Study on Segmentation Methods of Pancreas and Related Regions from Volumetric CT Images

Chen Shen

Abstract

Computed Tomography (CT) is a widely used modality for acquiring tomographic images in clinical practice. It is the preferred choice for the non-invasive diagnosis of complex anatomical structures, due to its exceptional spatial accuracy and extensive imaging range. The rapid development of CT technology has led to the prominence of 3D CT volumes, replacing the previous reliance on 2D imaging. While 3D volumes offer more valuable information and spatial context of human anatomy, interpreting these volumes has imposed a significant workload for radiologists. The development of Computer-Aided Diagnosis (CAD) systems is desired, as they can assist radiologists in reducing interpretation time and simultaneously enhancing interpretation accuracy. Since segmentation is a critical task for CAD in the field of computer science, its accuracy significantly influences the reliability of the CAD system.

This thesis presents three topics related to the pancreas and its associated regions, aiming to assist in pancreatic diagnosis in clinical usage: segmentation of a healthy pancreas and its surrounding abdominal organs, segmentation of the pancreas and pancreatic tumors, and segmentation of the dilated pancreatic duct. Segmenting larger objects, such as abdominal organs, is relatively straightforward, but the task becomes significantly more challenging when dealing with smaller tissues like the pancreatic duct. The presented research follows a progression from handling simpler tasks to addressing more complex segmentation challenges.

The first topic focuses on the segmentation of a healthy pancreas and its multiple associated abdominal organs. Segmentation of the pancreas in abdominal organs is challenging, especially when compared with larger organs like the liver and stomach. The pancreas has notable variations in shape compared to other abdominal organs. Additionally, the low contrast of the pancreas and its boundaries adds to the difficulty. Anatomically, the pancreas is closely situated to neighboring organs like the stomach, duodenum, liver, and spleen. The segmentation of multiple abdominal organs aids in providing relational information for pancreas segmentation. Two distinct methods are proposed for segmenting a healthy pancreas and its associated abdominal organs and tissues. Both methods have successfully improved the segmentation accuracy of the pancreas.

The second topic focuses on pancreatic pathology, where the segmentation targets include the pancreas and pancreatic tumors. Pancreatic cancer, marked by a high mortality rate, demands precise diagnosis, often reliant on detecting pancreatic tumors. Accordingly, a segmentation method for pancreas and pancreatic tumors is proposed. Acquiring diverse abnormal datasets is crucial for robust models, but the challenge lies in collecting extensive medical imaging datasets due to cost and privacy constraints. To address this, an innovative Federated Learning (FL) framework is proposed for efficient pancreas and pancreatic tumor segmentation, offering a solution to the limitations of dataset collection in medical imaging. Two optimization methods for FL are investigated to address the issue of data heterogeneity in FL. The proposed method is shown to outperform traditional model aggregation methods in the segmentation of the pancreas and pancreatic tumor.

The third topic addresses the early detection of pancreatic pathology by focusing on the segmentation of the dilated pancreatic duct. In clinical studies, the dilation of pancreatic duct is regarded as a high-risk indicator for pancreatic cancer. Acknowledging the significance of this observation, the segmentation of dilated pancreatic ducts from

CT volumes holds valuable potential for the diagnosis of pancreatic cancer. The small size of the pancreatic duct in abdominal CT volumes poses a challenge for traditional organ segmentation methods based on Fully Convolutional Networks (FCNs). For this topic, a pancreatic duct segmentation framework employing a coarse-to-fine strategy is proposed. To further improve segmentation, a pancreatic anatomical attention method is proposed, and vesselness structure features are integrated, enhancing the precision crucial for timely intervention and improved patient outcomes. The proposed segmentation framework exhibits significant superiority compared to other methods in pancreatic duct segmentation using single-phase CT volumes.

In summary, this thesis presents segmentation methods for the pancreas and its related structures from CT volumes. Chapter 1 provides the aim of this thesis. Chapter 2 proposes two distinct methods for segmenting the healthy pancreas and its associated abdominal organs and tissues. Chapter 3 introduces a framework for the segmentation of the pancreas and pancreatic tumor using FL. Chapter 4 presents a framework for the segmentation of the pancreatic duct. Chapter 5 establishes a connection between the research presented in this thesis and Real-World Data Circulation (RWDC). Finally, Chapter 6 provides a summary and outlines directions for future work.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Computed tomography

Computed Tomography (CT) is one of the most important and popular modalities for acquiring tomographic images in clinical practice. The outstanding spatial accuracy and extensive imaging range of CT makes it an ideal choice for the non-invasive diagnosis of complex anatomical structures. In most cases, the term "CT" refers to X-ray CT, which is an advanced medical imaging technique that employs X-rays and computer processing to create highly detailed cross-sectional images of the body. In the early 1970s, the introduction of the Electric and Musical Industries (EMI) scanner by Godfrey Newbold Hounsfield marked a significant milestone in medical image analysis [3]. The EMI scanner utilizes a rotating X-ray tube and detector array to capture multiple X-ray images from various angles around the patient's body [4]. It is the first commercially available CT scanner on record. Since then, the development of CT scans has continued.

A CT scanner typically comprises three essential components that collaboratively generate intricate cross-sectional images of the body. An example of a modern CT scanner is shown in Fig. 1.1. The main components of a CT scanner include the gantry,

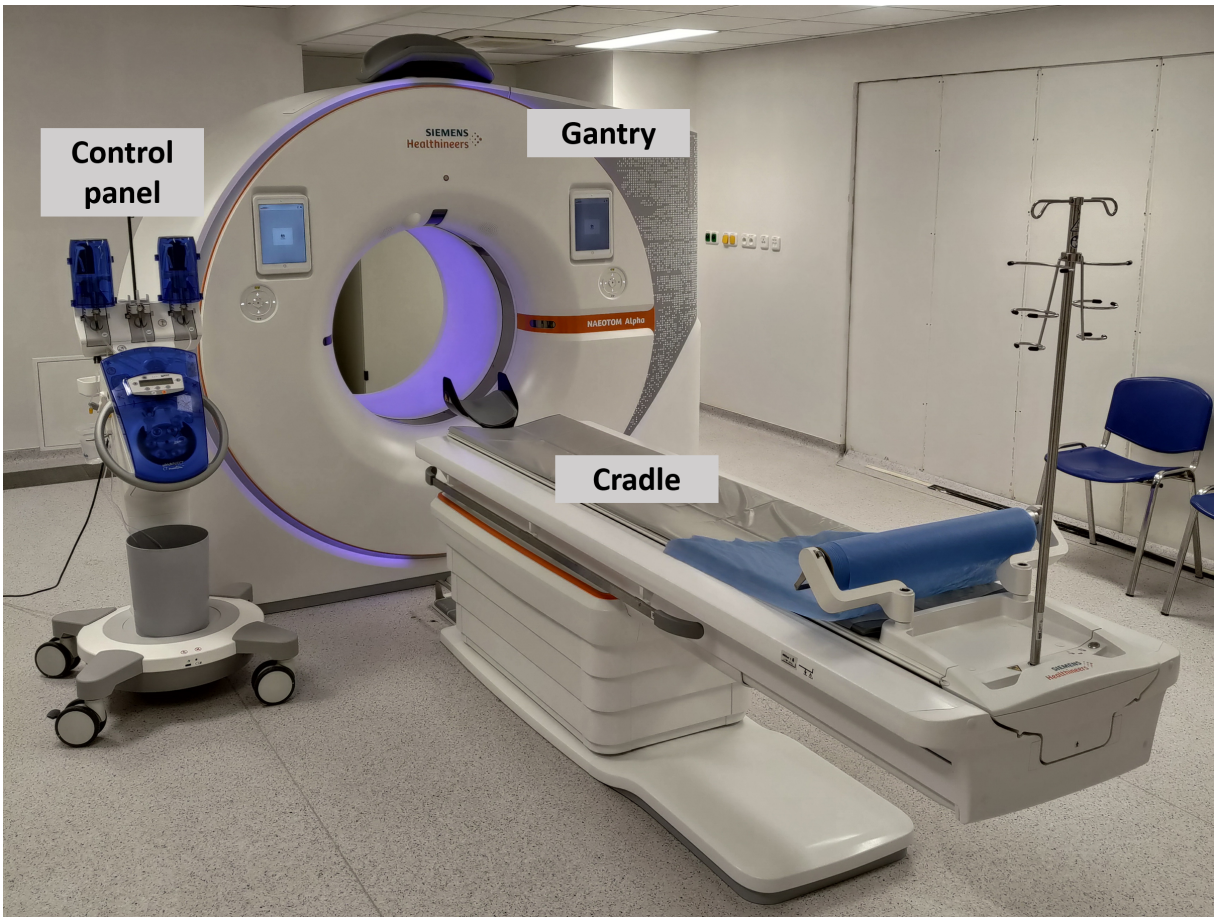


Figure 1.1: Modern CT scanner located at the Lochotín University Hospital in Pilsen, Czech Republic (Siemens NAEOTOM Alpha) [1].

cradle, and control panel. The gantry in a CT scanner is a circular or cylindrical structure that contains the X-ray tube. It rotates around the patient while scanning to obtain the X-ray images from various angles. The cradle, also known as the patient table or bed, is where the patient lies during the CT scanning process. The control panel manages the operation of the CT scanner [5].

CT images are displayed based on CT numbers, which are obtained through a linear transformation of the attenuation coefficients. The Hounsfield Unit (HU) is the unit of measurement used to represent the radiodensity of the tissues or materials in the CT

scan. The CT numbers can be defined as:

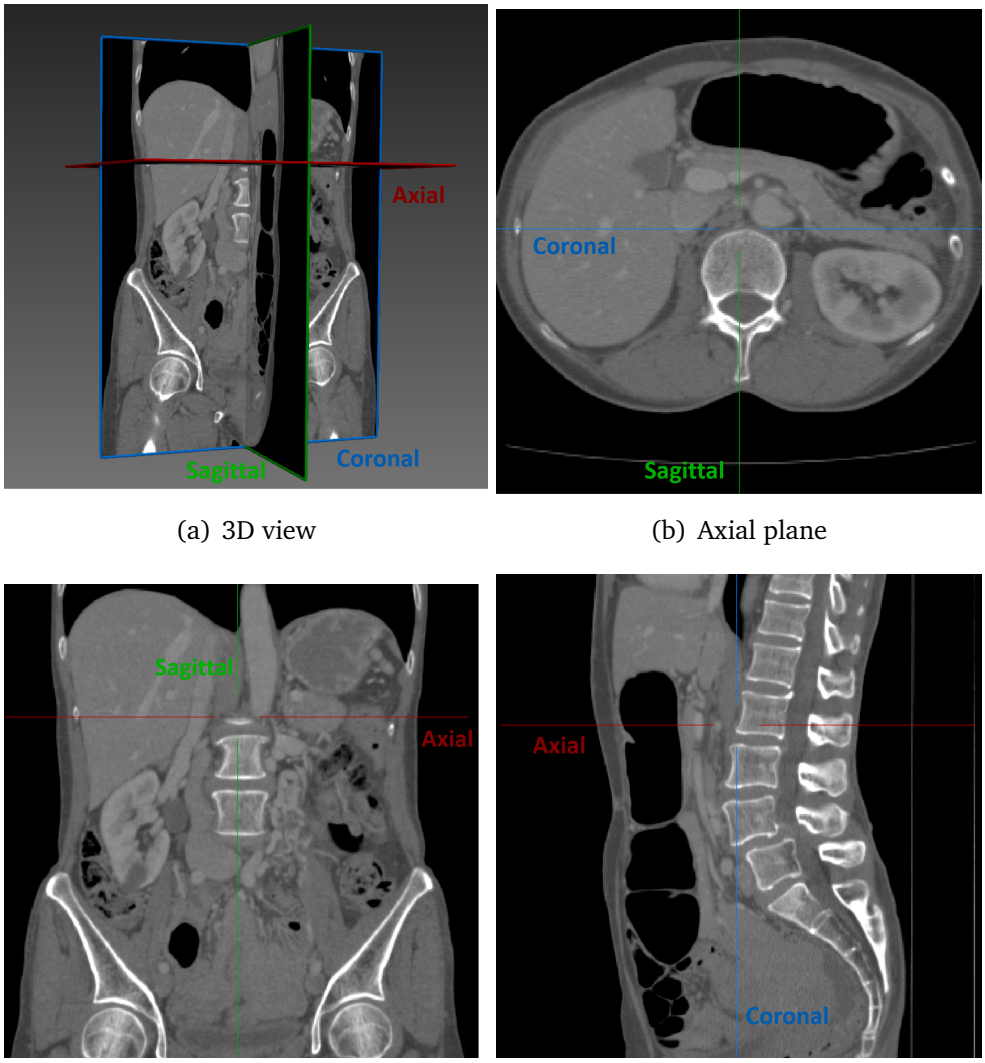
$$\text{HU} = \frac{\mu_t - \mu_w}{\mu_w} \times 1000, \tag{1.1}$$

where μ_t is the attenuation coefficient of material being scanned, and μ_w is the attenuation coefficient of water. Since $\mu_t - \mu_w = 0$, the assigned value of water is 0 HU. Air has a CT number of $-1,000$ HU because its attenuation coefficient is $\mu_t = 0$. Each organ or tissue typically has a specific CT number in the human body. These CT numbers can either be positive or negative, indicating how much more or less attenuate X-rays than water during scanning [6]. Low-density materials like fat typically exhibit negative CT numbers below 0 HU, whereas CT numbers of soft tissues, such as organs and muscles, are slightly higher than 0 HU. On the other hand, hard tissues, like bones and teeth, show much higher CT numbers.

After obtaining the CT numbers from the scanner, the CT images are commonly reconstructed into three main planes: axial plane, coronal plane, and sagittal plane. Each plane provides distinct internal information about the body. The axial plane shows views from top to bottom, while the coronal plane offers front-to-back views, and the sagittal plane provides side-to-side views. An example of the cross-sectional images from a 3D abdominal CT scan in axial, coronal, and sagittal views is shown in Fig. 1.2.

1.1.1 Computer-aided diagnosis for volumetric CT images

With the rapid development of CT technology, 3D volumetric CT images are predominantly utilized in clinical practice, replacing the previous reliance on 2D imaging. Although 3D volumetric images contain more valuable information and spatial context of human anatomy, interpreting them has become a substantial workload for radiologists. It becomes more challenging to catch important details within a limited image reading time, increasing the risk of overlooking crucial information [6]. The development



(a) 3D view (b) Axial plane

(c) Coronal plane (d) Sagittal plane

Figure 1.2: Example of the cross-sectional images of 3D abdominal CT in different views. Axial, coronal, and sagittal planes are represented by red, blue, and green lines, respectively.

of Computer-Aided Diagnosis (CAD) systems is desired as they can aid radiologists in shortening interpretation time and improving interpretation accuracy concurrently.

A CAD system quantitatively processes medical images and provides radiologists with the computer output that serves as a supplementary "second opinion" [7]. The

exploration of CAD commenced in the 1980s at the University of Chicago [8–11]. It is crucial to emphasize that while CAD systems can be valuable tools to assist radiologists, their purpose is not to replace human expertise. This distinction separates them from the concept of automated computerized diagnosis, a concept that was explored in the 1960s and 1970s [12–14]. The goal of CAD is to improve the accuracy and reproducibility of radiological diagnoses while simultaneously reducing the time required for radiologists to interpret images. The exploration of CAD system development dates back to the 1990s in Japan, marking the inception of research endeavors in this field [15]. Since then, continuous advancements and research have significantly contributed to the evolution and refinement of CAD systems in the realm of medical image analysis.

In the past few decades, extensive research has focused on lesion detection [16–18], disease prediction [19–21], and other related areas [22–24]. CAD systems serve as valuable tools, assisting healthcare professionals not only in identifying abnormalities but also in contributing to more efficient and accurate diagnostic workflows. The development of CAD systems in the medical field is influenced by various factors. Establishing a high-performance CAD system for medical diagnosis necessitates collaboration between medical professionals and computer scientists. The combination of medical insight and computational methods is crucial for CAD system success. The accuracy and reliability of the system depend on the quality of medical data, the expertise of ground-truth annotators, and the effectiveness of implemented models. In the field of computer science, research aims at developing computational methods for gaining a deeper understanding of medical data is crucial for advancing CAD capabilities.

Segmentation is a crucial task for CAD in the field of computer science. It helps precisely identifying specific regions or structures in medical images. Segmentation plays a pivotal role in CAD by enabling the precise identification of specific regions or structures within medical images. This step is crucial for accurate and detailed analysis, as it allows CAD systems to focus on relevant areas during the diagnostic process.

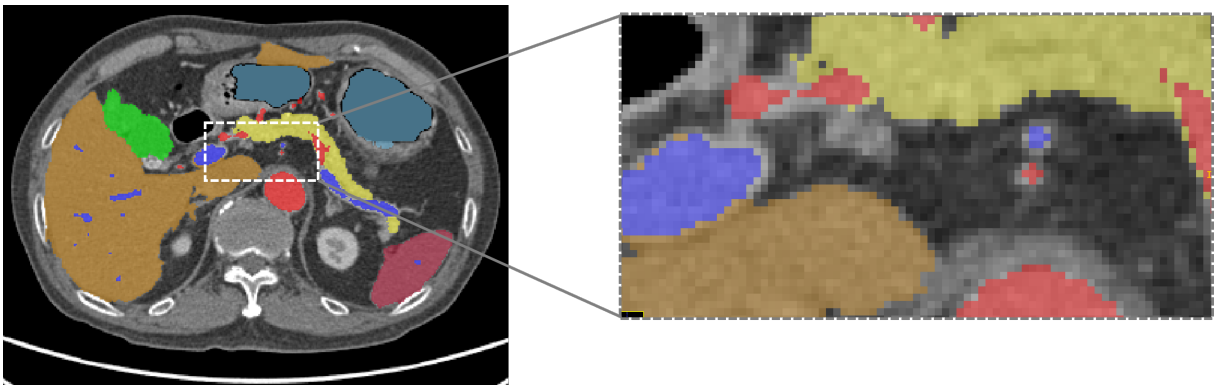


Figure 1.3: Illustration depicting the segmentation from a volumetric CT image, with each color representing distinct organ classes. Each voxel is identified as an individual organ.

Through segmentation, the CAD system can delineate and differentiate between various anatomical structures or abnormalities, contributing to a more refined and targeted interpretation of medical images. Moreover, segmentation aids in extracting quantitative data, facilitating the assessment of the size, shape, and characteristics of identified structures. This quantitative information enhances diagnostic accuracy and provides valuable insights for healthcare professionals. In essence, segmentation in CAD acts as a cornerstone, enabling the system to navigate and comprehend intricate details within medical images, ultimately contributing to more effective and reliable medical diagnoses.

1.1.2 Segmentation from volumetric CT images

In the context of volumetric CT images, segmentation involves the assignment of labels to individual voxels in 3D volumetric images [25]. This process is crucial for distinguishing and categorizing different tissues or organs within volumetric CT images, enabling a comprehensive analysis of medical images. An Illustration of segmentation from a volumetric CT image is shown in Fig. 1.3.

Traditional segmentation methods

Segmentation research in medical image analysis has a rich history that predates the widespread adoption of deep learning techniques. Before the emergence of deep learning, researchers have explored a variety of classical computer vision and machine learning methods to address medical image segmentation challenges.

