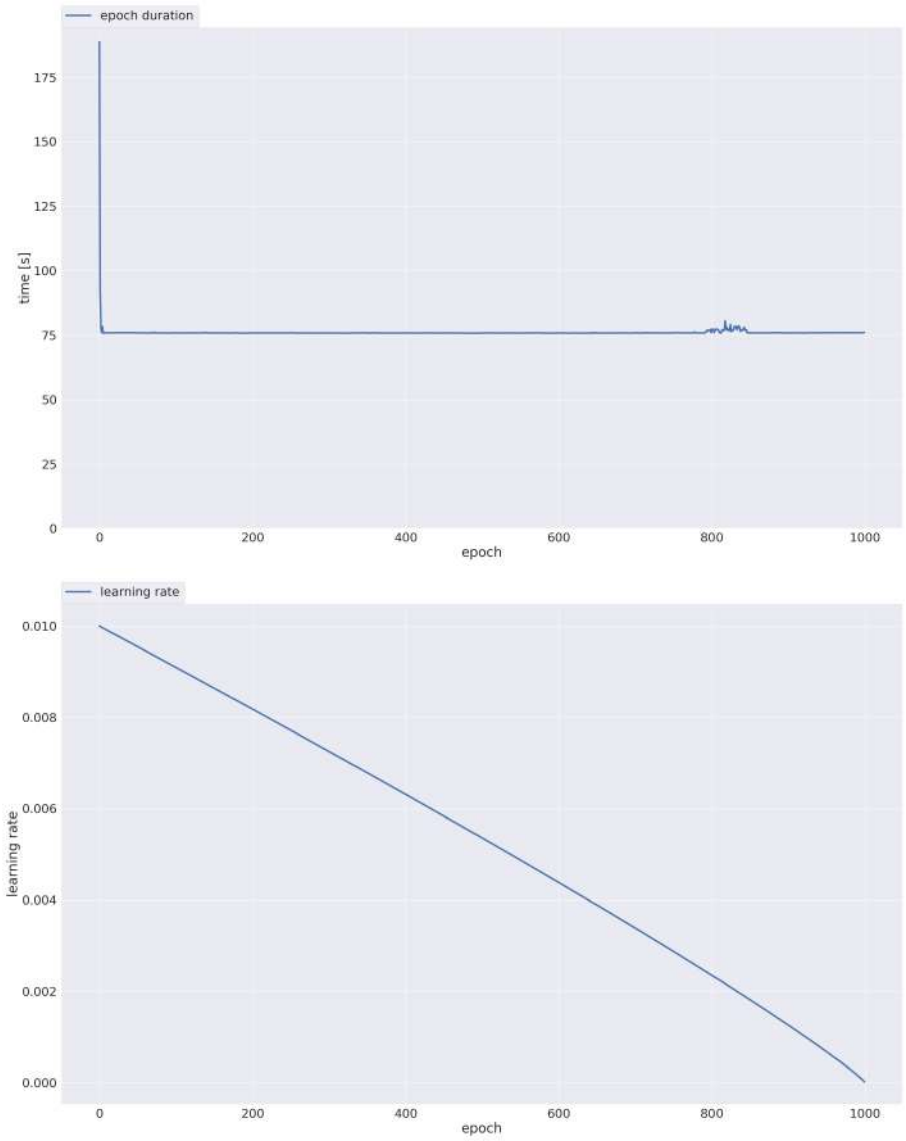

Figure 4.4: Epoch duration and learning rate schedule.

the U-Mamba model with LION optimizer achieves higher pseudo-dice scores, demonstrating superior segmentation performance. Additionally, the pseudo-dice scores for the U-Mamba model with LION optimizer exhibit less variability during training, indicating more stable and consistent performance.

The comparison result proves that integrating the LION optimizer with the U-Mamba model training outperforms the U-Mamba model with standard SGD optimizer in all measured aspects, providing enhanced convergence, more accurate segmentation, and more stable training compared to the U-Mamba model with SGD.

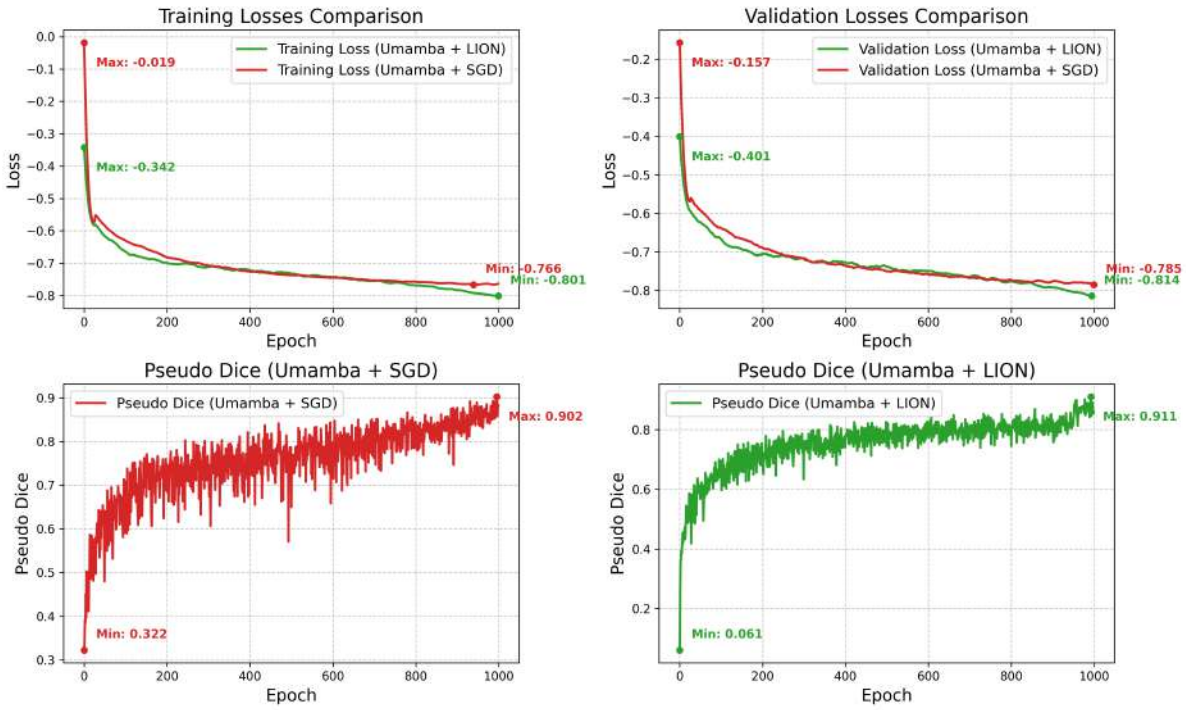

Figure 4.5: Comparison of training and validation losses, and Pseudo Dice scores for U-Mamba with SGD optimizer (red) and U-Mamba with LION optimiser (green) over 1000 epochs.