The atlas-based method is one of the widely used traditional segmentation methods in medical images. In this method, the organ's existence probability atlas is initially computed using volumetric CT images and ground truth. This anatomical statistical information is then leveraged for organ segmentation. Shimizu et al. [26] introduced a technique for simultaneous segmentation of 12 organs from non-contrast 3D abdominal CT images. They implemented atlas-guided segmentation with the expectation maximization algorithm to address the large fluctuation problem in feature distribution parameters. The segmentation process incorporates multiple level sets to delineate organ boundaries. Park et al. [27] devised a probabilistic atlases thresholding method on the segmentation of 24 abdominal volumetric CT images. Statistical information is used from multiple abdominal volumetric CT images to generate a probabilistic atlas, representing the likelihood of voxel locations corresponding to different anatomical structures.

Graph-cuts [28] is another traditional image segmentation method to formulate segmentation as an optimization problem on a graph. In this method, the image is represented as nodes, and the edges between nodes represent the relationships between pixels. The segmentation is then obtained by finding the minimum cut in the graph. Boykov et al. [29] developed a graph-cuts method to allow users to interactively separate the foreground and background. An approximation to multi-way graph-cuts was introduced, making it applicable to multi-class segmentation problems. Chen et al. [30] integrated the active appearance model and live wire cost function into the parameters

of graph cuts for 3D abdominal organ segmentation. A live-wire cost function was proposed to specify paths within the image. This was integrated into the graph-cuts framework. Linguraru et al. [31] indicated a combination of statistical modeling and a graph-cuts based approach for organ segmentation. Graph-cuts methods are effective in incorporating both local and global information, making them suitable for refining segmentation results obtained from statistical models.

Conditional Random Fields (CRFs) are also a well-known approach for organ segmentation. Unlike the atlas-based methods, CRFs-based methods can adjust empirical parameters to determine the organ regions automatically. It is used to model the relationships and dependencies between pixels from 3D volumetric CT images. CRFs and Histograms of Oriented Gradients (HOG) features were first combined by Bhole et al. [32] on medical images. They evaluated the influence on parameter learning in CRFs which compared to the Markov Random Fields (MRF) counterparts. Also, they indicated the importance of introducing HOG features in anatomical segmentation problems. Nimura et al. [33] used CRFs as a probabilistic graphical model for the right lung, left lung, heart, liver, spleen, right kidney, and left kidney segmentation. Yang et al. [34] proposed a voxel-connection structure based on CRFs in multi-organ segmentation. They utilized the stochastic gradient descent algorithm in the learning phase and Maximum A Posteriori (MAP) inference in the prediction. Furthermore, they introduced FCNs into CRFs. A fine-tuned 3D FCN is utilized to generate the probability maps for each organ. Beyond organ segmentation, CRFs have been applied to various other types of segmentation in medical images, including blood vessels [35] and brain tumors [36].

Deep learning methods

With the rapid advancement of machine learning, deep learning has emerged as the predominant method for medical image segmentation tasks. Convolutional Neural

Networks (CNNs) have played a pivotal role, allowing the training of complex models from extensive datasets and delivering state-of-the-art results with remarkable efficiency [37, 38]. CNNs enbale to learn hierarchical representations of features directly from the training data [39, 40]. This intrinsic ability alleviates the need for manual feature engineering, where human-defined features are specified in advance. Typically, deep learning based segmentation can be divided into two main categories: supervised segmentation and unsupervised segmentation. In supervised segmentation, the neural network is trained on annotated training data, where each input has a corresponding ground-truth segmentation. The performance of supervised methods is often high, especially when ample annotated data are available for training. However, a limitation of supervised methods is their dependence on large annotated training datasets, which can be extremely challenging to obtain in certain fields. On the other hand, unsupervised segmentation involves the model learning patterns or relationships from the training data without relying on annotated segmentation ground truth. However, achieving high performance on complex segmentation tasks with unsupervised methods can be challenging due to the absence of explicit guidance from annotated data. This thesis mainly focuses on supervised learning methods.

The introduction of FCNs by Long et al. [41] has significantly enhanced the feasibility of pixel-wise semantic segmentation. FCNs, being the most extensively employed architecture in supervised semantic segmentation, have played a pivotal role in advancing this field. U-Net proposed for 2D biomedical image segmentation is the most well-known structure in the medical image segmentation field. The architecture introduced an encoder-decoder structure to process feature maps, utilizing deconvolution for up-sampling the feature maps, and incorporating skip connections to integrate features from the encoder path. It was initially introduced by Ronneberger et al. [42] for 2D biomedical image segmentation and has subsequently been extended to 3D volumetric applications, as demonstrated in works such as 3D U-Net [43] and V-Net [44].

These extensions have led to significant performance enhancements in 3D medical image segmentation. Isensee et al. made significant improvements to the segmentation performance of the U-Net-shaped structure through meticulous hyperparameter tuning and pre-processing strategies [45].

1.2 Anatomical background

1.2.1 Abdomen

The term “human abdomen” denotes the anterior region situated between the thoracic diaphragm and the pelvis [46]. This area plays a pivotal role in the overall functioning of the body. The abdominal cavity contains a variety of organs and blood vessels essential for daily bodily functions, including digestion, metabolism, and other fundamental processes. Fig. 1.4 depicts an illustration of the anatomy of the posterior abdomen. In medical contexts, having a thorough understanding of the structure and components of human abdominal anatomy is crucial. Numerous diseases are directly related to the abdominal organs. Due to the presence of many critical organs, there is a significant focus on studying the abdomen. This understanding offers a fundamental viewpoint necessary for studying and diagnosing conditions and disorders associated with the abdomen. This thesis focuses primarily on the pancreas, while also taking into account relevant areas like blood vessels (artery and portal vein), liver, spleen, stomach, and gallbladder. Subsequently, a concise introduction to the anatomical background of these tissues and organs will be presented.

Artery and portal vein

The artery is depicted in red, while the portal veins are represented in blue in Fig. 1.4. They are two major blood vessels in the abdomen that contribute significantly to blood

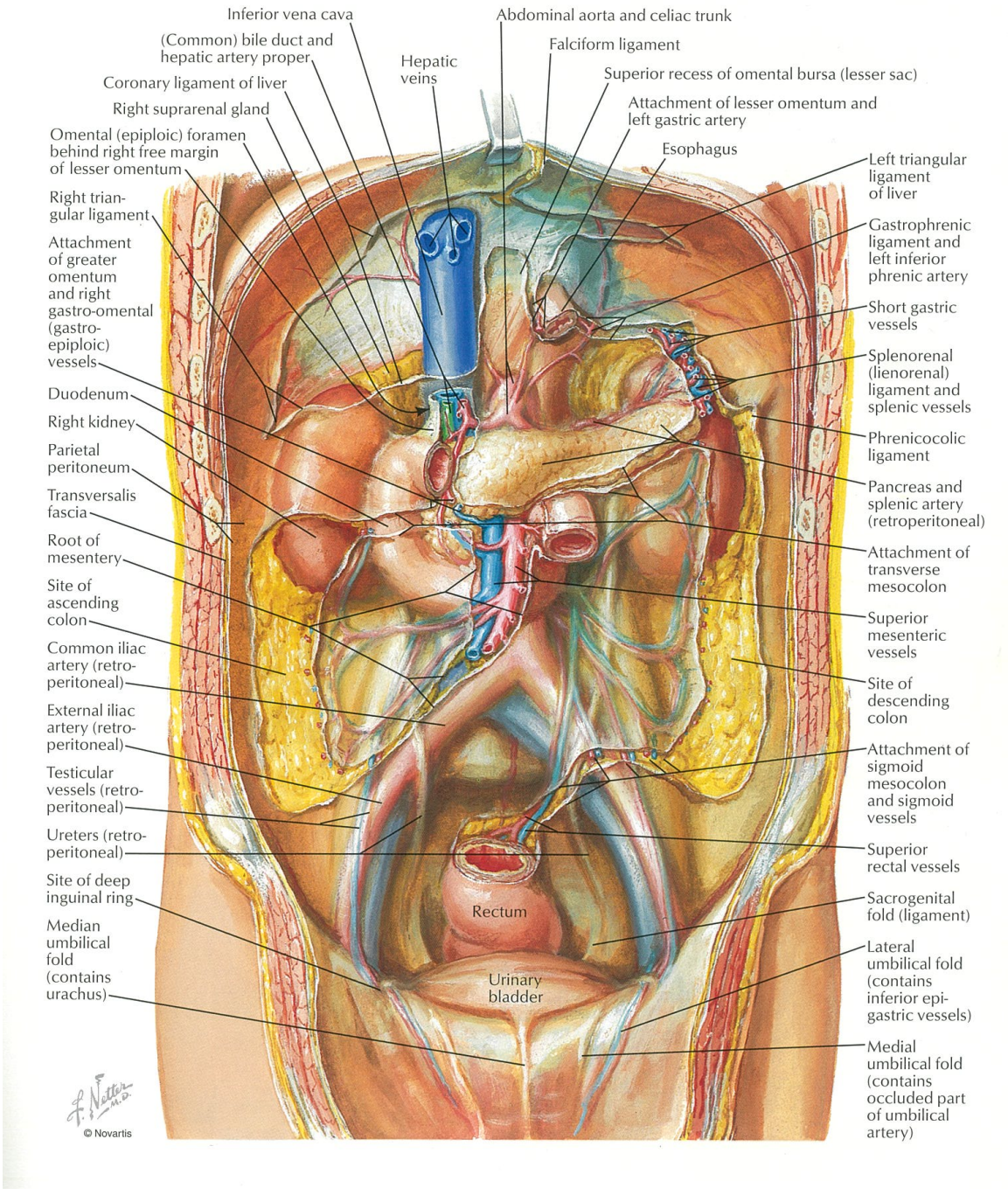


Figure 1.4: Illustration of the anatomy of the posterior abdomen [2].

circulation throughout the entire body. The artery transports oxygenated blood from the heart to all parts of the body, including the abdominal organs. Conversely, the portal vein plays a crucial role in nutrient transport. It carries blood with nutrients and toxins from the digestive organs to the liver [47].

Liver

The liver is the largest and heaviest solid organ in the body. It is located in the right upper quadrant of the abdominal cavity. The liver's substantial size and strategic placement emphasize its pivotal role in facilitating various essential physiological functions that are vital for maintaining life and overall well-being. The function of the liver is versatile. It is a multifunctional organ that plays a crucial role in various physiological processes. Its functions span across metabolic, synthetic, hemopoietic, and immunological activities, making it a central hub for maintaining overall health [48].

Stomach

The stomach is an essential digestive organ with a sac-like structure that distinguishes itself as the broadest within the abdomen. It is situated in the left hypochondrial region, beneath the lower rib cage. The size, shape, and location of the stomach can vary significantly, influenced by factors such as an individual's posture and the level of fullness in the stomach [49]. The stomach serves multiple functions, including aiding in digestion, regulating nutrition, and influencing appetite control [50].

Spleen

The spleen is an organ located in the left upper corner of the abdomen, right next to the stomach. It is part of the lymphatic system and serves multiple functions in the body. The spleen plays a crucial role in the body as a filter of the blood, removing damaged

blood cells and bacteria, and is involved in immune responses [51]. Additionally, it actively participates in immune responses, contributing significantly to the body's defense mechanisms.

Gallbladder

The gallbladder is a small organ located on the inferior surface of the liver. It plays a crucial role in the digestive process. In between meals, it acts as a reservoir for bile, a digestive fluid produced by the liver. Bile is essential for breaking down fats in the small intestine during digestion. When a person consumes a meal, especially one that contains fats, the gallbladder contracts, releasing the stored and concentrated bile into the small intestine. This process aids in emulsifying fats, facilitating their digestion and absorption into the body [52].

1.2.2 Pancreas and pancreatic duct

The pancreas is an organ located in the upper abdomen, situated behind the stomach and in front of the spine. It is connected with several abdominal organs including the spleen, stomach, duodenum, and colon. For a healthy adult, the pancreas typically weighs around 100 grams on average and measures between 14–25 centimeters in length [53]. It performs both exocrine and endocrine functions for the human body. In its exocrine function, it secretes digestive enzymes essential for the digestion and absorption of food, facilitating the intake of nutrients into the body. In its endocrine function, it releases hormones, primarily insulin and glucagon, into the bloodstream, playing a key role in regulating blood sugar levels [54].

The detailed anatomical structure of the pancreas is illustrated in Figure 1.5. It can be divided into three main components: head, body, and tail. Additionally, there is a crucial duct structure inside it known as the pancreatic duct. This duct connects to both

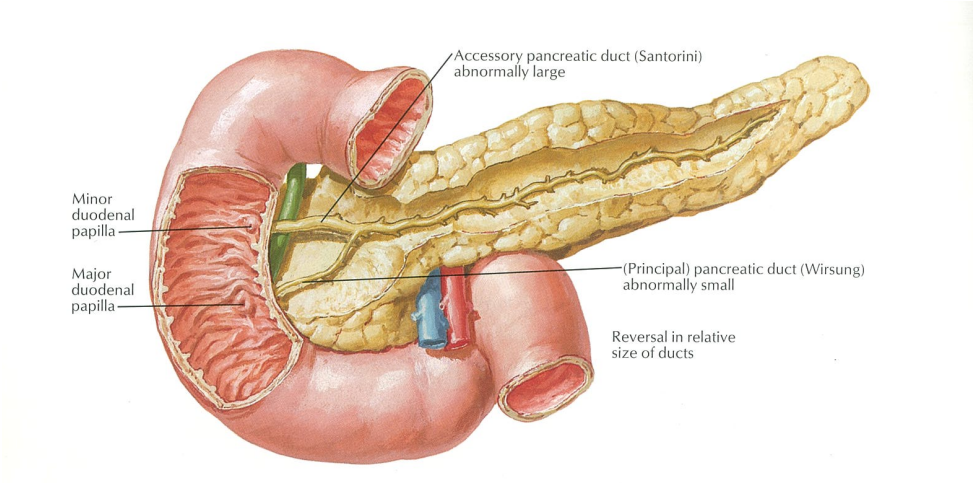


Figure 1.5: Illustration of the anatomy of the pancreas [2].

the first part of the small intestine and the bile duct [55]. It plays a pivotal role in transporting digestive enzymes produced by the pancreas to the duodenum, where they contribute to the breakdown of food.

Pancreatic pathology

Pancreatic cancer, a pathology that has garnered considerable attention, stands as one of the most lethal malignancies, claiming the lives of hundreds of thousands of individuals globally each year. Due to its high fatality rate, it is often referred to as the “king of carcinoma” [56]. The American Cancer Society estimates that pancreatic cancer will result in approximately 49,830 fatalities among adults in the United States in 2022 [57]. It has the lowest 5-year survival rate, which is around 10%, when compared to other malignancies [58]. The primary treatment for pancreatic cancer is resection. A critical factor contributing to the low survival rate is the limited manifestation of symptoms, often leading to detection at an advanced stage. Consequently, only a minority of patients are eligible for surgical resection at the time of diagnosis [58]. Therefore, it is essential to find it early when it is still treatable.

Pancreatic Ductal AdenocarCinoma (PDAC), which develops in the main duct, accounts for more than 90% of pancreatic cancer [59]. Several clinical studies suggest that dilatation of the main pancreatic duct indicates an increased risk of pancreatic cancer [60–62]. This thesis centers on the pancreas and related structures, including pancreatic tumors and ducts. The goal is to contribute to the development of a CAD system tailored for pancreatic conditions.

1.3 Aim of this study

This thesis aims to introduce innovative segmentation methods for the pancreas and its related structures, holding substantial promise for future clinical applications. Addressing the challenges related to accurate and efficient segmentation of these critical anatomical structures contributes to the development of CAD, treatment planning, and patient care in healthcare scenarios.

In typical medical scenarios, when performing noninvasive examinations of the abdomen, a variety of diagnostic imaging techniques are commonly employed. Among these, the predominant methods in clinical practice include CT, Ultrasonography (US), and Magnetic Resonance Imaging (MRI) [63–66]. These advanced imaging modalities play a central role in providing comprehensive insights into abdominal structures, enabling precise and detailed assessments across a broad spectrum of medical conditions. Each imaging modality comes with its strengths and weaknesses. MRI, with its outstanding soft tissue contrast, makes it superior for imaging the brain and joints [63]. However, its less frequent use is mainly attributed to its higher cost compared to other modalities. US stands out as the most cost-effective modality, but it may not provide the same level of detailed soft tissue imaging as CT or MRI. CT, on the other hand, is the most widely used modality in abdominal imaging, owing to its advantageous balance between cost and image quality. The American College of Radiology (ACR) Appropri-

ateness Criteria has proposed guidelines for the appropriate investigation of various disease entities [67]. In clinical practice, CT is the predominant diagnostic modality for diagnosing pancreatic diseases compared to other imaging techniques [61, 65]. Given this clinical perspective, the thesis is primarily centered on the analysis of abdominal volumetric CT images.

The research presented in this thesis centers around the pancreas, a crucial organ situated in the abdomen. Previous studies have extensively explored larger abdominal organs, such as the liver [68–70], spleen [70, 71], and stomach [72, 73], resulting in commendable achievements in understanding and segmenting these larger anatomical structures. However, when it comes to smaller organs like the pancreas, there still exists a noticeable gap in segmentation performance that needs to be improved [74, 75].

Besides exploring the pancreas itself, the research extends the exploration to the surrounding abdominal regions. Additionally, this research delves into the study of abnormal conditions, specifically focusing on pancreatic tumors and dilated pancreatic ducts. Through this exploration, this thesis tries to contribute to a comprehensive understanding of the pancreas, its surroundings, and the implications of disorders such as tumors and duct dilation. In a broader aspect, this research aims to provide valuable insights into the field of medical image analysis and healthcare.

1.4 Research overview

The work included in this thesis aims to propose segmentation approaches for the pancreas and its associated region to assist in pancreatic diagnosis in clinical usage. It can be primarily categorized into three main topics: (1) Segmentation of the pancreas with abdominal multi-organ, (2) Segmentation of the pancreas and pancreatic tumor, and (3) Segmentation of the pancreatic duct. Segmenting large objects like abdominal organs is relatively easy, but dealing with tiny tissues such as the pancreatic duct poses a

significant challenge. The research progresses from tackling simpler tasks to addressing more complex ones.

1.4.1 Topic 1: Pancreas and multiple associated abdominal organs segmentation

The first topic involves studying healthy pancreas cases. In this topic, two approaches to the segmentation of the pancreas and its multiple associated abdominal organs are presented. Segmenting the pancreas from abdominal organs is challenging, especially when compared to larger organs like the liver and stomach. Unlike other abdominal organs, the pancreas shows significant individual shape differences. Additionally, the low contrast of the pancreas and its boundaries adds to the difficulty. Anatomically, the pancreas is closely situated to neighboring organs like the stomach, duodenum, liver, and spleen. This proximity can lead to unclear boundaries in imaging studies, making it challenging to distinguish it from surrounding structures [5]. In actual clinical practice, radiologists can enhance the identification and separation of the pancreas from surrounding structures by making informed judgments based on the relative positions and anatomical connections between organs. This experiential knowledge can also be applied to automated pancreas segmentation situations. Here, two segmentation methods for the pancreas with multi-organ segmentation are presented. The first method is a traditional supervised deep learning-based approach. When employing FCNs for segmentation, organs with small voxel sizes often exhibit poor performance. This is because they learn less, and crucial information may be disregarded during the learning process. The input for an FCN is a crop of the entire volumetric CT image, constrained by computational resource limitations. In this process, positional information is disregarded. To tackle this issue, this work proposes a modification to an FCN that considers the spatial information of input sub-volumes. The second method involves leveraging

Automatic Machine Learning (AutoML) techniques for this task. This eliminates the need for carefully selecting hyperparameters for organ segmentation.

1.4.2 Topic 2: Pancreas and pancreatic tumor segmentation

The second topic focuses on pancreatic pathology. Pancreatic cancer is a significant pancreatic pathology that attracts high attention, mainly due to its high mortality rate. The presence of a pancreatic tumor is a primary indicator for diagnosing pancreatic cancer. In most cases, the identification of a tumor in the pancreas is a key outcome that leads to the diagnosis of pancreatic cancer. Here, a method for pancreas and pancreatic tumor segmentation is proposed. When addressing pancreatic pathologies such as pancreatic tumors, acquiring a substantial amount of abnormal datasets is crucial for developing a highly generalizable model. However, collecting a large number of datasets in medical image analysis is often challenging due to cost and privacy concerns. Federated Learning (FL) emerges as an innovation to address this issue. In this context, an efficient FL framework is proposed for pancreas and pancreatic tumor segmentation.

1.4.3 Topic 3: Pancreatic duct segmentation

The last topic proposes a research perspective aimed at facilitating the early detection of pancreatic pathology, which can serve as an entry point for the early detection of pancreatic cancer. Early detection enables prompt intervention and treatment, positively impacting patient outcomes and overall prognosis. The anatomical structure of the pancreatic duct is very small in abdominal volumetric CT images. Traditional organ segmentation methods based on FCNs may not perform well on this target. Thus, a pancreatic duct segmentation framework is investigated to overcome this issue. A coarse-to-fine strategy for pancreatic segmentation is proposed. Additionally, to enhance pancreatic duct segmentation performance, a pancreatic anatomical attention

method is introduced to guide the FCN to focus on the pancreas regions during training. In addition, vesselness structure features is incorporated to enhance the structure segmentation during segmentation.

1.5 Perspective of real-world data circulation

Real-World Data Circulation (RWDC) is a recently proposed concept by the Information Technology (IT) community [76]. Its central concept involves analyzing collected data, conducting a thorough analysis, and subsequently developing products or services based on the findings. The process then loops back to receiving feedback in the form of data derived from the utilization of these products and services. The key idea behind RWDC is that the analysis of collected data, followed by the integration of results into new designs and fabrication, forms the foundation for value creation. The paradigm of RWDC is illustrated in Fig. 1.6. The work in this thesis utilizes volumetric CT images obtained from hospitals. The data is analyzed to develop segmentation methods for CAD systems, with the potential application in clinical practice. This process aligns with the RWDC framework. Moreover, the three topics this thesis explored are interconnected, enhancing the potency of the circulation. This interconnectedness contributes to a thorough exploration and understanding within the scope of the study.

Acquisition

Acquisition of real-world data is the first step of this circulation. This thesis focuses on volumetric CT images obtained from real-world clinical scenarios. To employ the data for segmentation tasks, ground truth for each case needs to be obtained. The precision of the ground truth significantly impacts the quality of subsequent segmentation models. In the medical image analysis field, since manual annotation necessitates highly skilled

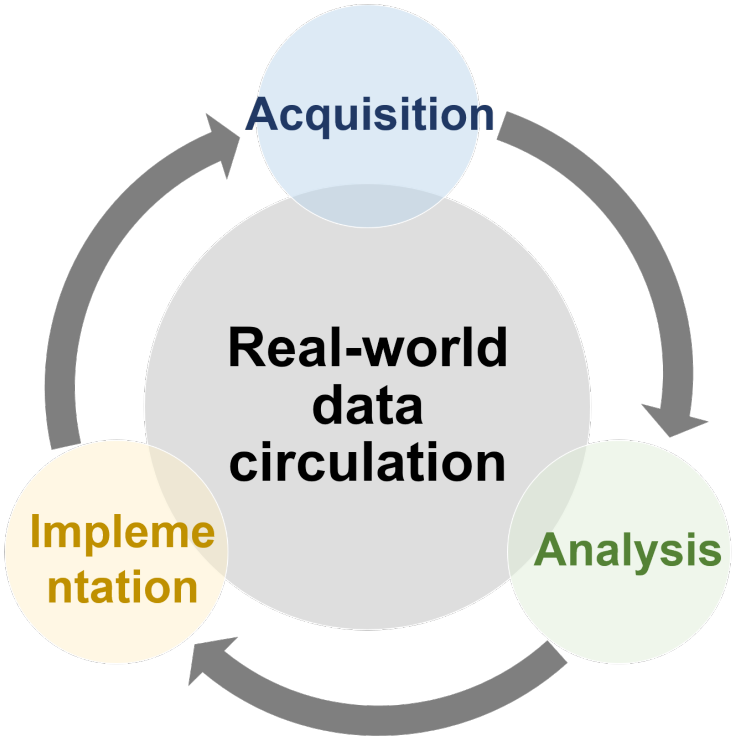


Figure 1.6: Illustration of the real-world data circulation paradigm.

experts, which can be cost-prohibitive, annotating ground truth is challenging. While using an implemented segmentation model is an alternative, it requires high accuracy.

Analysis

Data analysis in RWDC involves developing methods to address specific problems based on the acquired data from the real world. This thesis proposes three distinct segmentation topics for data analysis. Initially, volumetric CT images of the healthy pancreas are analyzed for the segmentation of the pancreas and its multiple associated organs. The other two topics focus on unhealthy pancreas, specifically those with pancreatic cancer or pancreatic duct dilation. Improving segmentation accuracy represents the primary challenge in this step. In this thesis, various deep learning-based methods are explored

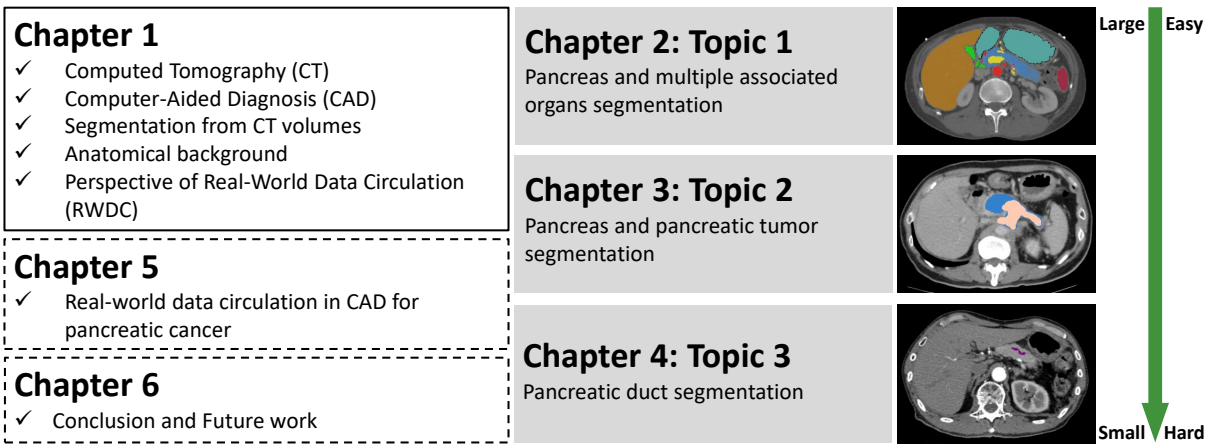


Figure 1.7: Overview of the chapters in this thesis.

to improve segmentation tasks. In real-world clinical applications, the significance lies not only in the methodology but also in the effectiveness of the approach.

Implementation

Implementation involves actively creating products or services based on insights derived from data analysis, and this aspect is tied to the "value creation" of RWDC. The research outcomes of this thesis have the potential to be applied in CAD systems, aiding radiologists in their diagnoses. In the clinical scenario, the accuracy of the analysis markedly influences the quality of the implementation. Additionally, the feedback obtained from the implementation can serve as valuable data for re-evaluating the analysis, thereby completing the circulation.

1.6 Thesis structure

This thesis comprises five chapters. An overview of the relationship between each chapter is depicted in Fig. 1.7. Chapter 1 served as an introduction, providing background

information on anatomy and modalities. Additionally, the research was introduced from the perspective of RWDC. Chapter 2 introduces two methods for pancreas segmentation within the context of multiple abdominal organs analysis. Chapter 3 covers the method for segmentation of the pancreas and pancreatic cancer. Chapter 4 introduces the methods for pancreatic duct segmentation. Chapter 5 presents the RWDC process and its societal value in this research. Finally, Chapter 6 concludes with a summary and a discussion regarding the future directions of this study.

Chapter 2

Segmentation of pancreas and its multiple associated organs

2.1 Background

In this chapter, two segmentation methods for the pancreas and its multiple associated abdominal organs are presented. Semantic segmentation from volumetric Computed Tomography (CT) images is a critical and challenging task in medical image analysis, as it aids in understanding the major anatomical structures of humans. Moreover, accurate segmentation results have the potential to be applied in Computer-Aided Diagnosis (CAD) systems for clinical workflows. In the pancreatic CAD system scenario, the pancreas segmentation results are fundamental, aiding in comprehending the location and condition of the pancreas. However, there are many challenges in pancreas segmentation. One limitation is the relatively small size of the pancreas compared to other abdominal organs. Moreover, significant individual variations in shape, size, and appearance further complicate the segmentation process. Additionally, the low contrast of the pancreas boundary with other organs adds to the complexity of the task. These

factors often limit segmentation accuracy, potentially impacting diagnostic quality when incorporating results into CAD systems in real-world clinical settings. Utilizing information from the multiple associated organs of the pancreas could be valuable for improving pancreas segmentation.

The increasing capacity of computing power supports the rapid development of deep learning in the last decade. Nowadays, numerous segmentation approaches based on deep learning have been introduced with the rapid development of Convolutional Neural Networks (CNNs). Among them, the Fully Convolutional Network (FCN) has demonstrated its advantages for medical image tasks [41], achieving higher segmentation performance. Various FCN architectures have been developed for medical image analysis, of which U-net [42] designed for 2D biomedical image segmentation, stands out as one of the most famous FCN architectures widely used in medical image analysis. It was originally proposed for 2D image processing and later extended to 3D volumetric processing by Çiçek et al. [43]. Moreover, numerous 3D U-Net-like FCNs have been proposed, showcasing their significant advantages in the field of medical image analysis [44, 77, 78]. The encoder-decoder network architecture is proven to achieve acceptable performance on pixel-wise segmentation tasks.

When training with FCNs, the network architecture, and its corresponding hyperparameters will influence the performance. The network architectures are carefully designed in traditional machine learning settings. The choices of hyperparameters are typically made by experts through a trial-and-error process. The emergence of Automated Machine Learning (AutoML) introduces a novel concept that involves automating the entire pipeline of machine learning, making it more accessible for individuals to conduct machine learning experiments. In this chapter, two methods for the segmentation of the pancreas and its multiple associated abdominal organs are introduced. The first method aims to enhance the traditional FCNs, while the other leverages AutoML for the automatic design of the FCN architecture and hyperparameter selection. Experiments

on a large annotated multiple abdominal organ dataset demonstrate the effectiveness of the proposed methods.

2.2 Contribution

In this chapter, two distinct segmentation methods for the pancreas and its multiple associated abdominal organs are introduced. One method focuses on enhancing a traditional FCN, while the other utilizes AutoML. The primary contribution can be summarized as follows:

1. The first method improves the FCN segmentation performance by incorporating coordinate spatial information into the network. The sub-volume of CT may discard crucial spatial information that is beneficial for the segmentation. Gamma correction is introduced in data augmentation to improve the robustness of FCNs.

2. The second method introduces an AutoML framework for segmentation based on reinforcement learning. Instead of randomly splitting a subset of data, a proxy dataset is obtained representing the entire dataset for AutoML.

3. Experiments on a large abdominal CT dataset with healthy pancreas conditions demonstrate the effectiveness of the proposed methods.

2.3 Method 1: Spatial-embedded FCN

2.3.1 Motivation

When training with volumetric images using FCNs, it is necessary to crop sub-volumes from the entire images as the input due to Graphics Processing Unit (GPU) memory limitations. However, this process may result in the loss of spatial information in the

sub-volumes compared to utilizing the entire volumetric CT images. Recovering dismissed spatial information can potentially enhance segmentation when using FCNs. Y-Net [79] is a novel FCN architecture that incorporates central coordinate values into the U-Net structure. The inclusion of coordinate information has demonstrated effectiveness in segmentation tasks, leading to improved performance. To further exploit the information retained in random cropping, Wang et al. introduced multiple selected coordinate values for x, y, and z as spatial position information for sub-volumes in kidney segmentation [80]. This approach remarkably improved the kidney segmentation accuracy. Since it is expected that the spatial information-embedded FCNs would also have a positive effect on multi-class organ segmentation, its effectiveness is evaluated. When training an FCN, another challenge is the need for a large amount of training data to prevent overfitting. Generating a substantially large number of datasets is particularly challenging for segmentation tasks, especially in the medical image analysis field where labeling volumetric CT images is time-consuming and expensive. To address this issue, data augmentation is widely employed. This technique effectively increases the training data and enhances the robustness of trained models.

2.3.2 Network architecture

The network architecture of the proposed spatial-embedded FCN is illustrated in Fig. 2.1. Similar to the original 3D U-Net [43], it is composed of symmetric analysis and synthesis paths with four resolution levels each. Each level consists of two convolutional layers with a size of $3 \times 3 \times 3$ in both analysis paths and synthesis paths. A Rectified Linear Unit (ReLU) is followed by each convolution layer. In the analysis path, max-pooling is adopted to downsample the feature maps. For synthesis paths, $2 \times 2 \times 2$ up-convolution layers are utilized to remap the lower-resolution feature maps. The shortcut connections from layers of equal resolution in the analysis path to the synthesis path are changed

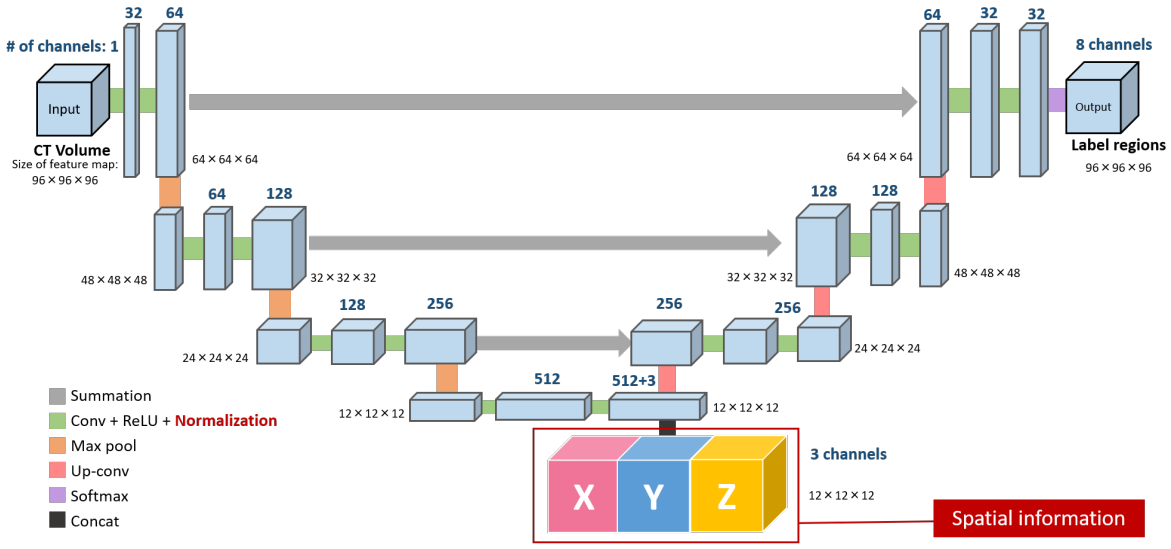


Figure 2.1: Network architecture of the proposed spatial-embedded FCN. Spatial position information is incorporated at the bottom of the FCN.

from concatenation to summation. The input of FCN is volumetric CT image and the output is eight classes of voxel-wise prediction results.

A three-channel feature map is concatenated to the bottom to introduce the spatial position information to the FCN. This information includes the value of x, y, and z coordinates, respectively. Fig. 2.2 indicates how to extract the spatial position information from the volumetric CT images. Each channel includes one pair of coordinates respectively. The size of each channel is $12 \times 12 \times 12$ voxels to suit the bottleneck layer. As the size of the input sub-volume is $96 \times 96 \times 96$ voxels, the coordinates are extracted for every 8 voxels. Firstly, the coordinates of each voxel in a sub-volume of the entire CT are extracted which are defined as x_n, y_m, and z_l, where $n = i + 8, i + 16, ..., i + 96$, $m = j + 8, j + 16, ..., j + 96$, $l = k + 8, k + 16, ..., k + 96$. Then, the values are normalized into the range of $[0, 1]$. Each value of coordinates x, y, and z of sub-volumes is divided by the corresponding values of width (W), height (H), and depth (D) of the entire volumetric CT image. This information is utilized as the spatial information in

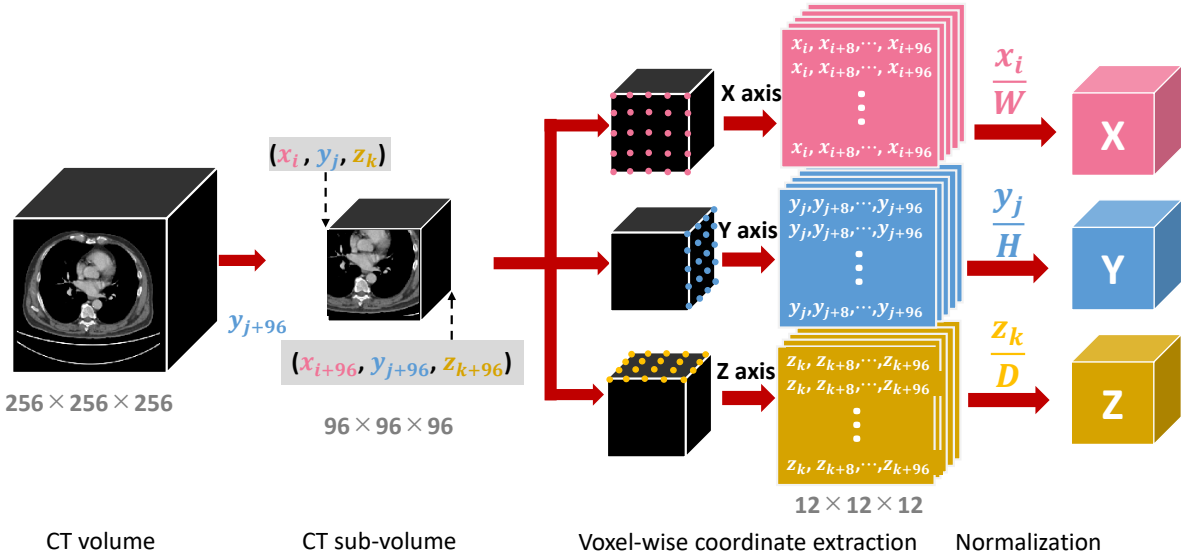


Figure 2.2: Calculation of the embedding of spatial information. Subvolumes are randomly cropped from volumetric CT image and then the voxel-wise coordinate values are calculated and normalized into the range of $[0, 1]$. W, H, and D indicate the width, height, and depth of the volumetric CT image, respectively.

the feature map. The coordinates of all voxels for each patch are used. Spatial position information of all voxels is adopted, instead of only considering the center coordinates of input sub-volume as [79].