4.2.7 Framework Performances

In this section, we present a comprehensive evaluation of our universal lesion segmentation framework across various lesion types. The performance metrics include the Confusion Matrix, the Dice coefficient and Long Axis Error (LAE), Intersection over Union (IoU), Recall, Precision, and Accuracy. These metrics are critical for understanding the model's effectiveness and reliability in segmenting lesions accurately.

Below, Figure 4.6 shows the confusion matrix, which provides a global overview of the model's predictions by comparing the predicted segmentation results to the ground truth.

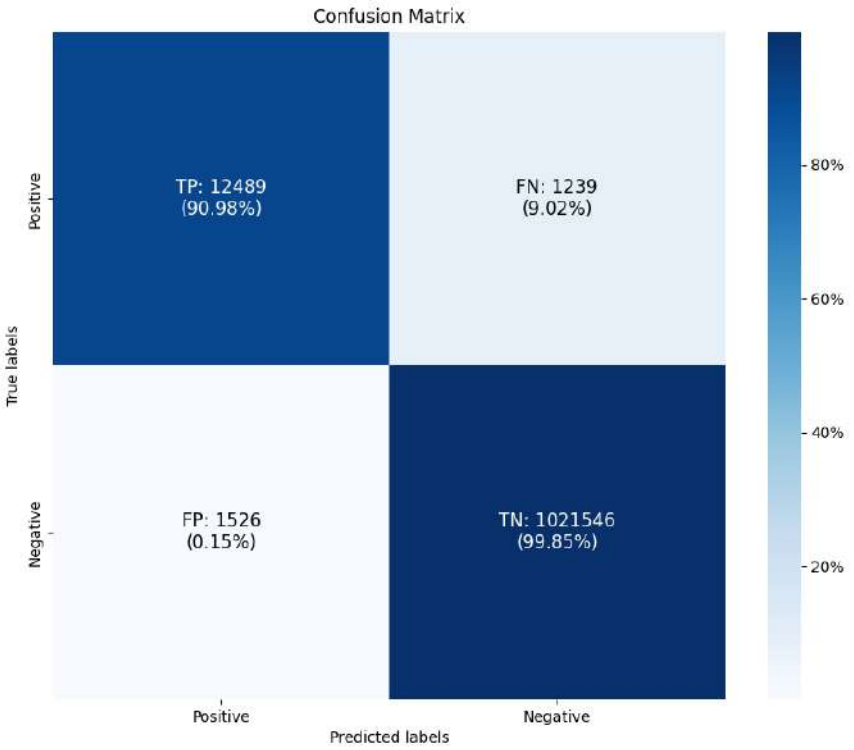

Figure 4.6: Confusion matrix for the universal lesion segmentation framework.

Figure 4.7 presents scatter plots of Dice coefficients versus Long Axis Error (LAE) for various lesion types, including mixed (Deeplesion 3D), kidney, lung, liver, pancreas, lymph nodes, colon, and bone. Each subplot includes the average metrics for the respective lesion type.

The scatter plots depict individual test samples, with the vertical axis representing the Long Axis Error (percentage) and the horizontal axis representing the Dice coefficient (percentage). The horizontal dashed line at LAE = 10% highlights the threshold for acceptable geometric error. Most data points are clustered in the lower right quadrant, indicating high Dice scores and low geometric errors, which reflects the model's high performance across different lesion types.

The evaluation metrics demonstrate the robustness and accuracy of our lesion segmentation model across various lesion types. The high Dice coefficients and IoU values, coupled with low Long Axis Errors, suggest that the model performs well in identifying and accurately segmenting lesions. This detailed evaluation provides confidence in the model's applicability in clinical settings.