2.3.3 Improved data augmentation

A large amount of data is always necessary to gain high performance in deep-learning-based methods. However, annotated volumetric CT images are hard to acquire because of the high cost. Moreover, if the volumetric CT images come from the same hospital and settings, the trained model may be imbalanced which would lead to poor performance on unseen data. Data augmentation is widely utilized to effectively solve these problems.

This work adopts non-rigid deformations including random 3D translation, rotation,

and B-spline deformation in both volumetric CT images and the ground truth, which is based on the method proposed by Özgün et al. [43]. Besides the general data augmentation mentioned above, gamma correction inspired by No New-Net [45] is also adopted. It is a nonlinear operation, which is widely used to adjust the contrast of images. With random gamma correction, volumetric CT images with different contrast conditions could be generated. The intensity of each pixel can be defined as $\boldsymbol{I}$, thus the results $\boldsymbol{I'}$ of gamma correction could be obtained as:

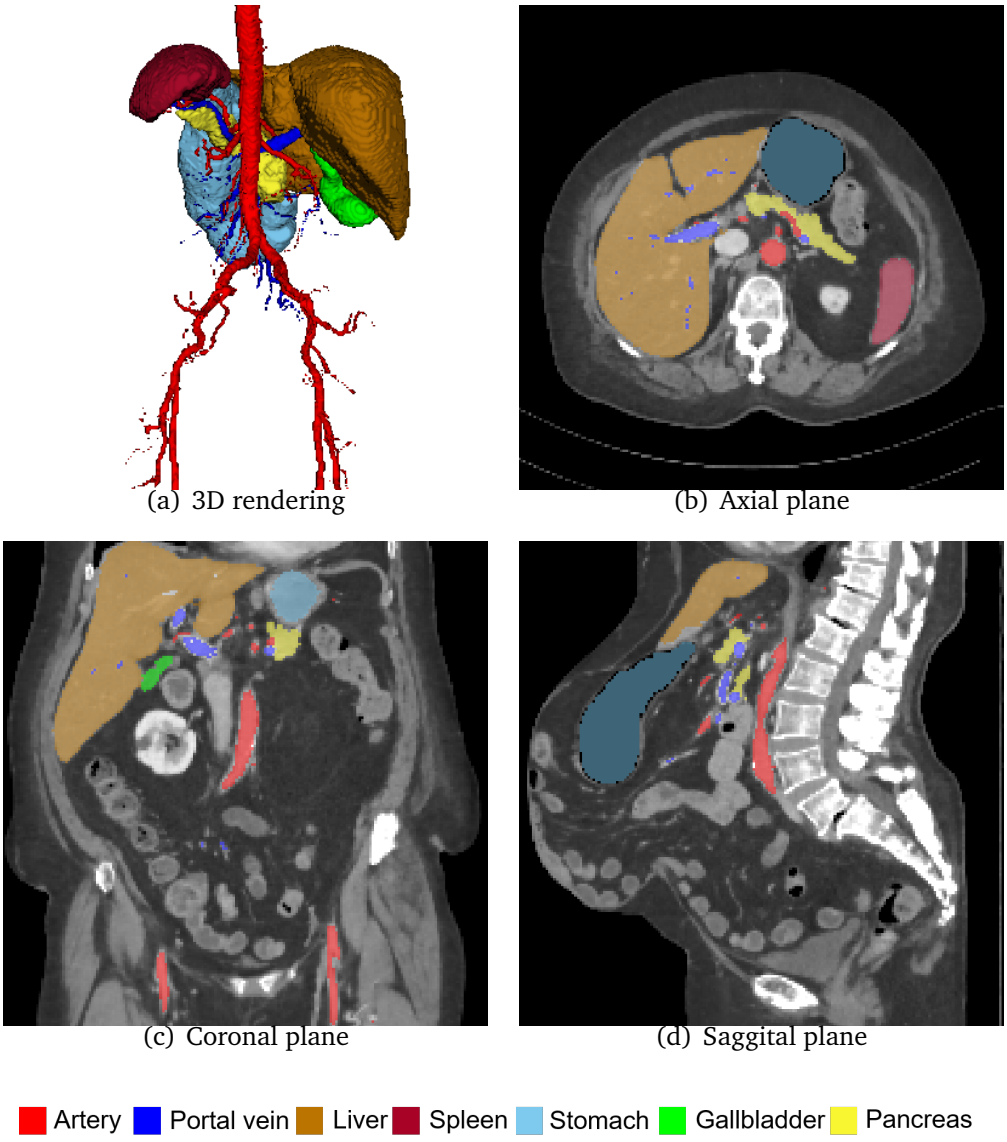
$$\boldsymbol{I'} = i_{\max} \left(\frac{\boldsymbol{I}}{i_{\max}} \right)^{\frac{1}{\gamma}}, \tag{2.1}$$

where $i_{\max}$ is the largest intensity of CT. When $\gamma > 1$, the difference in dark areas tends to be large and the image will turn brighter, and when $\gamma < 1$, the difference in bright areas tends to be large and the image will turn darker.

2.3.4 Experimental results

Dataset

400 abdominal clinical volumetric CT images in the portal-venous phase collected for pre-operative planning in gastric surgery were prepared. Each volumetric image contained 460–1,177 slices with the size of 512×512 pixels per slice. The dataset was randomly divided into subsets of 360 training volumetric images and 40 validation volumetric images. The ground truth including seven abdominal organs or tissues (artery, portal vein, liver, spleen, stomach, gallbladder, and pancreas) was annotated manually using semi-automated segmentation tools. Examples of CT data with the corresponding labels are shown in Fig. 2.3.



(a) 3D rendering (b) Axial plane

(c) Coronal plane (d) Saggital plane

Figure 2.3: Examples of CT data and corresponding ground truth in the axial, coronal, and sagittal planes, as well as 3D rendering.

Implementation

FCNs were implemented using the Keras[1] with TensorFlow[2] backend. For training, a mini-batch size of 3 was used with randomly cropped sub-volumes from different vol-

[1] `https://keras.io/` (Accessed on 2024/1/15)

umetric CT images. Each volumetric image was re-sampled to an isotropic resolution of $1 \times 1 \times 1$ mm^2. For optimization, the input and output sizes of the network were $96 \times 96 \times 96$ voxels, which were chosen based on the memory consumption of the 3D FCNs. For gamma correction, a range of γ from 0.5 to 2.0 was set to improve the data augmentation. The network was trained with an initial learning rate of 0.01 using the Adaptive moment estimation (Adam) [81], which is based on stochastic gradient descent. The loss function employed was the average Dice loss. The experiments were executed on a Deeplearning BOX with an NVIDIA Quadro P6000 GPU with 24GB memory.

Experimental results

A 3D U-Net-like architecture without spatial information presented in 2.3.2 was used as the baseline FCN, and compared with the inclusion of spatial position information (Spatial) and data augmentation (Aug.). Figure 2.4 shows examples of segmentation results of original FCN [78], FCN+Spatial, and FCN+Spatial+Aug.

Table 2.1 compares the Dice Similarity Coefficient (DSC) of each organ with the original FCN, FCN+Spatial, and FCN+Spatial+Aug. We can see that the spatial position information improved the overall segmentation performance, especially on blood vessels. Data augmentation showed a positive influence on pancreas segmentation for the testing data from the same hospital and CT scanner as the training data.

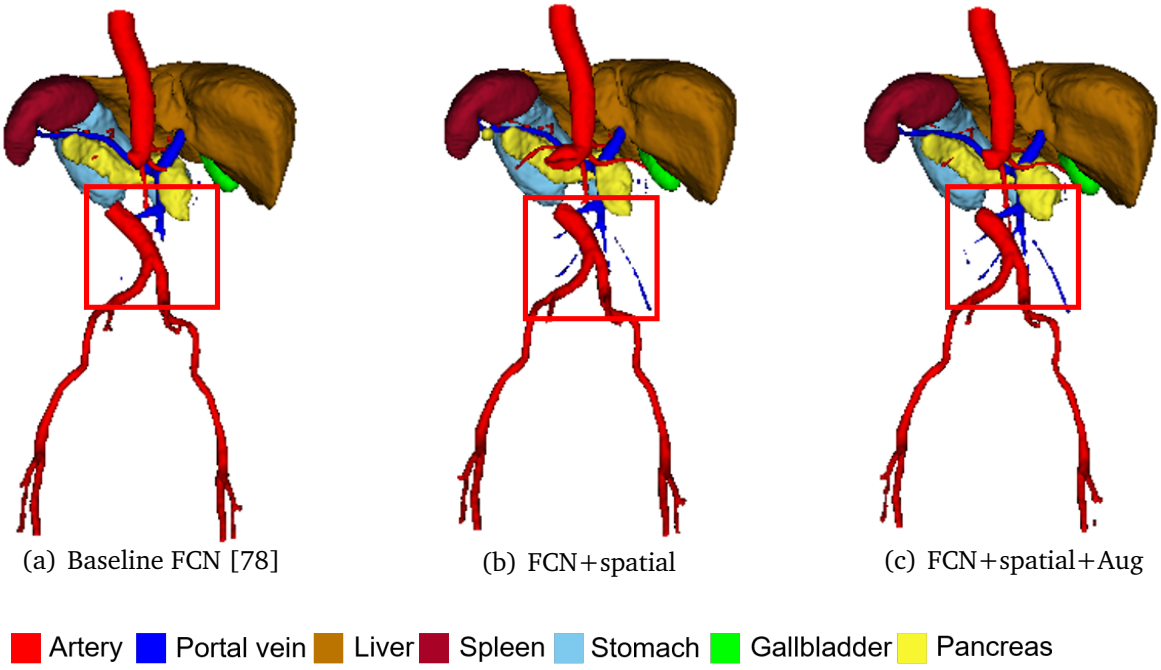
2.3.5 Discussion

The main breakthrough of this work was to introduce the spatial information-embedded FCN from single organ segmentation to multi-class organ segmentation. With the introduction of spatial position information into the FCNs, the segmentation accuracy in-

[2] `https://www.tensorflow.org/` (Accessed on 2024/1/15)

Table 2.1: Multi-class DSC of each organ with original FCN, FCN+Spatial, and FCN+Spatial+Aug. The best average score for each organ is shown in **bold**.

Organ	FCN [78]	FCN+Spatial	FCN+Spatial+Aug.
Artery	83.9%	**85.7%**	83.8%
Portal Vein	74.3%	**75.9%**	74.2%
Liver	96.0%	**96.3%**	96.1%
Spleen	96.0%	**96.8%**	96.4%
Stomach	**89.6%**	87.9%	88.8%
Gallbladder	85.2%	**88.1%**	87.9%
Pancreas	78.3%	79.4%	**81.0%**
Average	86.2%	**87.2%**	86.9%



(a) Baseline FCN [78] (b) FCN+spatial (c) FCN+spatial+Aug

Figure 2.4: Comparison of prediction results. Regions in red rectangles show that spatial information has a positive influence on the branch parts of portal vein segmentation.

creased for almost all classes. Especially for the gallbladder, the performance increased by about 3%. For blood vessel segmentation, like artery and portal veins, the spatial position information also had positive influences. Four types of data augmentation were

adopted to artificially increase the training data. The segmentation performance for the pancreas gained the highest accuracy, which achieved 81%. However, it seemed less help for other organs or tissues. One of the reasons could be that the data came from the same hospital and CT scanner. Effectiveness of unseen data taken by different devices and contrast conditions should be examined in the future.

2.4 Method 2: AutoML with proxy data

2.4.1 Motivation

Despite the fact that U-Net is commonly used as a benchmark for a variety of tasks in medical image analysis, researchers always need to adjust the network architecture or tune the hyperparameters carefully to fit each task and dataset. To fully boost the potential of a neural network and achieve State-Of-The-Art (SOTA) performance in deep-learning-based methods, a significant amount of experience and effort is necessary to select the proper training strategy for each certain task. The training strategy involves the design of network architecture, as well as the choice of initial learning rate, loss function, optimization, data augmentation, and so on.

Recent work on AutoML shows promise for designing a proper deep learning based training strategy without the demand for manual hyperparameters tuning [82], making deep learning more accessible to beginners and experts from other domains. AutoML is especially suitable for medical image segmentation as only a few experts are equipped with both the clinical and computer science knowledge needed to build performant models. Yang et al. introduced a reinforcement learning-based AutoML strategy [83] for 3D medical image segmentation. One deficiency of AutoML is the considerable time taken in the hyperparameter optimization phase. Searching hyperparameters using an entire dataset in AutoML causes immense computational costs. A straightforward

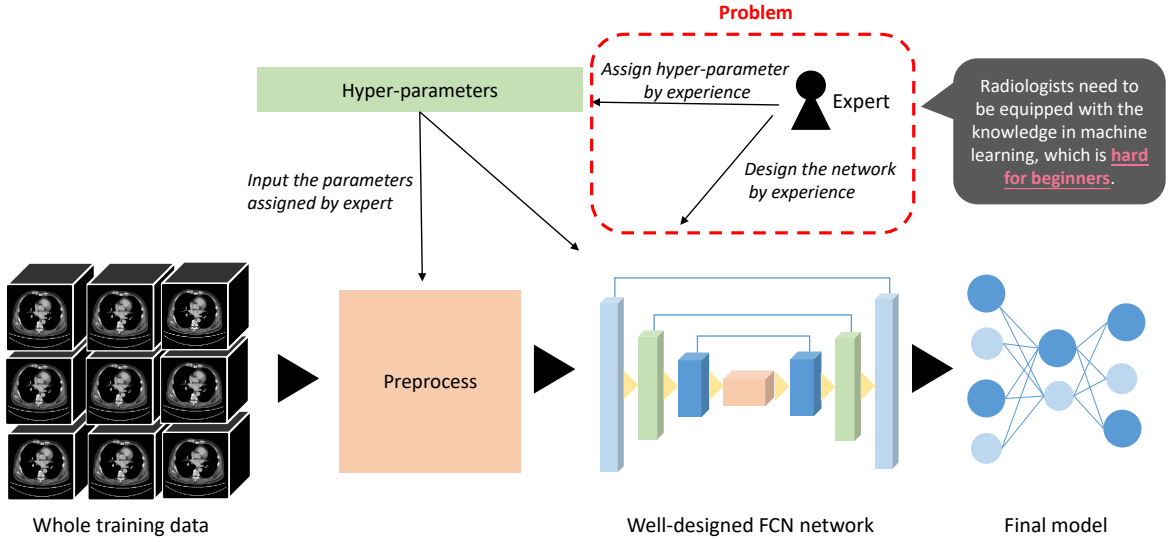
solution is to use only a small subset of training and validation data rather than the whole dataset to overcome this problem. However, randomly dividing the data into subsets sometimes results in a bias that may influence the AutoML's performance. Nath et al. [84] investigated an image ranking mechanism to select the most representative examples, a "proxy dataset", from the whole dataset.

2.4.2 Framework overview

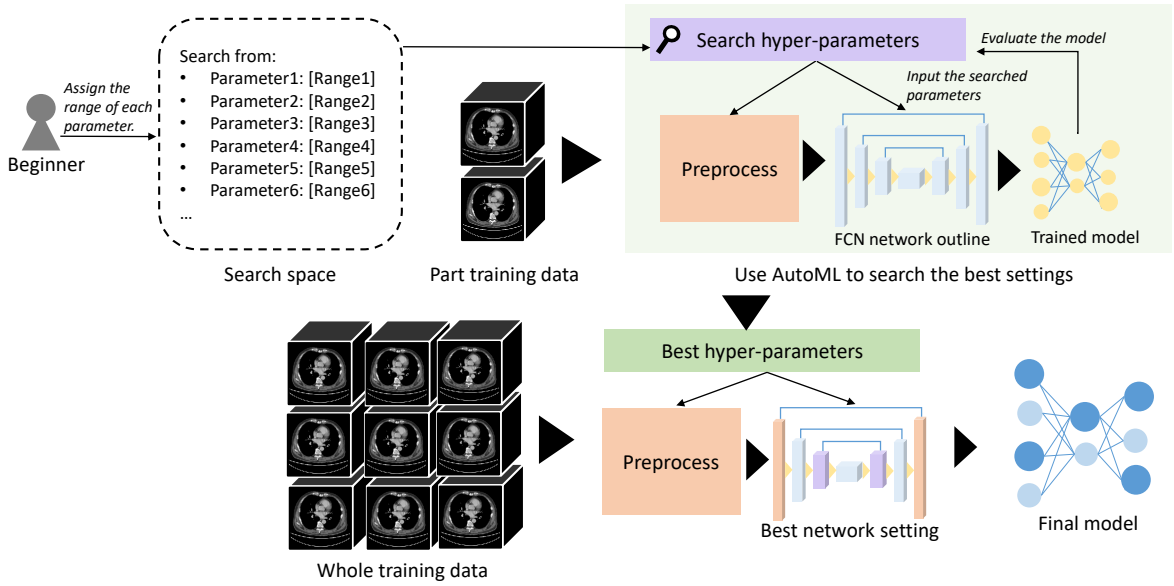
A comparison between traditional machine learning and the AutoML framework is shown in Fig. 2.5. Firstly, the importance of the entire dataset is ranked by measuring the similarity of each data. The best proxy data for AutoML is chosen from the top of the list. Then, the proxy dataset is used for network and hyperparameter optimization based on reinforcement learning to obtain the best settings. Finally, the whole training data and the best settings are used to train a final model. The segmentation performance of the final model is evaluated on testing data. The way of selecting the proxy data is demonstrated in 2.4.3 and the detailed AutoML strategies are explained in 2.4.4.

2.4.3 Proxy data selection

Using the entire dataset for hyperparameter search in AutoML will take a long time and a large amount of computational power. To reduce time and computation, proxy data are chosen instead of using the whole training dataset in AutoML. In previous studies, dividing a subset of training data randomly was considered as a solution to this problem [85]. However, the randomly divided subset sometimes has a bias which might influence the AutoML performance. This bias can be reduced by selecting an appropriate proxy dataset to reflect the entire dataset. Nath et al. investigated two measurements to indicate the importance of certain images among the entire dataset [84]. Figure 2.6



(a) Workflow of traditional machine learning.



(b) Workflow of AutoML.

Figure 2.5: Comparison between traditional machine learning and the AutoML framework. In traditional machine learning, radiologists need to be equipped with knowledge in machine learning, which is hard for beginners. AutoML could alleviate this problem.

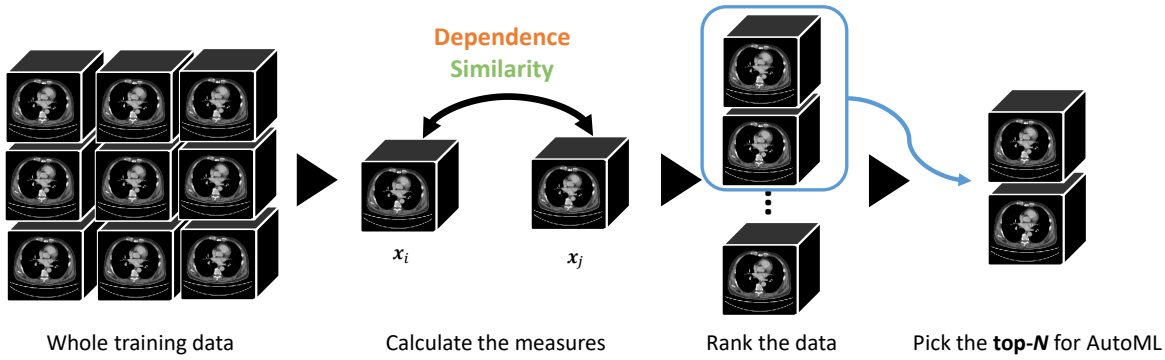


Figure 2.6: Illustration of how to choose the proxy data for AutoML.

shows an illustration of the proxy data selection procedure in this framework.

Suppose there is a dataset $S = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n\}$, the relation of two different examples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ can be denoted as $m(\boldsymbol{x}_i, \boldsymbol{x}_j)$. To explore the importance of data $\boldsymbol{x}_i$, the mean of the relation measures could be calculated as $m(\boldsymbol{x}_1, \boldsymbol{x}_i)$, $m(\boldsymbol{x}_2, \boldsymbol{x}_i)$, ..., and $m(\boldsymbol{x}_n, \boldsymbol{x}_n)$. As indicated in [84], the Mutual Information (MI) and Normalized local Cross-Correlation (NCC) can be used as the 3D pair-wise measurements. MI indicates the mutual dependence between two data. Its value will be high if one data has a big amount of information than the other data.

The MI of two data can be defined as:

$$\mathcal{MI}\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \sum_{\boldsymbol{x}_i} \sum_{\boldsymbol{x}_j} P\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) \log \frac{P\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)}{P\left(\boldsymbol{x}_i\right) P\left(\boldsymbol{x}_j\right)}, \tag{2.2}$$

where $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are flattened vectors of volumetric CT images. $P\left(\boldsymbol{x}_i\right)$ is marginal probability distribution of $\boldsymbol{x}_i$, and $P\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)$ is the joint probability distribution of two vectors $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. On the other hand, NCC specifies the correlation between two data. A higher value is a sign that two data are more similar.

The NCC of two data can be defined as:

$$\mathcal{NCC}\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \frac{1}{\Omega} \sum_{\boldsymbol{d} \in \Omega} \frac{\left[\sum_{d_i} \left\{d_i(\boldsymbol{x}_i) - \overline{\boldsymbol{d}}(\boldsymbol{x}_i)\right\} \left\{d_i(\boldsymbol{x}_j) - \overline{\boldsymbol{d}}(\boldsymbol{x}_j)\right\}\right]^2}{\sum_{d_i} \left\{d_i(\boldsymbol{x}_i) - \overline{\boldsymbol{d}}(\boldsymbol{x}_i)\right\}^2 \sum_{\boldsymbol{d}_i} \left\{d_i(\boldsymbol{x}_j) - \overline{\boldsymbol{d}}(\boldsymbol{x}_j)\right\}^2}, \tag{2.3}$$

where d_i is a voxel position in $9 \times 9 \times 9$ voxels local window and $\boldsymbol{d}$ are the voxels around it. Local mean around position d_i within the window in $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ can be written as $\overline{\boldsymbol{d}}(\boldsymbol{x}_i)$ and $\overline{\boldsymbol{d}}(\boldsymbol{x}_j)$.

MI is used as measurement in this study because Nath et al. [84] found that MI beats NCC in terms of picking proxy data for AutoML for organ segmentation. The training images are then ranked in order of lower MI value, which indicates that the image volume is less similar to other images. A certain percentage of the ranked data are selected as the proxy data for the AutoML experiment.

2.4.4 Network and hyperparameter searching strategy

The performance of the segmentation model highly relies on the network structure and hyperparameter settings. Training data may affect the design of networks and hyperparameters. Deep-learning-based methods require a great deal of effort to find the proper settings manually. AutoML is an alternative for this limitation that explores the appropriate settings within the specified search spaces.

For a searching parameter λ, either a value range $[\lambda_{\text{min}}, \lambda_{\text{max}}]$ or a finite searching set $V = \{\lambda', \lambda'', ..., \lambda'^{\cdots'}\}$ could be specified. In the value range, a floating target number λ^* can be acquired for each training. Given a set of hyperparameters $\Lambda = \{\lambda_1, \lambda_2, ..., \lambda_n\}$, n is the number of hyperparameters to be searched in AutoML. The searching objective $\Lambda^* = \{\lambda_1^*, \lambda_2^*, ..., \lambda_n^*\}$ can be obtained jointly in each training.

The network structure used for this work is SegResNet [86], which is designed for 3D medical segmentation. The variational autoencoder branch is eliminated and only the

U-Net-like backbone is kept as the baseline for the network architecture search. Block number combinations for both the encoder and decoder phases and the kernel number of the first convolution layer are searched for. Each block consists of a normalization layer, ReLu, and a $3 \times 3 \times 3$ convolution layer. A summation connection joins the output of the normalization and convolution layer. The normalization methods searched for are Batch Normalization [87] and Group Normalization [88].

2.4.5 Experimental results

Dataset

420 abdominal volumetric CT images in portal-venous phases were prepared, including annotations of ten classes (artery, portal vein, liver, spleen, stomach, gallbladder, pancreas, left kidney, right kidney, and background). The slice numbers were between 250 and 2,533 slices with the size of 512×512 pixels per slice. The resolution of each volumetric image was [0.576–0.977, 0.576–0.977, 0.160–1.000] mm. Each volumetric CT image was re-sampled to an isotropic resolution of $1.0 \times 1.0 \times 1.0$ mm^3, and the CT Hounsfield Unit (HU) intensities were re-scaled into $[0, 1]$ with the window range of $[-1,000, 1,000]$ HU. The whole dataset was randomly split into training, validation, and testing sets with 352, 44, and 44 volumetric images, respectively.

Implementation details

The proposed reinforced learning-based AutoML implementation was implemented using NVIDIA Clara Train SDK 3.0 [3]. For one AutoML trial, all experiments employed an NVIDIA Tesla V100 with 32 GB memory, with a total of 8 GPUs running in parallel. A subset of 42 training volumetric images and 5 validation volumetric images were uti-

[3] `https://docs.nvidia.com/clara/clara-train-archive/3.0/index.html/` (Accessed on 2024/1/10)

Table 2.2: Hyperparameters in carefully tuned model (Baseline), standard AutoML (AutoML (rand.)), and AutoML with the proxy data (AutoML (proxy)).

Hyperparameter	Baseline [86]	AutoML (rand.)	AutoML (proxy)
Crop size	$128 \times 128 \times 128$	$128 \times 128 \times 128$	$128 \times 128 \times 128$
Learning rate	0.001	0.000532845	0.00044437
Loss function	Dice	Focal Dice	Focal Dice
Optimization	Adam	Adam	Adam
Probability of adjust contrast	0.5	0.503098488	0.459003747
Probability of scale intensity	0.1	0.424087077	0.418601304
Blocks numbers	[1,2,2,4][1,1,1]	[1,2,2,4][1,1,1]	[1,2,2,4][1,1,1]
Initial kernel number	16	16	16
Normlization	Group norm	Batch norm	Batch norm

lized for hyperparameter searching in AutoML. In the standard AutoML approach, the data used for searching are chosen randomly for a subset, but in the proposed AutoML framework, the top 42 cases as ranked by the proxy data method were chosen.

Parameters searching

The searching space includes the initial learning rate $\mu \in [1e-3, 1e-6]$, crop size $\phi = \{64^3, 80^3, 96^3, 112^3, 128^3\}$, loss function $\chi = \{$Dice loss, Focal loss, Focal Dice loss$\}$, and the optimization method $\varsigma = \{$Adam, NovoGrad, Momentum$\}$. For the network structure, the block numbers of SegResNet in both encoder and decoder phases within the the searching set $\theta = \{[1,1,1,1,1,1,1], [1,2,2,4,1,1,1], [2,2,2,3,2,2,2], [1,2,4,1,1]\}$ were explored. The convolution kernel number $k = \{8, 16\}$ of the first convolution layer and the normalization method $\delta = \{\text{batch}, \text{group}\}$ for each layer were also searched for SegResNet. In terms of the data augmentation, the probability $p \in [0, 1]$ was employed for adjusting contrast and magnitude $m \in [0, 1]$ for scale intensity oscillation for volumetric CT images.

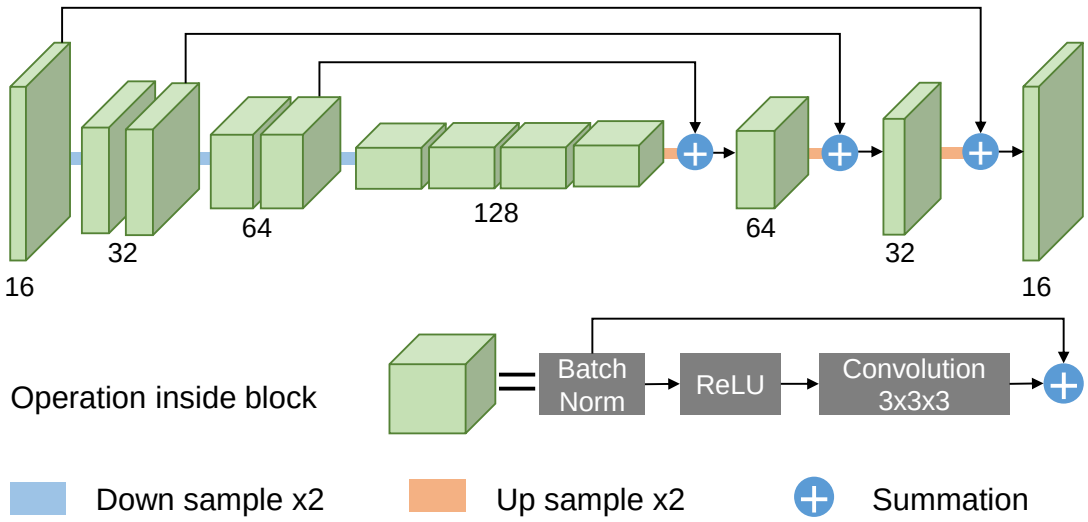


Figure 2.7: Network architecture searched by AutoML. Green boxes represent feature maps and the numbers below the boxes are the number of channels.

Experimental results on multi-organ segmentation

Table 2.2 compares the hyperparameters used for segmentation of the original well-tuned SegRestNet (Baseline), standard AutoML method with the subset selected randomly (AutoML (rand.)) and AutoML with the proxy data (AutoML (proxy)). For all three models, the crop size and optimization method were the same. The network structure searched by AutoML was identical to that of the original SegResNet [86], except for the normalization type. Figure 2.7 shows an illustration of the network architecture obtained by AutoML. The loss function utilized in AutoML was Focal Dice loss, which was different from the baseline Dice loss function.

Table 2.3 compares the DSC by Baseline, AutoML (rand.), and AutoML (proxy). We can see that AutoML helps the achievement of better segmentation performance than the baseline model. When using the randomly selected subset for AutoML, the segmentation accuracy was comparable to the well-tuned SegResNet model. The proxy data improved the overall segmentation accuracy for AutoML, especially for the spleen

Table 2.3: Comparison of multi-class DSC of each organ by using carefully tuned model (Baseline), standard AutoML (AutoML (rand.)), and AutoML with the proxy data (AutoML (proxy)). Similar Dice scores show the effectiveness of AutoML. The best score for each organ is shown in **bold**.

Organ	Baseline [86]	AutoML (rand.)	AutoML (proxy)
Artery	85.6%	85.0%	**86.0%**
Portal Vein	**68.6%**	67.9%	68.3%
Liver	95.7%	95.9%	**96.0%**
Spleen	92.1%	95.0%	**95.1%**
Stomach	85.0%	87.1%	**87.6%**
Gallbladder	84.0%	83.4%	**85.7%**
Pancreas	78.0%	78.8%	**79.2%**
Left kidney	**81.5%**	80.9%	81.4%
Right kidney	93.8%	93.3%	**94.1%**
Average	84.9%	85.2%	**85.9%**

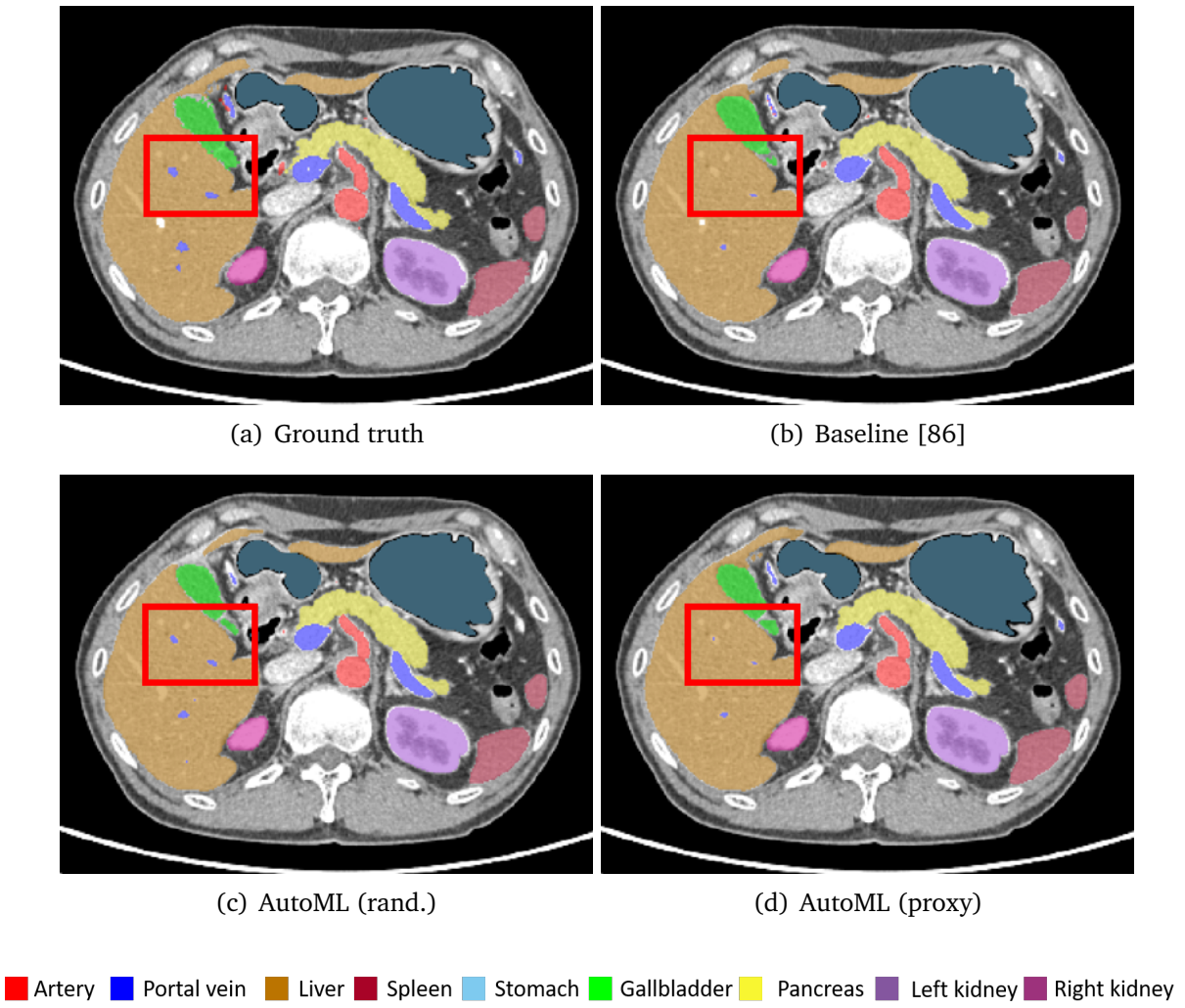
and stomach.

Fig. 2.8 and Fig. 2.9 show examples of segmentation in axial and coronal views. AutoML improved the segmentation performance of small blood vessels and stomach regions. The proxy data showed a positive influence on the gallbladder segmentation.

2.4.6 Discussion

The main contribution of this work was to introduce a practical application of AutoML for multi-organ segmentation, which helps to fully boost the deep-learning-based network performance with a little machine learning experience. The proxy data was selected automatically in the first step to represent the whole dataset, instead of a randomly chosen subset as in the standard AutoML approach. Therefore, the run-time was vastly reduced compared to performing AutoML on the full data while also effectively finding a suitable set of hyperparameters for multi-organ segmentation.

Experimental results demonstrated the benefits of the proposed approach. Baseline and AutoML (rand.) showed similar average DSC of 84.9% and 85.2%, respectively.

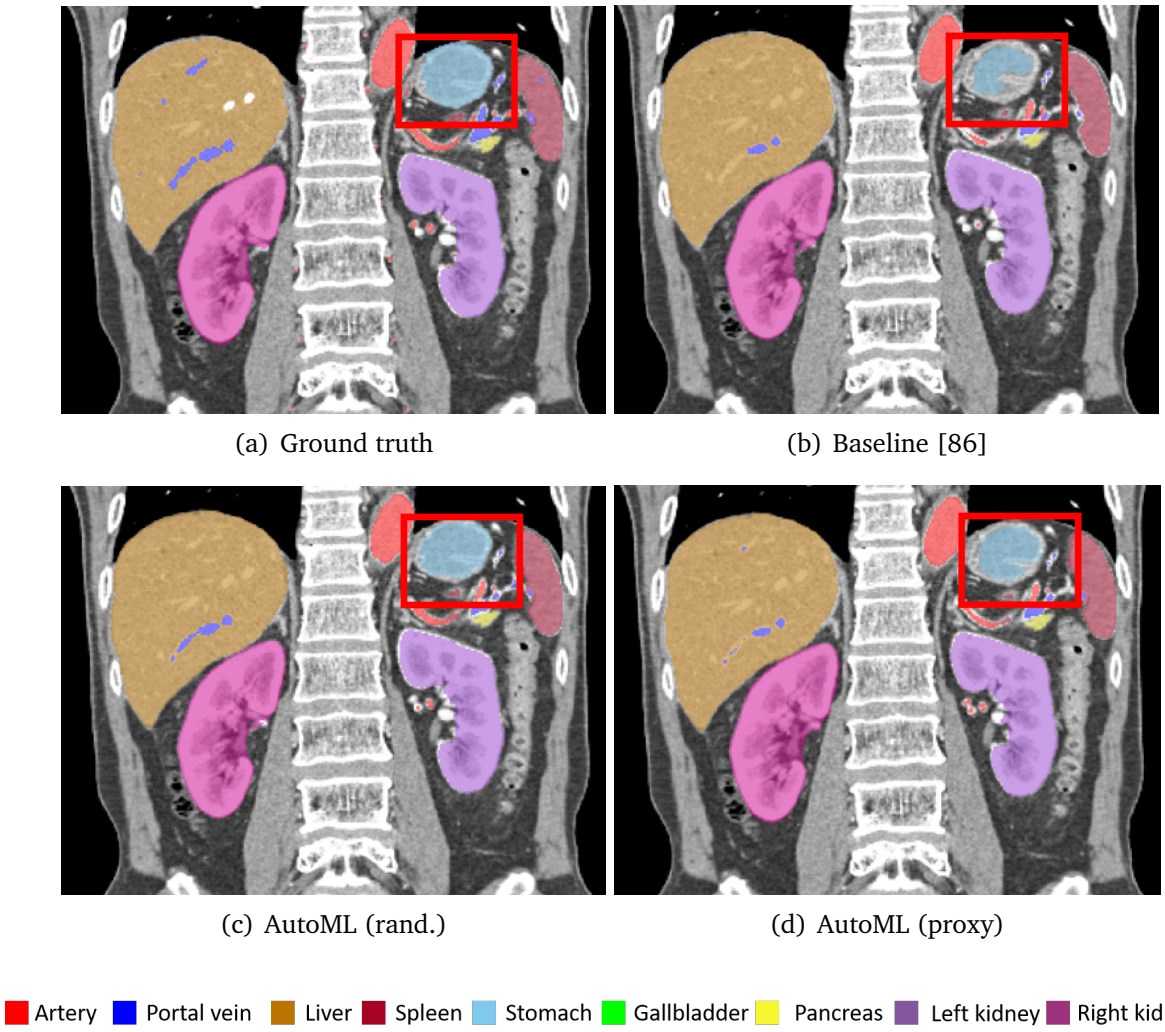


(a) Ground truth (b) Baseline [86]

(c) AutoML (rand.) (d) AutoML (proxy)

Figure 2.8: Comparison of segmentation results in axial slice. Regions in red rectangles show the effectiveness of using AutoML for organ segmentation.