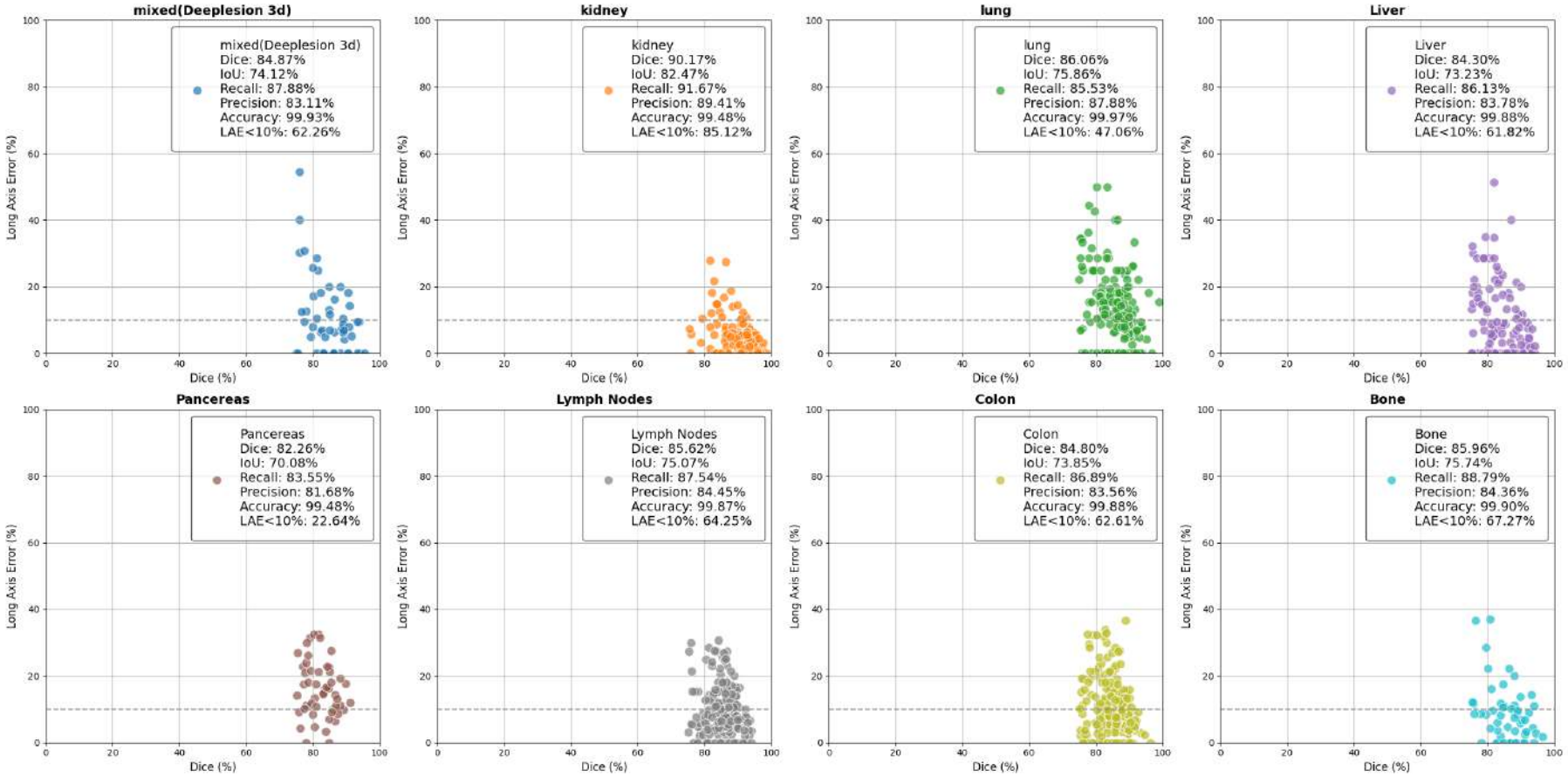

Figure 4.7: Scatter plots of Dice coefficients versus Long Axis Error (LAE) for various lesion types. Each subplot includes the average metrics for the respective lesion type.

Tableau 4.2: Comparison of Dice Similarity Coefficients (DSC) for Various Models Across Different Lesion Types

Methods	Bone	Kidney	Colon	Lymph Nodes	Lung	Liver	Pancreas
	DSC	DSC	DSC	DSC	DSC	DSC	DSC
nnUNet	0.6957±0.1940	0.8306±0.1199	0.7700±0.1326	0.7700±0.1326	0.8017±0.1182	0.7373±0.1607	0.7386±0.0985
U-Mamba+SGD	0.7637±0.1500	0.8839±0.0660	0.7892±0.1235	0.7892±0.1235	0.8213±0.1054	0.7606±0.1545	0.7601±0.0948
U-Mamba+LION	**0.7852±0.1501**	**0.8936±0.0571**	**0.8012±0.1168**	**0.8012±0.1168**	**0.8252±0.1008**	**0.7696±0.1438**	**0.7753±0.0810**

4.2.8 Comparative Analysis

In this section, we present a comparative analysis of different models based on the evaluation metrics for lesion segmentation. The analysis includes a visual comparison of segmentation results and a detailed table of Dice scores for various lesion types across different organs.

Figure 4.8 shows the segmentation results from different models across various lesion types. The first row represents the Ground Truth annotations, while the subsequent rows represent the nnUNet+SGD, U-Mamba+SGD, and U-Mamba+LION models. Each column represents a different lesion type within an organ (Bone, Kidney, Lymph Nodes, Lung, Colon, Liver, Pancreas). The red contours indicate the predicted lesion boundaries.

The visual comparison indicates that U-Mamba+LION produces more precise and smoother lesion boundaries, which are closer to the ground truth annotations demonstrating superior performance in terms of lesion segmentation accuracy compared to nnUNet and U-Mamba+SGD.

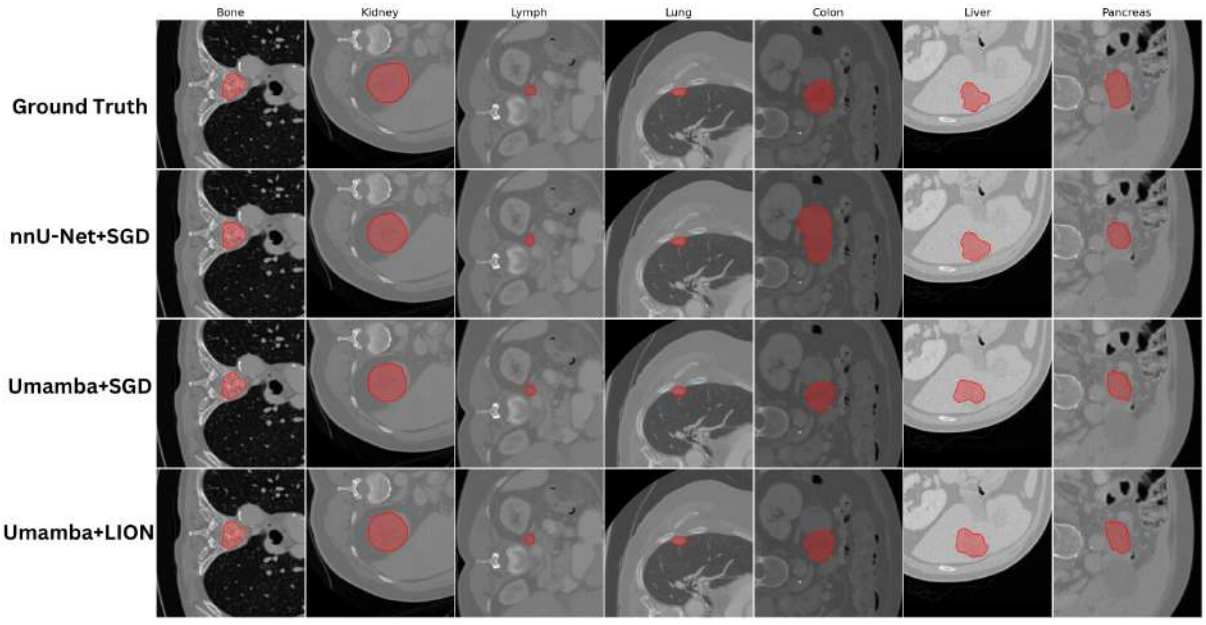

Figure 4.8: Illustration of segmentation results by different models. The rows represent Ground Truth, nnUNet+SGD, U-Mamba+SGD, and U-Mamba+LION, while the columns represent different lesion types in various organs (Bone, Kidney, Lymph Nodes, Lung, Colon, Liver, Pancreas).