AutoML eliminated the importance of manually constructing the network architecture and tuning the hyperparameters, demonstrating the benefits of utilizing it for multi-organ segmentation. The average DSC improved by about 1% after employing the proxy data strategy compared to using the randomly selected subset. For the spleen, when using the proxy data for AutoML, the DSC was 3% higher than the baseline,



(a) Ground truth (b) Baseline [86]

(c) AutoML (rand.) (d) AutoML (proxy)

Figure 2.9: Comparison of segmentation results in coronal slice. Regions in red rectangles show the effectiveness of using AutoML for organ segmentation.

despite the fact that the network structures of the baseline and AutoML were almost identical. AutoML significantly improved the network performance by assigning the proper hyperparameters of pre-processing transforms and augmentations.

2.5 Conclusion

This chapter proposed two separate segmentation methods for the segmentation of a healthy pancreas and its multiple associated organs segmentation. In the first method, a spatial-embedded FCN was introduced to enhance the segmentation accuracy, especially for organs with small voxel sizes. The spatial information showed a positive influence on organs with small voxel sizes. Additionally, an improved data augmentation method was implemented, including original non-rigid data augmentation and gamma correction, to mitigate overfitting during training. While data augmentation might provide less assistance for datasets from the same hospital and CT scanner on most organs, it showed improvement in pancreas segmentation. In the second method, an efficient application of AutoML was presented, simplifying the development of multi-organ segmentation for volumetric CT images. Proxy data, effectively representing the entire training dataset were utilized, for network structure and hyperparameter searching in AutoML instead of randomly selecting data. Experiments demonstrated that proxy data positively influenced boosting the final AutoML model's performance. Although the two methods cannot be directly compared due to differences in training data, they provide two dominant viewpoints for organ segmentation using supervised methods: one utilizing traditional machine learning and the other employing AutoML.

Chapter 3

Segmentation of pancreas and pancreatic tumor

3.1 Background

Segmentation of the healthy pancreas and its multiple associated organs from volumetric Computed Tomography (CT) images was introduced in Chapter 2. In this chapter, attention turns to a pancreatic pathology known as pancreatic cancer, recognized as one of the deadliest types of cancer. The main segmentation target in this chapter is the pancreas and pancreatic tumors. Automated segmentation of pancreatic tumors plays a crucial role in the development of Computer-Aided Diagnosis (CAD) systems. The main challenge arises from the poor contrast of the tumor boundary and significant appearance variations among individuals and different cancer stages, as highlighted in [89]. In clinical settings, a CAD system for pancreatic cancer detection not only requires high accuracy but also needs to demonstrate high generalizability to handle different types of pancreatic tumors.

Deep Neural Networks (DNNs)-based methods have recently been widely used for

medical image analysis research. High-performing clinically useful models always require vast, varied, and high-quality datasets. One limitation of pancreas segmentation using deep learning-based approaches is the lack of annotated training data. Medical images from a single resource can be biased toward specific pathologies, equipment, acquisition protocols, and patient populations. The low generalizability of DNN models trained on insufficient datasets is critical when applying deep learning methods for clinical usages. Distinct from natural images, collecting extensive training data from various resources may lead to multiple technical, legal, and privacy issues in healthcare applications.

To improve the robustness when dealing with limited data, fine-tuning is an alternative way to learn the knowledge from pre-trained DNNs. The fine-tuning technique starts training from a pre-trained weight instead of random initialization, which has been proved helpful in medical image analysis [90, 91], which exceeds the performance on training a DNN from scratch. However, fine-tuned models can still have high deficiencies in generalizability [92]. When a model is pre-trained on one dataset (source data) and subsequently fine-tuned on another dataset (target data), the trained model often adapts well to the target data but may lose its ability to represent the source data, as noted by Li et al. [93]. Federated Learning (FL) provides an innovative solution to tackle all the aforementioned issues. It is a technique that allows for the collaborative learning of a deep learning-based model among distributed devices, referred to as clients, without the need to centralize the training data in a single location, as outlined by McMahan et al. [94]. In FL, clients might have different categories of labels represented in their data. For example, one client might have patient data with “healthy” pancreases only while datasets from other clients may contain cases with pancreatic tumors. When working with different types of training data, the problem of heterogeneity can pose a substantial impact on the model’s performance when it is aggregated on the server.

In this chapter, efforts are made to investigate an FL framework to effectively address the issue of heterogeneity using three distinct pancreatic datasets. For the pancreas segmentation task, three publicly annotated datasets are utilized to represent three heterogeneous clients. One of the datasets consists of both pancreas and pancreatic tumors, while the other two exclusively consist of healthy pancreas cases. Two heterogeneous optimization methods are proposed for FL model aggregation. Experimental results demonstrate notable improvements in the automated segmentation of pancreas and pancreatic tumors in abdominal CT images with FL settings.

3.2 Related works

Thanks to the rapid development of deep learning, significant progress has been achieved in enhancing pancreas segmentation performance over the past decade, largely owing to the utilization of Convolutional Neural Network (CNN)-based approaches [95–98]. When addressing both pancreas and pancreatic tumors, single-agent learning encounters a limitation due to the constraints of the training dataset, primarily because of the significant variations of tumors. However, in the medical image analysis field, sharing the training dataset among different clients is difficult. FL is a technique that enables collaborative model training using datasets from different sites without the necessity of sharing the actual dataset. It has demonstrated significant effectiveness on segmentation tasks in abdominal organs [99, 100], brain tumors [101, 102], and COVID-19 image analysis [103–106], both in simulation and real-world FL applications.

In FL, a central server aggregates models trained on different datasets. The most common method used for this aggregation is Federated Averaging (FedAvg). It combines models from each client and updates the global model on the server using a weighted sum. These weights are typically determined based on the number of the local training data and remain constant throughout the training process [94]. Meanwhile,

FedProx is a well-known model aggregation method introduced to address data heterogeneity in FL. It achieves this by incorporating a regularization loss on the client that penalizes deviations from the current global model [107]. Prior studies have suggested that techniques developed for multi-task learning can serve as an alternative approach for addressing heterogeneous statistical challenges [108]. Building on this inspiration, the primary motivation of this work is to investigate effective model aggregation methods to address the issue of heterogeneity.

3.3 Contribution

This study utilizes three publicly annotated datasets for pancreas segmentation, with each dataset representing a heterogeneous client in the context of FL. Building upon these datasets, two FL optimization methods are introduced, inspired by multi-task learning principles. The main contribution can be summarized as follows:

1. The first method is Dynamic Task Prioritization (DTP), where priority is assigned to each task in multi-task learning using task-specific metrics. It allocates a higher prioritization to the client with superior performance. The method explores the scenario where a client model with superior performance contributes more to the server.

2. The second method is Dynamic Weight Averaging (DWA). In traditional FL, the model from each client is aggregated based on a fixed weight, typically calculated as the proportion of the client's training data size relative to the total data. In contrast, DWA calculates the weight for each client dynamically at each round, allowing the adaptation of weights while keeping the data distribution fixed.

3. The validation of the proposed methods is performed using three publicly annotated datasets for pancreas segmentation. It is worth noting that these datasets

exhibit some variations. One of the datasets comprises cases involving both pancreas and tumors, while the remaining two datasets exclusively include cases of healthy pancreas. The effectiveness of the enhancements are compared in comparison to FedAvg and FedProx for pancreas and pancreatic tumor segmentation tasks.

3.4 Federated learning

FL can be categorized into three main types based on the distribution characteristics of data, as discussed by Yang et al. [109]: horizontal FL, vertical FL, and federated transfer learning. In horizontal FL, datasets from each client share the same feature dimensions but contain different examples. In vertical FL, datasets from each client share the same examples but differ in their feature dimensions. In federated transfer learning, both samples and feature dimensions are distinct for each client.

A standard FL system comprises a server and multiple clients, as illustrated in Fig. 3.1. The server manages the training process and generates a *global model*, and the client trains with local data to produce a *local model*. The server receives trained weights from each client and aggregates them into the global model. The clients train with the local dataset and send the weights to the server. The process of generating one global model is counted as one round.

A typical FL workflow involves the following sequential steps:

1. **Server setup:** The process is initiated by launching the server. On the server side, Google Remote Procedure Call (gRPC) communication ports and Secure Sockets Layer (SSL) certificate are configured, and the allowable ranges for the minimum and maximum numbers of clients are defined.

2. **Client initialization**: Each client is started using client-side configuration settings

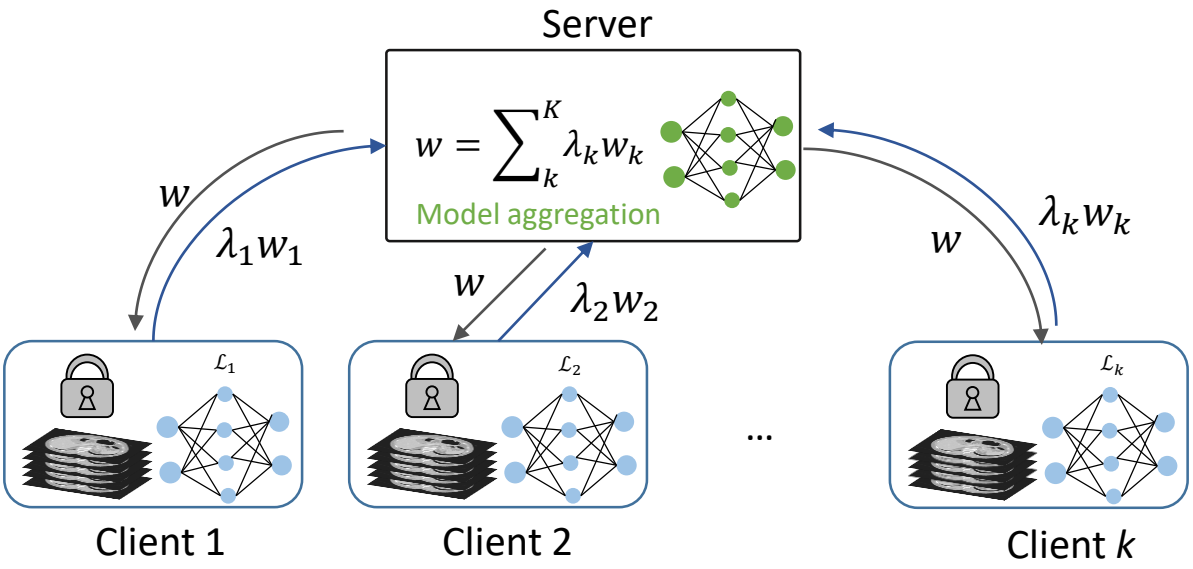


Figure 3.1: Illustration of FL in medical image analysis. The server only receives model updates and the training data stays on the client sites privately.

for initialization and authenticated by sending a login request to the server.

3. **Local training:** The current global model is downloaded from the server to the client and fine-tuned by local datasets. After fine-tuning, only the trained model is sent from each client to the server and contributions from other clients are waited.

4. **Model aggregation:** Once the updated models are received from a predefined minimum number of clients, they are aggregated into a new global model.

5. **Global model update:** The global model in the server is updated and one round is finished.

6. **Repeat:** Return to Step 3 to initiate another round of FL.

The models shared among the server and clients are only weight parameters, protecting the privacy of local data. To establish trust between the server and clients, the

server employs tokens throughout the process. Additionally, gRPC communication ports and SSL certificate authority are consistently utilized in FL to enhance security.

3.5 FL model aggregation

In a FL round, each client receives the global model from the server and fine-tunes it on their local dataset. Subsequently, the client shares a weight update with the server after completing the local training. A standard FL aims to minimize an overall loss, which can be defined as:

$$\mathcal{L} = \min \sum_{k}^{K} \eta_k \mathcal{L}_k, \tag{3.1}$$

where the k-th client tries to optimize the local loss function $\mathcal{L}_k$. The total number of clients is K and the proportion that each client contributes to the global model update is $\eta_k \geq 0$, where $\sum_k^K \eta_k = 1$.

This work aims to investigate an FL optimization method for the challenging task of heterogeneous pancreas segmentation, where clients possess diverse types of images and labels in their datasets.

In this section, firstly, two widely adopted FL methods, FedAvg [94] and FedProx [107] are revisited. Then, two optimization methods are adapted from the multi-task learning literature to the FL framework: DTP and DWA.

3.5.1 FedAvg

FedAvg is the most employed aggregation method in FL. In a standard FL scenario, when the FL system comprises a substantial total of N clients, it employs the practice of selecting only a subset of K clients (where $K \ll N$) for global model updates in each round. This selection is a strategy aimed at mitigating the significant communication costs and effectively managing potential client dropouts [94].

In the FedAvg scheme, the model aggregator conducts a weighted average of the model parameters from each client. These weights can be calculated by considering factors like the proportion of the size of training datasets held by each client or set manually as needed. If the weights are not specified manually, the weight η_k of client $k \in K$ is typically a constant number that can be defined as:

$$\eta_k = \frac{n_k}{n}, \tag{3.2}$$

where n_k is the number of local training data in client k. The total number of training data in all clients can be derived as $n = \sum_k^K n_k$. In this way, clients with larger local training datasets contribute more significantly to the updated global model.

3.5.2 FedProx

FedProx is an improved federated optimization algorithm proposed to address the challenges of learning from distributed and heterogeneous datasets [107], which is an extension of the standard FedAvg scheme. The FedProx algorithm adopts the aggregation scheme and adds another learning constraint for each client, namely regularization. The regularization term can help the local client model stay close to the server model. The local client tries to minimize

$$\hat{\mathcal{L}}_k = \mathcal{L}_k + \frac{\mu}{2} \left\| \boldsymbol{w}_k - \boldsymbol{w} \right\|^2, \tag{3.3}$$

where $\hat{\mathcal{L}}_k$ specifies the learning target of client k, and $\boldsymbol{w}_k$ stands for the local model parameters. The $\boldsymbol{w}$ is the model parameter from the FL global model, and $\|\cdot\|^2$ indicates the L2 normalization.

This learning constraint ensures the consistency of gradients from different clients. A more consistent gradient can prevent model divergence of client models and improve

the convergence of the global model. FedProx is designed to achieve a balance between global model aggregation and client-specific model stability, particularly when dealing with heterogeneous data.

3.5.3 Dynamic task prioritization

Dynamic Task Prioritization (DTP) allows for the adjustment of weights between different tasks by estimating the Key Performance Index (KPI) κ [110] in multi-task learning. KPI κ is a monotonic increasing function ranging from 0 to 1; a larger value of κ stands for better performance of the specific task. DTP concentrates on challenging tasks by increasing corresponding weights and lowering the weights of easier tasks. This technique enables the handling of the imbalance problem associated with different task difficulties. It ensures that the training process prioritizes challenging tasks while allocating less attention to easier ones.

In standard FL, each client's prioritization is uniform. In this work, the DTP is generated for FL by considering each client as a different task. This enables the dynamic specification of weights for each client during the training process based on task difficulty. Specifically, the KPI of client k could be defined as:

$$\kappa_{k,i} = \mathcal{L}_{k,i}^{r}, \tag{3.4}$$

where $\mathcal{L}_{k,i}^{r}$ indicates the loss value $\mathcal{L}$ in iteration i of round r. The weight of client k can be defined as:

$$w_{k,i} = -(1 - \bar{\kappa_{k,i}})^{\gamma} \log \bar{\kappa_{k,i}}, \tag{3.5}$$

where $\bar{\kappa_{k,i}}$ is an exponential average, which is used to stabilize the weights between

each batch. The exponential average $\bar{\kappa_{k,i}}$ can be derived as:

$$\bar{\kappa_{k,i}} = (1-\alpha)\,\kappa_{k,i} + \alpha\bar{\kappa_{k,i-1}}, \tag{3.6}$$

where α is a value between 0 and 1 and γ is a tunable hyperparameter.

3.5.4 Dynamic weight averaging

Dynamic Weight Averaging (DWA) tries to optimize the FL procedure by focusing on server model aggregation instead of applying a constraint on the loss function. This method is inspired by optimization approaches from classical multi-task learning methods [111]. In FL, finding a suitable balance to aggregate the model updates from heterogeneous clients is challenging. The weights of each client for global model aggregation in FedAvg is based on the local iteration numbers as stated in 3.5.1. An additional constant weight can be assigned for each client as well. However, specifying the proper weight requires a large number of experiments and priority knowledge. For DWA, this work investigates a method that defines the client weights on each round automatically. The server learns to weigh each client based on the variation of loss values from the previous round. The weight of client k in round r can be defined as:

$$\lambda_{k,r} = \frac{\xi \exp(\rho_{k,r-1}/T)}{\sum_{i=1}^{K} \exp(\rho_{i,r-1}/T)}, \tag{3.7}$$

where $\rho_{k,r-1} \in (0, +\infty)$ represents the dynamic proportion of the loss value $\mathcal{L}$ changes in client k from round $r-2$ to round $r-1$, which can be defined as:

$$\rho_{k,r-1} = \frac{\mathcal{L}_{k,r-1}}{\mathcal{L}_{k,r-2}}. \tag{3.8}$$

To control the effectiveness of dynamic proportion, T is defined similarly to [111]. When $T \to +\infty$, the weight of each client tends to be $\rho_k \to 1$. Here $\xi \in \mathbb{N}$ is introduced to adjust the impact of weights in DWA. Different from the way to calculate loss value $\mathcal{L}_{k,r}$ in [111], the local loss value of each iteration is averaged in each round as:

$$\mathcal{L}_{k,r} = \frac{1}{J} \sum_{j=1}^{J} \mathcal{L}_{k,j}, \tag{3.9}$$

where j is the local iteration number within the total J iterations. The averaging will make the loss value of each round more stable. For the first round ($r = 1$), $\mathcal{L}_{k,r-1}$ and $\mathcal{L}_{k,r-2}$ are initialized to 1, so that $\rho_{k,1} = 1$ could be obtained.

3.6 Experiments

Experiments on three public datasets were conducted, evaluating standalone models trained individually on each dataset, as well as the FL server model with various aggregation methods.

3.6.1 Datasets

This experiment assumed one federated server for model aggregation and three clients for local training. The server did not own any validation datum and only aggregated the client's model parameters. Each client owned a dataset from a different source. Three publicly available datasets were employed for pancreas segmentation tasks. It is important to note that one dataset was sourced from patients with pancreatic cancer, where the segmentation targets included both pancreas and pancreatic tumor segmentation. In contrast, the other two datasets were derived from cases with healthy pancreas, exclusively intended for pancreas segmentation.

Table 3.1: Data split details for each dataset.

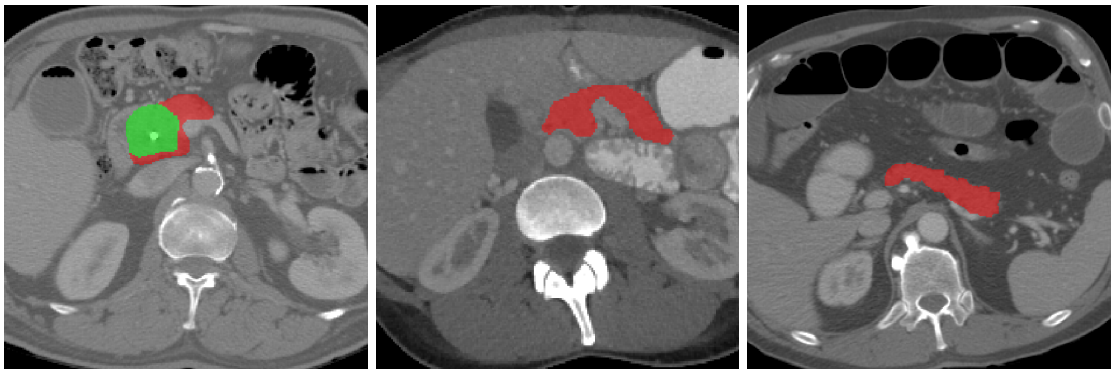
Dataset	Training	Validation	Testing	Total
MSD [112]	165	58	58	281
TCIA [95]	48	17	17	82
Synapse [113]	18	6	6	30

The first dataset was the Task07 pancreas segmentation dataset from the Medical Segmentation Decathlon (MSD) challenge[1] [112]. This dataset contained 281 portal venous phase CT scans with manual labels for the pancreas and pancreatic tumors (intraductal mucinous neoplasms, pancreatic neuroendocrine tumors, or Pancreatic Ductal AdenoCarcinoma (PDAC)). The second dataset was The Cancer Imaging Archive (TCIA) dataset from The National Institutes of Health Clinical Center [2] [95]. This dataset contained 82 abdominal contrast-enhanced CT scans with manual segmentation labels for the healthy pancreas. The third dataset was from the MICCAI Multi-Atlas Labeling Beyond the Cranial Vault challenge (Synapse)[3] [113]. This dataset contained 30 portal venous contrast phase CT scans with manual labels for 13 abdominal organs including the pancreas. This dataset specifically retained the pancreas labels, while the labels for the other 12 organs were discarded. The three datasets were independently shuffled randomly, and split into training, validation, and testing sets with a ratio of 60%, 20%, and 20%, respectively. Among the total 231 training cases, 165 cases had both pancreas and pancreatic tumor labels. The specification of each subset is shown in Table 3.1. Visual examples of the three datasets are shown in Fig. 3.2.

[1] `http://medicaldecathlon.com/` (Accessed on 2024/1/10)

[2] `https://wiki.cancerimagingarchive.net/display/Public/Pancreas-CT/` (Accessed on 2024/1/10)

[3] `https://www.synapse.org/#!Synapse:syn3193805/wiki/217785/` (Accessed on 2024/1/10)



(a) MSD [112] (b) TCIA [95] (c) Synapse [113]

Figure 3.2: Visual examples of the datasets. The pancreas regions are shown in red, while the pancreatic tumor region is shown in green.

3.6.2 Implementation details

NVIDIA Clara Train SDK 3.1[4] was used as the FL framework. As this was a collaborative effort between Nagoya University and National Taiwan University, the experiments were conducted using two types of deep learning machines. Specifically, for DWA, a DGX-1 with 40 CPU cores, 512GB system memory, and 8 V100 32GB GPUs was utilized at Nagoya University. Meanwhile, for DTP, a DGX station equipped with 20 CPU cores and 256GB system memory was employed at National Taiwan University. During the FL simulation, the server and associated clients were physically on the same machine but running in individual Docker containers. This setup in Docker containers was consistent across all experiments, ensuring that the results were not influenced by using different machines. The server had no access to a GPU, while each client was assigned one V100 32GB GPU. All volumetric CT images were resampled to isotropic spacing with $1.0 \times 1.0 \times 1.0\ \mathrm{mm}^3$. To ensure that the volumetric CT images were in the same orientation, the voxel axes were arranged as close as possible to the RAS+ orientation. The Hounsfield Unit (HU) intensity in the range $[-200, 250]$ HU was rescaled and clipped

[4] `https://docs.nvidia.com/clara/clara-train-archive/3.1/index.html/` (Accessed on 2024/1/10)

into $[-1, 1]$. A network found by Coarse-to-Fine Architecture Search (C2FNAS) [114] using a TensorFlow implementation was employed in all experiments. The training loss function was the sum of Dice loss and cross-entropy. The Adam optimizer [81] with cosine annealing learning rate scheduler was adopted with the initial learning rate of 5×10^{-4}. The input patch size of this network was $96 \times 96 \times 96$. The total round number of FL was set to 60 with a local epoch number of 10. The minimum client number was 3. Despite running FL in simulation on public datasets, a percentile sharing protocol was employed as a privacy-preserving measure [102]. Only 25% of the model updates with the largest absolute values were shared, a measure taken to ensure the approach's applicability in real-world settings.

3.6.3 Results on pancreas and pancreatic tumor segmentation

The experimental results included the standalone models trained individually on each dataset (MSD local, TCIA local, and Synapse local), as well as the FL global best models for FedAvg, FedProx, DTP, and DWA. The global best models were determined using the average client validation scores throughout each FL round. Table 3.2 compares the Dice Similarity Coefficient (DSC) across all experiments with different hyperparameter settings. For the standalone training model, the performance on other datasets was quite unsatisfactory. Specifically, the local model trained on the Synapse dataset exhibited poor performance on the other two datasets. FL global models have markedly better generalizability than standalone models. The average DSC for each client was 56.9% for FedAvg, and 58.1% for FedProx. Both DTP and DWA methods relied on precise hyperparameter settings; DTP performed optimally with $\gamma = 1$, $\alpha = 0.5$, and $r = 1$, achieving an average DSC of 46.1%. DWA delivered its best performance when $T = 2$ and $\xi = 2$, achieving a DSC of 61.4%. FedAvg achieved the highest DSC on the MSD dataset, but the performance was not ideal on the Synapse dataset. Meanwhile, the

Table 3.2: Comparison of DSC for the pancreas and pancreatic tumor segmentation on *local models* trained from scratch with single datasets (MSD, TCIA and Synapse); and on *FL server best global model* with FedAvg, FedProx, DTP, and DWA optimization. The best scores of FL server model are shown in **bold**, while the performance of the local model on its individual test data are shown in *italic*. Non-italic numbers show the lack of generalizability of local models evaluated on other clients' test data.

	MSD		**TCIA**	**Synapse**	**All**
	Pancras	**Tumor**	**Pancreas**	**Pancreas**	**Average**
MSD local ($N_{\text{train}} = 165$)	*77.8%*	*31.1%*	61.9%	4.4%	*40.3%*
TCIA local ($N_{\text{train}} = 48$)	71.8%	0.0%	*79.4%*	5.0%	40.1%
Synapse local ($N_{\text{train}} = 18$)	0.4%	0.0%	9.8%	*61.1%*	23.7%
FedAvg [94]	**75.1%**	**20.2%**	80.6%	42.6%	56.9%
FedProx [107]	75.0%	19.5%	80.6%	47.6%	58.5%
DTP ($\gamma = 1, \alpha = 0.9, r = 1$)	54.8%	14.4%	64.1%	34.5%	44.4%
DTP ($\gamma = 2, \alpha = 0.9, r = 1$)	57.3%	12.4%	46.0%	27.6%	36.1%
DTP ($\gamma = 1, \alpha = 0.5, r = 1$)	57.2%	12.4%	64.4%	39.2%	46.1%
DTP ($\gamma = 1, \alpha = 0.5, r = 2$)	55.7%	12.4%	64.3%	39.3%	45.9%
DWA ($T = 1.0$)	59.8%	9.5%	65.3%	49.8%	49.9%
DWA ($T = 1.5$)	71.7%	13.7%	76.1%	53.0%	57.3%
DWA ($T = 2.0$)	72.4%	6.8%	78.2%	56.2%	58.0%
DWA ($T = 2.0, \xi = 2$)	73.4%	13.9%	**80.9%**	**59.6%**	**61.4%**
DWA ($T = 2.0, \xi = 3$)	59.8%	7.7%	68.7%	39.3%	47.3%

average DSC with FedProx improved over FedAvg. The TCIA and Synapse datasets have shown the highest DSC with DWA ($T = 2, \xi = 2$). The average DSC with DWA was 4.5% and 2.9% higher than FedAvg and FedProx, respectively.

Axial visualizations of the segmentation results for standalone training are shown in Fig. 3.3, while the segmentation results of FL are shown in Fig. 3.4. The standalone models only showed acceptable performance on their respective datasets. Because the TCIA and Synapse datasets exclusively contained healthy pancreas cases, their standalone models were incapable of segmenting pancreatic tumor regions. FL models exhibited significantly better generalizability across all three datasets compared to standalone models. Both FedAvg and FedProx yielded similar performance, with acceptable segmentation results on MSD and TCIA datasets but suboptimal results on the smaller Synapse dataset. However, the DWA method showed a more stable performance on the three different datasets.

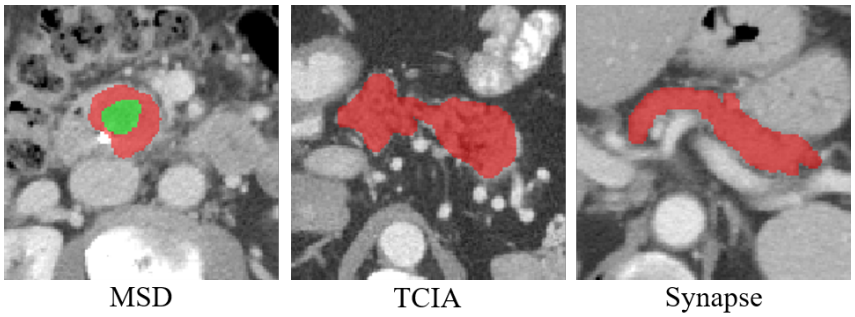
3.6.4 Visualization of the weights

The dynamically chosen weights of each client, visualized using the DTP and DWA approaches, are presented in Fig. 3.5. The weights of clients with the TCIA dataset, MSD dataset, and Synapse dataset are represented in red, blue, and green, respectively. DTP maintained a consistent weight across the FL rounds, whereas DWA exhibited significant weight fluctuations.

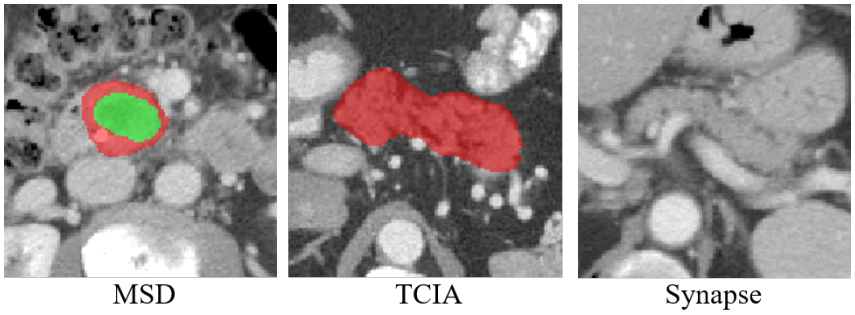
3.7 Discussion and conclusion

3.7.1 Discussion

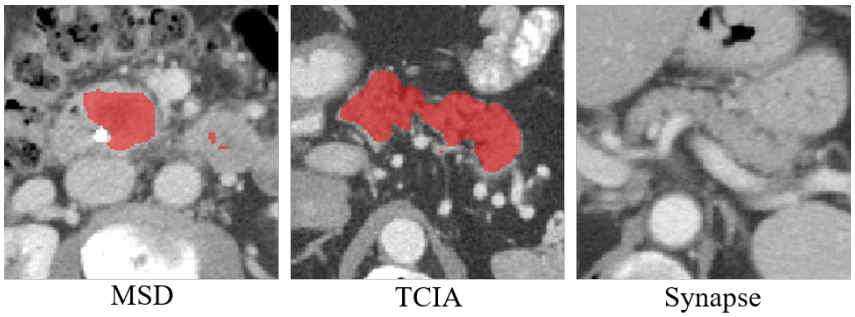
In Table 3.2, FedAvg was the standard FL baseline to be compared with other methods. Three local models were standalone training results for their corresponding datasets.



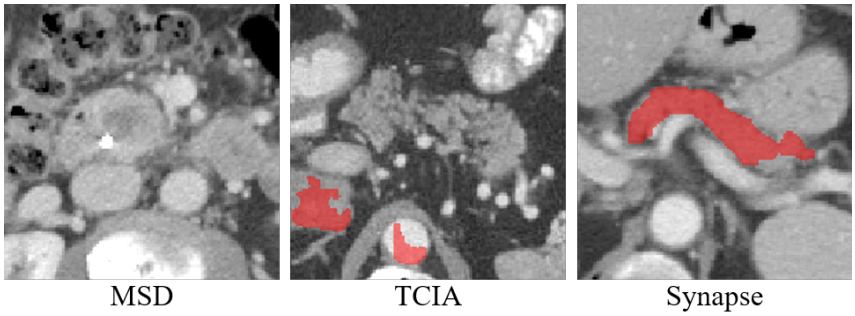
(a) Ground truth



(b) MSD

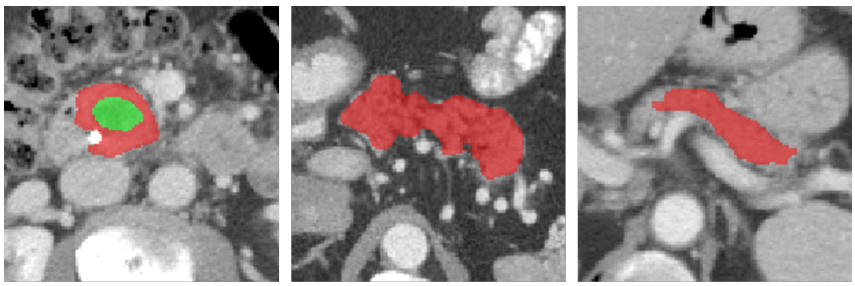


(c) TCIA

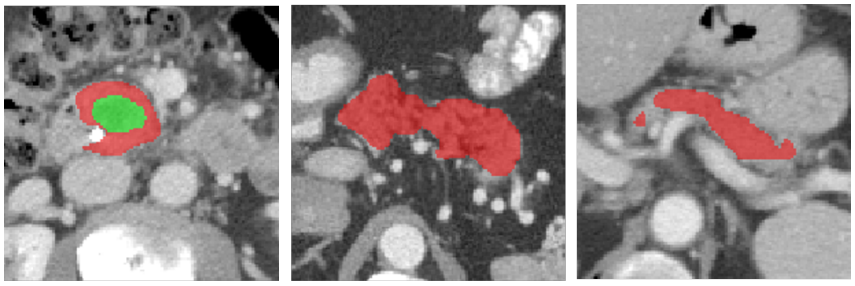


(d) Synapse

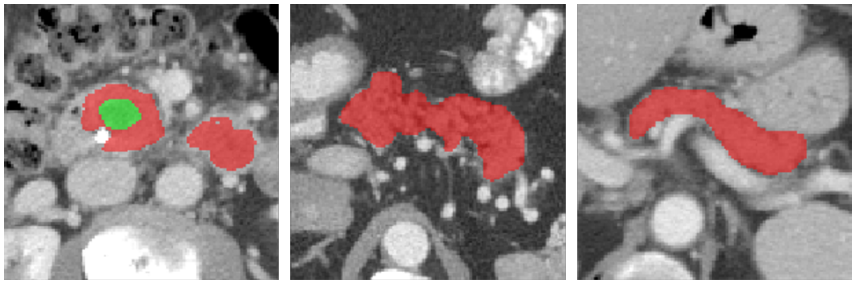
Figure 3.3: Axial visualization of pancreas (red) and pancreatic tumor (green) segmentation results obtained using standalone training models.