Table 4.2 presents the Dice Scores for different models and lesion types. The Dice Similarity Coefficient (DSC) is reported for each lesion type. The results show that the U-Mamba+LION model consistently achieves higher Dice scores across all lesion types compared to nnUNet and U-Mamba with standard SGD optimizer. The improvements are particularly noticeable in challenging cases such as bone and pancreas lesion segmentation.

Next, compared to most relevant related works, our model outperforms all algorithms from the ULS23 challenge. As summarized in Table 4.3, our model achieves a Dice Score of 0.807 ± 0.109, significantly higher than the first-place winner model, which achieved a score of 0.708 ± 0.235. This substantial improvement highlights the effectiveness of our proposed framework in universal lesion segmentation.

Tableau 4.3: Dice Scores from the ULS23 Challenge Leaderboard [2].

Rank	**Algorithm**	**Dice Score**
11	Maiol-UPF team models	0.433 ± 0.307
10	AIMI_Team6	0.566 ± 0.291
9	ULS_JHU	0.677 ± 0.258
8	One Click Lesion Segmentation	0.690 ± 0.272
7	Baseline AIMI Group 12	0.691 ± 0.249
6	Finetuning The Official Baseline	0.694 ± 0.235
5	Universal Lesion Segmentation [ULS23 Baseline]	0.703 ± 0.240
4	CW nnUnet	0.703 ± 0.240
3	DeepMax	0.704 ± 0.223
2	Junma3	0.708 ± 0.235
1	**Our Model**	**0.807 ± 0.109**

the comparative analysis demonstrates that the U-Mamba+LION optimizer significantly enhances the performance of lesion segmentation models. The visual and quantitative results confirm that this model is more accurate and reliable for clinical applications.

2. Source: https://uls23.grand-challenge.org/evaluation/test-phase-leaderboard/leaderboard/
3. Source: https://uls23.grand-challenge.org/evaluation/test-phase-leaderboard/leaderboard/

Conclusion

In this chapter, we have introduced and rigorously evaluated our universal lesion segmentation framework, which incorporates the U-Mamba model with the LION optimizer. Through extensive experimental studies and comparisons, our model demonstrated superior performance across various lesion types and sizes in thoracic-abdominal computed tomography scans.

The key contributions of this research include the use of the U-Mamba model, a robust architecture designed for accurate lesion segmentation. This model incorporates innovative techniques such as encoder-decoder design and residual blocks. Furthermore, by optimizing with LION, we established that this optimizer significantly outperforms the standard SGD optimizer, providing enhanced convergence, more accurate segmentation, and more stable training. To overcome the challenge of limited annotated data, our model training methodology integrates various CT image datasets, selected from recent research, each designed for segmenting specific organ lesions. These encompass bone, liver, pancreas, lung, lymph node, and colon lesions.

To ensure a comprehensive evaluation, we utilized a diverse set of metrics, including Dice Coefficient, Intersection over Union (IoU), Recall, Precision, Accuracy, and Long Axis Error (LAE). This provided a multifaceted and clinically relevant assessment of our model's performance. Additionally, by leveraging our hardware infrastructure and meticulous training protocols, we ensured the model's robustness and computational efficiency, achieving inference times as low as three seconds per volume.

Our results indicate that the U-Mamba model with LION optimizer not only achieves higher accuracy and consistency in lesion segmentation but also generalizes well across different lesion types. This makes it a valuable tool for real-world medical imaging applications. This framework paves the way for future research and improvements in automated lesion segmentation, potentially aiding radiologists in clinical decision-making and enhancing patient outcomes.

General Conclusion

Medical imaging plays a crucial role in the diagnosis and treatment of various diseases. With the advanced imaging technologies, there has been a significant increase in the volume of medical scans generated worldwide. Accurate and efficient analysis of these scans is essential for effective clinical decision-making. Traditional methods of medical image analysis, which heavily rely on manual annotations by radiologists, are becoming increasingly impractical due to the volume of data and the inherent limitations of human capabilities.

Deep learning has emerged as a powerful tool in medical image analysis, offering the potential to automate and enhance the accuracy of lesion segmentation. Lesion segmentation is a critical task in medical imaging, as it helps in identifying and delineating pathological regions within the body. Accurate segmentation is particularly important for the diagnosis and treatment planning of various diseases, including cancer.

In this Master thesis, we have designed and evaluated a universal lesion segmentation framework that is capable of segmenting multiple types of lesions in the thoracic-abdominal area. This work has encompassed the design, implementation, and thorough evaluation of our proposed universal lesion segmentation framework including the use the U-Mamba, optimized using the LION optimizer. The key contributions and findings of this research are summarized as follows:

Summary of Contributions

— We proposed a universal lesion segmentation framework that can effectively segment various types of lesions in thoracic-abdominal CT scans, including bone, liver, pancreas, lung, lymph node, and colon lesions.

— The U-Mamba model was exploited as robust architecture incorporating innovative techniques such as encoder-decoder design and residual blocks to enhance segmentation accuracy.

— The LION optimizer was demonstrated to significantly outperform the standard

SGD optimizer, providing better convergence, more accurate segmentation, and more stable training for segmentation task.

— A comprehensive evaluation was conducted using a diverse set of metrics, including Dice Coefficient, Intersection over Union (IoU), Recall, Precision, Accuracy, and Long Axis Error (LAE), to assess the model's performance in a clinically relevant manner.

— Our model was trained using an integrated dataset from various CT image datasets, which included annotated data for specific organ lesions, overcoming the challenge of limited annotated data.

— The U-Mamba model with LION optimizer achieved higher accuracy and consistency in lesion segmentation compared to existing models, as evidenced by its superior performance in the ULS23 challenge leaderboard.

— The framework demonstrated robust computational efficiency, with inference times as low as three seconds per volume, making it suitable for real-world clinical applications.

Future Work

While this research has made significant strides in the field of automated universal lesion segmentation, there are several areas for future exploration:

— **Expanding the Dataset:** Incorporating more diverse and extensive datasets could further enhance the model's generalizability and robustness, particularly for rare or less common lesion types.

— **Improving Model Architecture:** Further refinement of the model architecture, such as incorporating advanced deep learning techniques and exploring other novel optimizers, could lead to even better segmentation performance.

— **Explainability and Interpretability:** Enhancing the explainability and interpretability of the segmentation results would help gain trust from clinicians and improve the overall usability of the model.

— **Longitudinal Studies:** Conducting longitudinal studies to assess the model's performance over time and its impact on patient outcomes would provide valuable insights into its clinical effectiveness.

Bibliography

[1] Giles W Boland, Alexandre S Guimaraes, and Peter R Mueller. The radiologist's conundrum: benefits and costs of increasing ct capacity and utilization. *European radiology*, 19:9–11, 2009.

[2] Robert J McDonald, Kevin M Schwartz, Lawrence J Eckel, et al. The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. *Academic radiology*, 22(9):1191–1198, 2015.

[3] Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, et al. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 71(3):209–249, 2021.

[4] Elizabeth A Eisenhauer, Patrick Therasse, Jan Bogaerts, et al. New response evaluation criteria in solid tumours: revised recist guideline (version 1.1). *European journal of cancer*, 45(2):228–247, 2009.

[5] Andreas Hering, Florian Peisen, Teresa Amaral, et al. Whole-body soft-tissue lesion tracking and segmentation in longitudinal ct imaging studies. In *Medical Imaging with Deep Learning*, pages 312–326. PMLR, 2021.

[6] Jingru Cai, Yifan Tang, Le Lu, et al. Accurate weakly-supervised deep lesion segmentation using large-scale clinical annotations: Slice-propagated 3d mask generation from 2d recist. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV*, pages 396–404. Springer International Publishing, 2018.

[7] Yifan Tang, Ke Yan, Jing Xiao, et al. One click lesion recist measurement and segmentation on ct scans. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV*, pages 573–583. Springer International Publishing, 2020.

[8] Yifan Tang, Ke Yan, Jingru Cai, et al. Lesion segmentation and recist diameter prediction via click-driven attention and dual-path connection. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II*, pages 341–351. Springer International Publishing, 2021.

[9] Ke Yan, Xiaohuan Wang, Le Lu, et al. Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *Journal of medical imaging*, 5(3):036501–036501, 2018.

[10] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, volume 25, 2012.

[11] Marvin Minsky and Seymour Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, 1969.

[12] Justin Ker, Lipo Wang, Jai Rao, and Tchoyoson Lim. Deep learning applications in medical image analysis. *Ieee Access*, 6:9375–9389, 2017.

[13] Mariusz Bojarski, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D. Jackel, Mathew Monfort, Urs Muller, Jiakai Zhang, Xin Zhang, Jake Zhao, and Karol Zieba. End to end learning for self-driving cars, 2016.

[14] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

[15] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT press, 2016.

[16] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer, 2010.

[17] Ning Qian. On the momentum term in gradient descent learning algorithms. *Neural networks*, 12(1):145–151, 1999.

[18] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. In *Proceedings of the 24th International Conference on Neural Information Processing Systems*, pages 257–265, 2011.

[19] Tijmen Tieleman and Geoffrey Hinton. *Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude*, volume 4. 2012.

[20] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[21] Adrian Remonda, Sarah Krebs, Eduardo Veas, Granit Luzhnica, and Roman Kern. Formula rl: Deep reinforcement learning for autonomous racing using telemetry data. *arXiv preprint arXiv:2104.11106*, 2021.

[22] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[23] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: Learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2016.

[24] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2):203–211, 2021.

[25] Ghislaine Scelo and Tricia L. Larose. Epidemiology and risk factors for kidney cancer. *Journal of Clinical Oncology*, 36(36):3574, 2018.

[26] Ryan D. Ward et al. 2017 aua renal mass and localized renal cancer guidelines: imaging implications. *Radiographics*, 38(7):2021–2033, 2018.

[27] Arpit Rao, Charles Wiggins, and Richard C. Lauer. Survival outcomes for advanced kidney cancer patients in the era of targeted therapies. *Annals of translational medicine*, 6(9), 2018.

[28] Alexander Kutikov and Robert G. Uzzo. The renal nephrometry score: a comprehensive standardized system for quantitating renal tumor size, location and depth. *The Journal of urology*, 182(3):844–853, 2009.

[29] Vincenzo Ficarra et al. Preoperative aspects and dimensions used for an anatomical (padua) classification of renal tumours in patients who are candidates for nephron-sparing surgery. *European urology*, 56(5):786–793, 2009.

[30] Matthew N. Simmons et al. Kidney tumor location measurement using the c index method. *The Journal of urology*, 183(5):1708–1713, 2010.

[31] Nicholas Heller et al. The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge. *Medical Image Analysis*, 67:101821, 2019.

[32] Alex Golts, Daniel Khapun, Daniel Shats, Yoel Shoshan, and Flora Gilboa-Solomon. An ensemble of 3d u-net based models for segmentation of kidney and masses in ct scans. In *International Challenge on Kidney and Kidney Tumor Segmentation*, pages 103–115. Springer, 2021.

[33] Qi Ming How and Hoi Leong Lee. Kidney and kidney tumour segmentation in ct images, 2022.

[34] Zhongchen Zhao, Huai Chen, and Lisheng Wang. A coarse-to-fine framework for the 2021 kidney and kidney tumor segmentation challenge. In *International Challenge on Kidney and Kidney Tumor Segmentation*, pages 53–58. Springer, 2021.

[35] Matej Gazda, Peter Bugata, Jakub Gazda, David Hubacek, David Jozef Hresko, and Peter Drotar. Mixup augmentation for kidney and kidney tumor segmentation. In *International Challenge on Kidney and Kidney Tumor Segmentation*, pages 90–97. Springer, 2021.

[36] Xiaoshuai Hou, Chunmei Xie, Fengyi Li, and Nan Yang. Cascaded semantic segmentation for kidney and tumor. 2019.

[37] Rebecca L. Siegel, Kimberly D. Miller, H. E. Fuchs, and Ahmedin Jemal. Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians*, 72(1):7–33, 2022.

[38] Vinod Cheppamkuzhi and Menaka Dharmaraj. Improved segmentation of pulmonary nodules using soft computing techniques with segnet and adversarial networks. *Applied Sciences*, 13(12):7281, 2023.

[39] Chandra Sekhara Rao Annavarapu, Samson Anosh Babu Parisapogu, Nikhil Varma Keetha, Praveen Kumar Donta, and Gurindapalli Rajita. A bi-fpn-based encoder–decoder model for lung nodule image segmentation. *Diagnostics*, 13(8):1406, 2023.

[40] Diganta Misra. Mish: A self regularized non-monotonic activation function. *arXiv preprint arXiv:1908.08681*, 2019.

[41] Muhammad Usman, Azka Rehman, Abdullah Shahid, Siddique Latif, Shi Sub Byon, Sung Hyun Kim, Tariq Mahmood Khan, and Yeong Gil Shin. Mesahanet: Multi-encoders based self-adaptive hard attention network with maximum intensity projections for lung nodule segmentation in ct scan. *arXiv preprint arXiv:2304.01576*, 2023.

[42] Xinying Zhang, Shanshan Kong, Yang Han, Baoshan Xie, and Chunfeng Liu. Lung nodule ct image segmentation model based on multiscale dense residual neural network. *Mathematics*, 11(6):1363, 2023.

[43] B.W. Stewart and C.P. Wild. *World Cancer Report 2014*. World Health Organization Technical Report. WHO Press, World Health Organization, Geneva, Switzerland, 2014.

[44] L.E. Hann, C.B. Winston, K.T. Brown, and T. Akhurst. Diagnostic imaging approaches and relationship to hepatobiliary cancer staging and therapy. *Seminars in Surgical Oncology*, 19:94–115, 2000.

[45] Heimann T et al. Comparison and evaluation of methods for liver segmentation from ct datasets. *IEEE Transactions on Medical Imaging*, 28(8):1251–1265, 2009.

[46] Elizabeth A. Eisenhauer, Patrick Therasse, Jan Bogaerts, Lawrence H. Schwartz, Daniel Sargent, Robert Ford, Janet Dancey, Stephen Arbuck, Susan Gwyther, Mary Mooney, et al. New response evaluation criteria in solid tumours: Revised recist guideline (version 1.1). *European Journal of Cancer*, 45(2):228–247, 2009.

[47] Shigeki Shiina, Kenji Sato, Ryo Tateishi, Makoto Shimizu, Hideaki Ohama, Tomohiko Hatanaka, Minoru Takawa, Hiroaki Nagamatsu, and Yoshito Imai. Percutaneous ablation for hepatocellular carcinoma: Comparison of various ablation techniques and surgery. *Canadian Journal of Gastroenterology and Hepatology*, 2018:1–9, 2018.

[48] Kathy S. Albain, Robert S. Swann, Valerie W. Rusch, Andrew T. Turrisi, Frances A. Shepherd, Carmen Smith, Ying Chen, Robert B. Livingston, Richard H. Feins, David R. Gandara, et al. Radiotherapy plus chemotherapy with or without surgical resection for stage iii non-small-cell lung cancer: A phase iii randomised controlled trial. *The Lancet*, 374(9687):379–386, 2009.

[49] Francesca Virdis, Ilaria Reccia, Salomone Di Saverio, Gregorio Tugnoli, Sharon Kwan, Jay Kumar, Jacopo Atzeni, and Mauro Podda. Clinical outcomes of primary arterial embolization in severe hepatic trauma: A systematic review. *Diagnostic and Interventional Imaging*, 100(2):65–75, 2019.

[50] Miroslav Todorov, Johannes C. Paetzold, Oliver Schoppe, Giles Tetteh, Sahil Shit, Vladimir Efremov, Kinga Todorov-Völgyi, Marco Düring, Martin Dichgans, Marie Piraud, et al. Machine learning analysis of whole mouse brain vasculature. *Nature Methods*, 17(4):442–449, 2020.

[51] Mehrdad Moghbel, Syamsiah Mashohor, Rozi Mahmud, and M. Iqbal Bin Saripan. Review of liver segmentation and computer assisted detection/diagnosis methods in computed tomography. *Artificial Intelligence Review*, pages 1–41, 2017.

[52] Weibin Liao, Yinghao Zhu, Xinyuan Wang, Cehngwei Pan, Yasha Wang, and Liantao Ma. Lightm-unet: Mamba assists in lightweight unet for medical image segmentation. *arXiv preprint arXiv:2403.05246*, 2024.

[53] Yue Zhang, Chengtao Peng, Liying Peng, Huimin Huang, Ruofeng Tong, Lanfen Lin, Jingsong Li, Yen-Wei Chen, Qingqing Chen, Hongjie Hu, et al. Multi-phase liver tumor segmentation with spatial aggregation and uncertain region inpainting. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*, pages 68–77. Springer, 2021.

[54] Qixin Hu, Yixiong Chen, Junfei Xiao, Shuwen Sun, Jieneng Chen, Alan Yuille, and Zongwei Zhou. Label-free liver tumor segmentation, 2023.

[55] Xiao Han. Automatic liver lesion segmentation using a deep convolutional neural network method. *arXiv preprint arXiv:1704.07239*, 2017.

[56] Grzegorz Chlebus, Andrea Schenk, Jan Hendrik Moltz, Bram van Ginneken, Horst Karl Hahn, and Hans Meine. Automatic liver tumor segmentation in ct with fully convolutional neural networks and object-based postprocessing. *Scientific reports*, 8(1):15497, 2018.

[57] E. Svensson, C.F. Christiansen, S.P. Ulrichsen, M.R. Rørth, and H.T. Sørensen. Survival after bone metastasis by primary cancer type: A danish population-based cohort study. *BMJ Open*, 7(e016022), 2017.

[58] G. Chu, P. Lo, B. Ramakrishna, H. Kim, D. Morris, J. Goldin, and M. Brown. Bone tumor segmentation on bone scans using context information and random forests. In *Springer International Publishing*, pages Cham, Switzerland, 2014.

[59] S.T.H. Peeters, E.J. Van Limbergen, L.E.L. Hendriks, and D. De Ruysscher. Radiation for oligometastatic lung cancer in the era of immunotherapy: What do we (need to) know? *Cancers*, 13:2132, 2021.

[60] K.L. Zeng, C.L. Tseng, H. Soliman, Y. Weiss, A. Sahgal, and S. Myrehaug. Stereotactic body radiotherapy (sbrt) for oligometastatic spine metastases: An overview. *Front. Oncol.*, 9:337, 2019.

[61] K.L. Spencer, J.M. van der Velden, E. Wong, E. Seravalli, A. Sahgal, E. Chow, J.J. Verlaan, H.M. Verkooijen, and Y.M. van der Linden. Systematic review of the role of stereotactic radiotherapy for bone metastases. *J. Natl. Cancer Inst.*, 111:1023–1032, 2019.

[62] M. Loi, J.J. Nuyttens, I. Desideri, D. Greto, and L. Livi. Single-fraction radiotherapy (sfrt) for bone metastases: Patient selection and perspectives. *Cancer Manag. Res.*, 11:9397–9408, 2019.

[63] D.A. Palma, R. Olson, S. Harrow, S. Gaede, A.V. Louie, C. Haasbeek, L. Mulroy, M. Lock, G.B. Rodrigues, B.P. Yaremko, et al. Stereotactic ablative radiotherapy for the comprehensive treatment of oligometastatic cancers: Long-term results of the sabr-comet phase ii randomized trial. *J. Clin. Oncol.*, 38:2830–2838, 2020.

[64] D. De Ruysscher, R. Wanders, A. van Baardwijk, A.M. Dingemans, B. Reymen, R. Houben, G. Bootsma, C. Pitz, L. van Eijsden, W. Geraedts, et al. Radical treatment of non-small-cell lung cancer patients with synchronous oligometastases: Long-term results of a prospective phase ii trial (nct01282450). *J. Thorac. Oncol.*, 7:1547–1555, 2012.

[65] L. Dercle, T. Henry, A. Carré, N. Paragios, E. Deutsch, and C. Robert. Reinventing radiation therapy with machine learning and imaging bio-markers (radiomics): State-of-the-art, challenges and perspectives. *Methods*, 188:44–60, 2020.

[66] C.K. Speirs, P.W. Grigsby, J. Huang, W.L. Thorstad, P.J. Parikh, C.G. Robinson, and J.D. Bradley. Pet-based radiation therapy planning. *PET Clin.*, 10:27–44, 2015.

[67] I. Vergalasova and J. Cai. A modern review of the uncertainties in volumetric imaging of respiratory-induced target motion in lung radiotherapy. *Med. Phys.*, 47:e988–e1008, 2020.

[68] Tao Wu, Renze Luo, Hongyu Lin, Hong Yu, Qingsong Wang, and Heng Liu. Research on focal segmentation of bone scan based on swin transformer. In *2023 4th International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, pages 426–430. IEEE, 2023.

[69] Junyu Chen, Ye Li, Licia P Luna, Hyun W Chung, Steven P Rowe, Yong Du, Lilja B Solnes, and Eric C Frey. Learning fuzzy clustering for spect/ct segmentation via convolutional neural networks. *Medical physics*, 48(7):3860–3877, 2021.

[70] Qiang Lin, Runxia Gao, Mingyang Luo, Haijun Wang, Yongchun Cao, Zhengxing Man, and Rong Wang. Semi-supervised segmentation of metastasis lesions in bone scan images. *Frontiers in Molecular Biosciences*, 9:956720, 2022.

[71] Noémie Moreau, Caroline Rousseau, Constance Fourcade, Gianmarco Santini, Ludovic Ferrer, Marie Lacombe, Camille Guillerminet, Mario Campone, Mathilde Colombié, Mathieu Rubeaux, et al. Deep learning approaches for bone and bone lesion segmentation on 18fdg pet/ct imaging in the context of metastatic breast cancer. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1532–1535. IEEE, 2020.

[72] Xiangbing Zhan, Jun Liu, Huiyun Long, Jun Zhu, Haoyu Tang, Fangfang Gou, and Jia Wu. An intelligent auxiliary framework for bone malignant tumor lesion segmentation in medical image analysis. *Diagnostics*, 13(2):223, 2023.

[73] A. Bert, I. Dmitriev, S. Agliozzo, and et al. An automatic method for colon segmentation in ct colonography. *Computerized Medical Imaging and Graphics*, 33(4):325–331, 2009.

[74] R. L. Siegel, K. D. Miller, and A. Jemal. Cancer statistics, 2016. *CA: A Cancer Journal for Clinicians*, 66(1):7–30, 2016.

[75] R. García-Figueiras, S. Baleato-González, A. R. Padhani, and et al. Advanced imaging techniques in evaluation of colorectal cancer. *RadioGraphics*, 38(3):740–765, 2018.

[76] M. A. Mazurowski, M. Buda, A. Saha, and M. R. Bashir. Deep learning in radiology: an overview of the concepts and a survey of the state of the art with focus on mri. *Journal of Magnetic Resonance Imaging*, 49(4):939–954, 2019.

[77] R. Tachibana, J. J. Näppi, J. Ota, and et al. Deep learning electronic cleansing for single- and dual-energy ct colonography. *RadioGraphics*, 38(7):2034–2050, 2018.

[78] Lisha Yao, Yingda Xia, Haochen Zhang, Jiawen Yao, Dakai Jin, Bingjiang Qiu, Yuan Zhang, Suyun Li, Yanting Liang, Xian-Sheng Hua, et al. Deepcrc: Colorectum and colorectal cancer segmentation in ct scans via deep colorectal coordinate transform. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 564–573. Springer, 2022.

[79] Yi-Jie Huang, Qi Dou, Zi-Xian Wang, Li-Zhi Liu, Ying Jin, Chao-Feng Li, Lisheng Wang, Hao Chen, and Rui-Hua Xu. 3-d roi-aware u-net for accurate and efficient colorectal tumor segmentation. *IEEE transactions on cybernetics*, 51(11):5397–5408, 2020.

[80] Junming Jian, Fei Xiong, Wei Xia, Rui Zhang, Jinhui Gu, Xiaodong Wu, Xiaochun Meng, and Xin Gao. Fully convolutional networks (fcns)-based segmentation method for colorectal tumors on t2-weighted magnetic resonance images. *Australasian physical & engineering sciences in medicine*, 41:393–401, 2018.

[81] Yankai Jiang, Shufeng Xu, Hongjie Fan, Jiahong Qian, Weizhi Luo, Shihui Zhen, Yubo Tao, Jihong Sun, and Hai Lin. Ala-net: Adaptive lesion-aware attention network for 3d colorectal tumor segmentation. *IEEE transactions on medical imaging*, 40(12):3627–3640, 2021.

[82] Yun Pei, Lin Mu, Yu Fu, Kan He, Hong Li, Shuxu Guo, Xiaoming Liu, Mingyang Li, Huimao Zhang, and Xueyan Li. Colorectal tumor segmentation of ct scans based on a convolutional neural network with an attention mechanism. *IEEE Access*, 8:64131–64138, 2020.

[83] Ying Liang, Diane Schott, Ying Zhang, Zhiwu Wang, Haidy Nasief, Eric Paulson, William Hall, Paul Knechtges, Beth Erickson, and X Allen Li. Auto-segmentation of pancreatic tumor in multi-parametric mri using deep convolutional neural networks. *Radiotherapy and Oncology*, 145:193–200, 2020.

[84] Zhihui Guo, Ling Zhang, Le Lu, Mohammadhadi Bagheri, Ronald M Summers, Milan Sonka, and Jianhua Yao. Deep logismos: Deep learning graph-based 3d segmentation of pancreatic tumors on ct scans. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 1230–1233. IEEE, 2018.

[85] Yuhei Iwasa, Takuji Iwashita, Yuji Takeuchi, Hironao Ichikawa, Naoki Mita, Shinya Uemura, Masahito Shimizu, Yu-Ting Kuo, Hsiu-Po Wang, and Takeshi Hara. Automatic segmentation of pancreatic tumors using deep learning on a video image of contrast-enhanced endoscopic ultrasound. *Journal of clinical medicine*, 10(16):3589, 2021.

[86] Yue Du, Xiaoying Zuo, Shidong Liu, Dai Cheng, Jie Li, Mingzhu Sun, Xin Zhao, Hui Ding, and Yabin Hu. Segmentation of pancreatic tumors based on multi-scale convolution and channel attention mechanism in the encoder-decoder scheme. *Medical Physics*, 50(12):7764–7778, 2023.

[87] Radhia Khdhir, Aymen Belghith, and Salwa Othmen. Pancreatic cancer segmentation and classification in ct imaging using antlion optimization and deep learning mechanism. *International Journal of Advanced Computer Science and Applications*, 14(3), 2023.

[88] Zhuotun Zhu, Dakai Jin, Ke Yan, Tsung-Ying Ho, Xianghua Ye, Dazhou Guo, Chun-Hung Chao, Jing Xiao, Alan Yuille, and Le Lu. Lymph node gross tumor volume detection and segmentation via distance-based gating using 3d ct/pet imaging in radiotherapy. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 753–762. Springer, 2020.

[89] DL Farfan Cabrera, Eloïse Grossiord, Nicolas Gogin, Dimitri Papathanassiou, and Nicolas Passat. Analysis of lymph node tumor features in pet/ct for segmentation. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 588–592. IEEE, 2021.

[90] Mariaulpa Sahalan, Aidil Munir Mazlee, Farah Nabila Mustafa Amirrudin, Nurafiqah Syazwani Mohd Jamil, Rahwani Nasir, Nusrah Athirah Suhaimi, Jareer Murtaza Amin, and Ahmad Naqib Mohd Qari. Auto segmentation of lymph node microscopy images. *Journal of Medical Device Technology*, 1(1):50–55, 2022.

[91] Hung Chu, Luis Ricardo De la O Arévalo, Wei Tang, Baoqiang Ma, Yan Li, Alessia De Biase, Stefan Both, Johannes Albertus Langendijk, Peter van Ooijen, Nanna Maria Sijtsema, et al. Swin unetr for tumor and lymph node segmentation using 3d pet/ct imaging: A transfer learning approach. In *3D Head and Neck Tumor Segmentation in PET/CT Challenge*, pages 114–120. Springer, 2022.

[92] Yanling Zhang, Yuejia Zhang, and Li Li. Lymph node image segmentation based on improved fcm clustering and multi-threshold. In *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013)*, pages 3002–3005. Atlantis Press, 2013/03.

[93] J. Ma, F. Li, and B. Wang. U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*, 2024.

[94] Xiangning Chen, Chen Liang, Da Huang, Esteban Real, Kaiyuan Wang, Yao Liu, Hieu Pham, Xuanyi Dong, Thang Luong, Cho-Jui Hsieh, Yifeng Lu, and Quoc V. Le. Symbolic discovery of optimization algorithms, 2023.

[95] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[96] Max de Grauw, Natália Alves, Megan Schuurmans, Henkjan Huisman, Bram van Ginneken, and Alessa Hering. The uls23 challenge public training dataset, 2023.

[97] Nicholas Heller, Fabian Isensee, Darya Trofimova, Rohan Tejpaul, Ziyue Zhao, Hewei Chen, and Christopher Weight. The kits21 challenge: Automatic segmentation of kidneys, renal tumors, and renal cysts in corticomedullary-phase ct. *arXiv preprint arXiv:2307.01984*, 2023.

[98] Colin Jacobs, E. Marleen van Rikxoort, Kieran Murphy, Mathias Prokop, Cornelia M Schaefer-Prokop, and Bram van Ginneken. Computer-aided detection of pulmonary nodules: a comparative study using the public lidc/idri database. *European radiology*, 26:2139–2147, 2016.

[99] Patrick Bilic, Patrick Christ, H. B. Li, Eugene Vorontsov, Avi Ben-Cohen, Georgios Kaissis, and Bjoern Menze. The liver tumor segmentation benchmark (lits). *Medical Image Analysis*, 84:102680, 2023.

[100] Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, Annette Kopp-Schneider, Bennett A Landman, and M. Jorge Cardoso. The medical segmentation decathlon. *Nature communications*, 13(1):4128, 2022.

[101] Holger Roth, Le Lu, Ari Seff, Kevin M Cherry, Judy Hoffman, S Wang, Jing Liu, Evrim Turkbey, and Ronald M Summers. A new 2.5 d representation for lymph node detection in ct. *The Cancer Imaging Archive*, 2015.