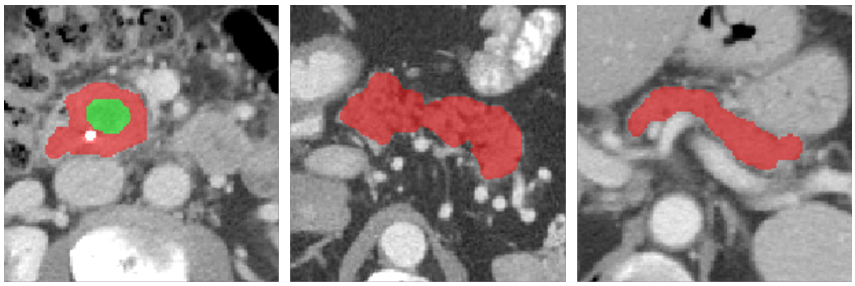
(a) FedAvg



(b) FedProx

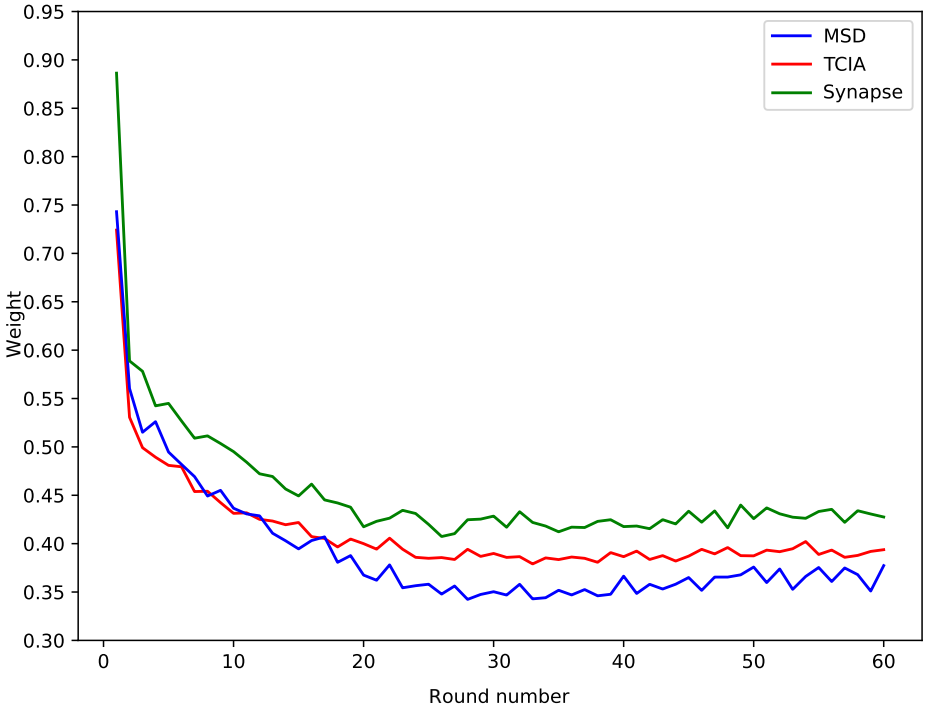


(c) DTP

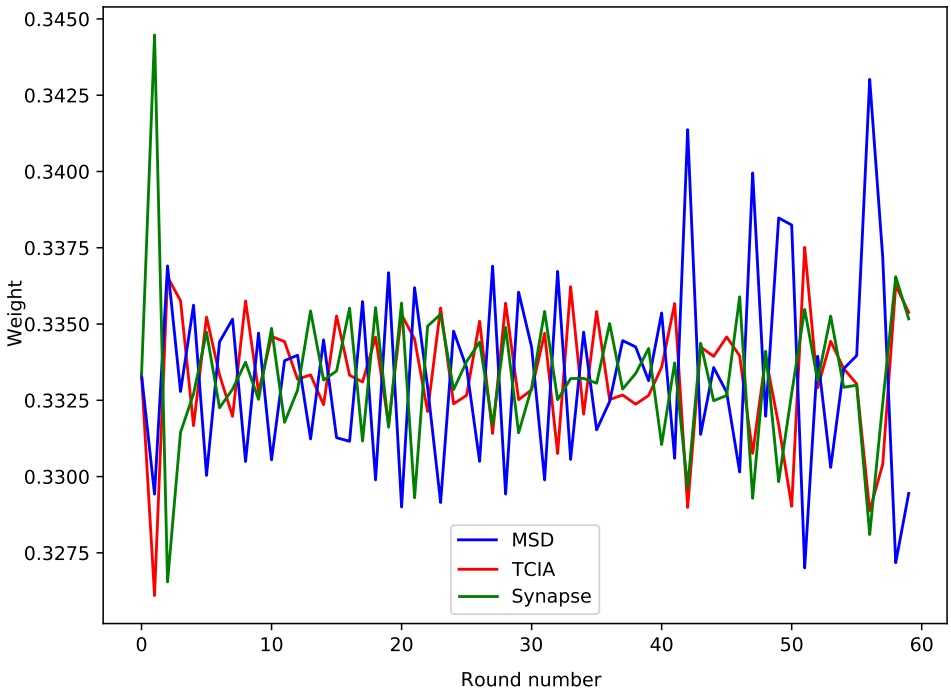


(d) DWA

Figure 3.4: Axial visualization of pancreas (red) and pancreatic tumor (green) segmentation results obtained from different FL model aggregation methods.



(a) DTP



(b) DWA

Figure 3.5: Client's weight chosen every round.

The global model resulting from FedAvg performed well for the TCIA dataset, and for the MSD pancreas dataset comparable to local models. Moreover, for the MSD tumor and Synapse datasets, although the performance was not as high as corresponding local models, there was still a significant improvement to other local models, indicating the improved generalizability of the global models. The FedProx model showed similar performance as the FedAvg model. For the MSD dataset, the average DSC of the pancreas and pancreatic tumor dataset was slightly lower than the result of the FedAvg model. However, the average DSC of the Synapse dataset was significantly higher than the result of the FedAvg model. The performance of DTP models was generally lower than the FedAvg baseline and DWA results. Nonetheless, in most settings, DTP models still outperformed the local models. Furthermore, compared to the DWA results, the performance on MSD tumors dataset was more consistent.

Both FedAvg and FedProx are commonly used in FL, and the experiments suggest that both methods already provided a strong baseline performance even on heterogeneous datasets.

In DTP, the dynamic prioritization weight focused on the most challenging tasks by adjusting the magnitude of the loss. However, each client only calculated the prioritization weights using local batch data. The lack of a global perspective of the training can therefore limit the performance of DTP. Also, DTP scaled the magnitude of the loss, disrupting the optimization and increasing the need for further hyperparameter tuning. In contrast, with most DWA configurations, the Synapse dataset's performance was markedly higher than the FedAvg baseline. The results showed that DWA can outperform both FedAvg and FedProx on average with properly selected hyperparameters.

3.7.2 Conclusion

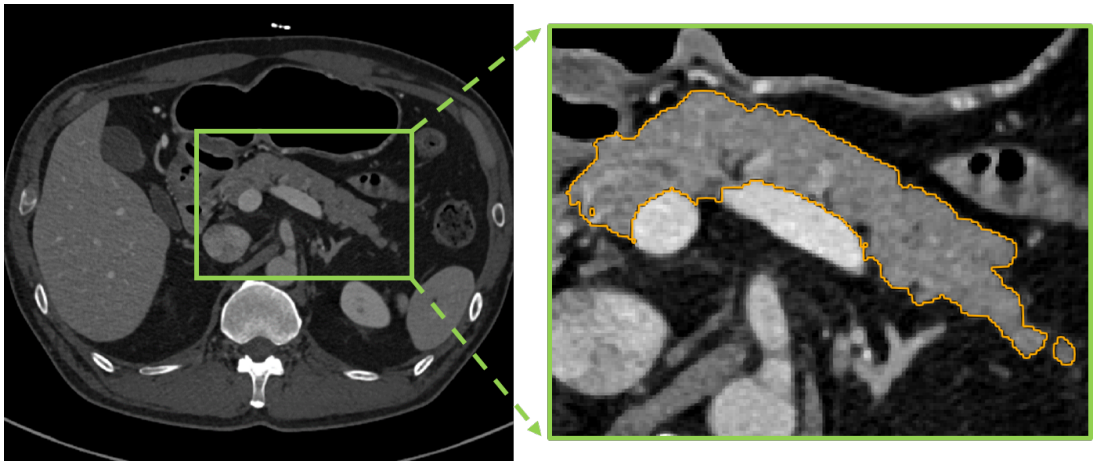
This chapter investigated two multi-task optimization methods for FL in medical image analysis with heterogeneous datasets: DTP and DWA. The application of both methods was inspired by the analogy and similarity between FL and multi-task learning. Each method was evaluated within an FL framework, and the performance of the global model was compared with FedAvg and FedProx. The DSC of DTP was lower than FedAvg and FedProx, likely because of limited manual tuning. However, the global model from DTP still outperformed the local models. DWA model aggregation method showed significant improvement, especially on the Synapse client whose training data was relatively smaller than the other two clients.

Chapter 4

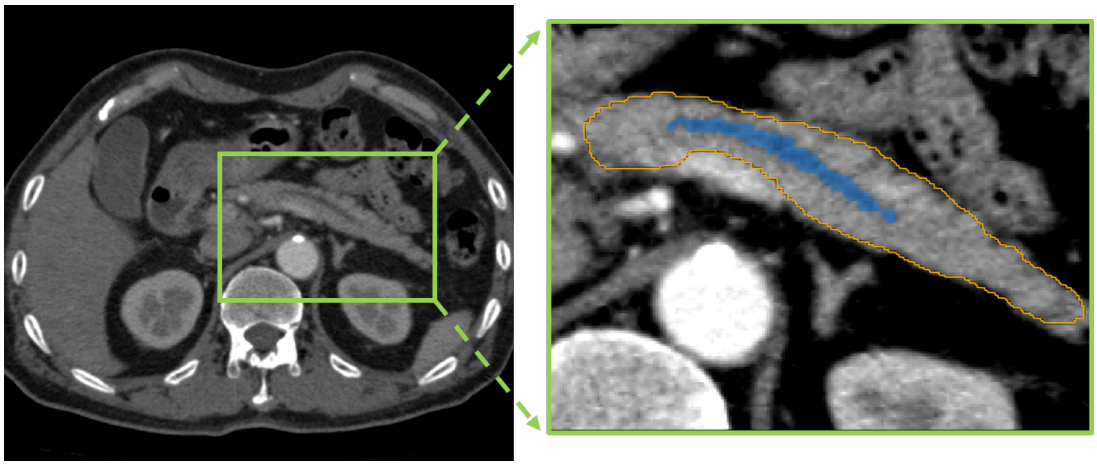
Dilated pancreatic duct segmentation

4.1 Background

The research on the segmentation of pancreas and pancreatic tumor from patients diagnosed with pancreatic cancer was introduced in Chapter 3. This chapter concentrates on individuals in the pre-pancreatic cancer stage. As introduced in Chapter 1, Pancreatic Ductal AdenoCarcinoma (PDAC) which develops in the main pancreaticduct, accounts for more than 90% of pancreatic cancer [59]. Detecting pancreatic cancer is challenging due to its mild symptoms, often going unnoticed until it reaches an advanced stage. Several clinical studies suggest that dilatation of the main pancreatic duct indicates an increased risk of pancreatic cancer [60–62]. Therefore, the appearance of pancreatic duct dilatation may serve as a useful entry point for diagnosing pancreatic cancer. In typical medical settings, Computed Tomography (CT) is the more commonly employed diagnostic imaging technique for diagnosing pancreatic diseases compared to other diagnostic modalities, as highlighted in the literature [61]. However, the main duct of a healthy pancreas is not apparent on the CT images, as seen in Fig. 4.1 (a). On the other hand, if the main pancreatic duct is dilated, a dark line structure can be observed



(a) Normal pancreas



(b) Pancreas with dilated pancreatic duct

Figure 4.1: Visual examples of axial CT slices of (a) normal pancreas and (b) pancreas with dilated pancreatic duct. The orange contour surrounds the pancreas region, and the pancreatic duct is indicated by blue.

inside the pancreas region as shown in Fig. 4.1 (b). Considering these facts, the automated segmentation of dilated pancreatic ducts from volumetric CT images is expected to aid in the early detection of pancreatic cancer. A cautious follow-up may be needed if the pancreatic duct can be segmented from a patient's CT scan, as it could indicate potential progression into PDAC in the future. Furthermore, the segmentation results

can potentially be used as a clinical screening tool in Computer-Aided Diagnosis (CAD) systems for pancreatic cancer in the future.

4.2 Related works

In the past few years, some studies have been devoted to the PDAC segmentation [115–119] and its surrounding anatomy such as blood vessels potentially useful for evaluation of treatment response [120]. However the automated segmentation of the pancreatic duct itself was excluded, and only a few consider it as a discrete segmentation target [116, 117]. Among them, Zhou et al. [116] suggested an approach for segmenting the pancreas, pancreatic tumor, and pancreatic duct using multi-phase CT. An investigation on a dual-path network was conducted to efficiently integrate data from multi-phase volumetric CT images. To make the most of the information collected from both the arterial and venous phases of the volumetric CT images, Xia et al. [117] suggested a CT alignment strategy and achieved good performance using the multiple phases of volumetric CT images. Both studies investigated pancreatic duct segmentation methods on a large number of precisely annotated venous and arterial phase volumetric CT images, which are extremely difficult to obtain. All volumetric CT images were obtained from patients who have already been diagnosed with PDAC. To the best of my knowledge, this work is the first to address pancreatic duct segmentation in individuals without any pre-existing pancreatic cancer.

4.3 Contribution

This chapter proposes an anatomical attention-based framework for dilated pancreatic duct segmentation. The contributions are summarized as follows:

1. Investigate the anatomical attention-based strategy to guide the FCN to focus on the pancreas region during dilated pancreatic duct segmentation.

2. Employ the multi-scale aggregation strategy before the final prediction to fully utilize the information produced at different scales.

3. Introduce the feature obtained from the tubular structure enhancement filter as an additional input channel of FCN. This inclusion aids the FCN in capturing additional connection information that may be lost during the convolution process.

4. Validate the proposed approach on a real-world dilated pancreatic duct dataset from patients who have not been detected as suffering from PDAC yet and show acceptable performance on dilated pancreatic duct segmentation using single-phase volumetric CT images only.

4.4 Coarse-to-fine pancreatic duct segmentation

4.4.1 Overview

Since the pancreatic duct only takes up a small portion of the abdominal volumetric CT image, it is particularly challenging to segment the target directly. This study investigates a coarse-to-fine framework for dilated pancreatic duct segmentation. As illustrated in Fig. 4.2, the framework mainly consists of two stages. In the first stage, a coarse pancreas segmentation model is created using a publicly available pancreas dataset. The details of coarse pancreas segmentation will be described in 4.4.2. The dilated pancreatic duct dataset is applied to a pancreas segmentation model for prediction. This process generates the Region Of Interest (ROI) crops of the pancreas region, which are then utilized for fine pancreatic duct segmentation described in 4.4.3. In the

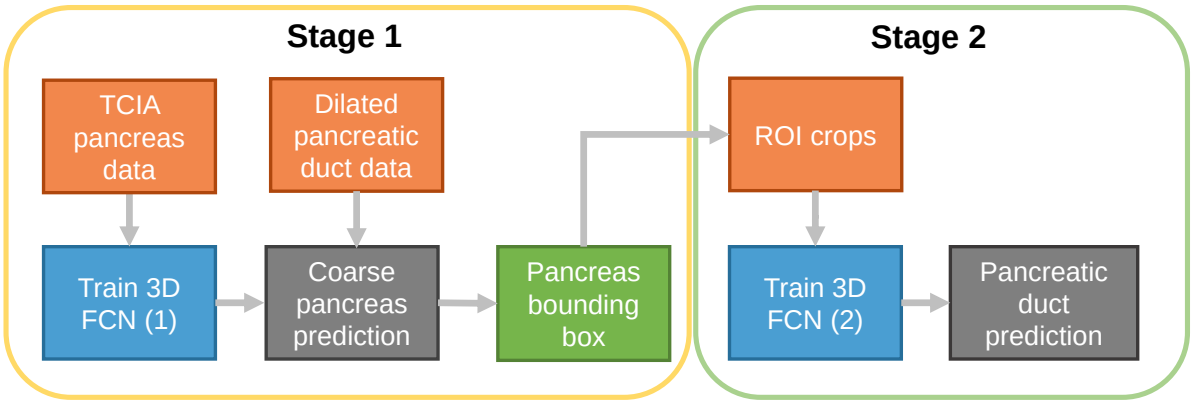


Figure 4.2: Overall workflow of coarse-to-fine pancreatic duct segmentation.

Table 4.1: GPU memory requirements for different input sizes when using the standard 3D U-Net and V-Net. The batch size is set to 2 in all the following settings.

Network	Input size	Parameters #	GPU memory usage
3D U-Net [43]	128^3	19.1M	~20.0 GB
V-Net [44]	128^3	41.2M	~5.8 GB
3D U-Net	160^3	19.1M	>24.0 GB
V-Net	160^3	41.2M	~9.7 GB

second stage, the dilated pancreatic duct segmentation model is trained exclusively using the ROI crops for fine prediction. The details of fine pancreatic duct segmentation are described in 4.4.4.

4.4.2 Coarse pancreas segmentation

Here, a straightforward yet effective pancreatic mask segmentation model is developed using a publicly accessible dataset [95] for coarse pancreas mask segmentation. U-Net [42] is a ready-to-use FCN that has proven to be useful in the field of medical image segmentation. 3D U-Net [43] and V-Net [44] are well-known 3D extensions for U-Net, which show considerably good ability in handling 3D volumetric images instead of 2D images. As shown in Table 4.1, the standard V-Net can hold larger input volumetric

image sizes with less memory usage than the standard 3D U-Net. Due to this fact, V-Net is used here as the baseline for the coarse pancreatic mask segmentation.

Scaling up the networks in depth, width, and resolution aspects has been shown to be beneficial in boosting segmentation performance [121]. When it comes to the V-Net, increasing the number of resolution levels of the network helps it to capture more specific information from the input volumetric images. This is especially important in medical image analysis, where the networks attempt to make the most of the limited training data. Extending the input size of FCNs can also benefit the network by allowing it to handle larger volumetric images and more detailed contexts. However, simply scaling up a neural network is not always the best approach because larger networks are more computationally expensive to train and may be more prone to overfitting. Model complexity must be carefully balanced with computational efficiency and generalization performance. Here, the standard V-Net is scaled up to discover the most efficient type for coarse pancreas segmentation and then is utilized as a baseline for pancreatic duct segmentation.

Different types of normalization techniques, including Batch Normalization (BN) and Instance Normalization (IN) are further introduced. Deep Neural Networks (DNNs) often use normalization to increase convergence stability, accelerate training speed, and mitigate issues like vanishing and exploding gradients. In 3D U-Nets, BN is a commonly used method [87]. It normalizes the activations of the previous layer by subtracting the batch mean and dividing by the batch standard deviation. When training with 3D FCNs, the number of batches is always constrained by the amount of available GPU memory. Smaller batch sizes, though, might affect how well batch normalization works [88]. On the other hand, IN is a technique that can be used to alleviate this issue. In contrast to BN, which normalizes the activations across the batch, it normalizes the activations for each unique sample in the batch. This lessens the reliance on batch size and boosts the model's robustness. In the context of 3D FCNs, IN can be particularly useful, as

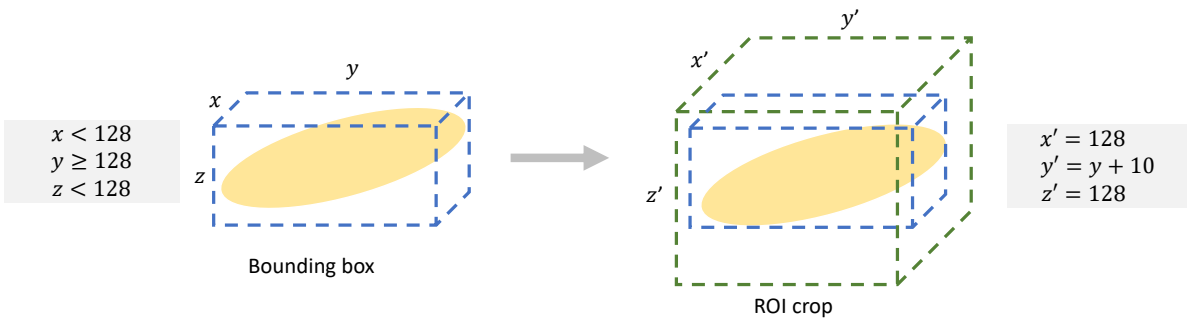


Figure 4.3: Example of ROI processing. Yellow ellipses represent the pancreas regions. Blue cubes are the bounding boxes of the pancreas, and the green cube is the ROI crop for the next step.

it allows for the use of larger batch sizes without requiring a significant increase in GPU memory. This can lead to faster training times and improved model performance. Furthermore, it has been shown to be effective in reducing the internal covariate shift, which is the change in the distribution of the network activations due to the changing parameters of the network. This can help to reduce the risk of overfitting and improve the generalization performance of the model. Overall, IN is a useful technique for 3D FCNs that can help to improve training performance and model accuracy, particularly when dealing with limited GPU memory and small batch sizes.

4.4.3 Pancreatic ROI generation

A pancreatic duct dataset is adapted to the model trained in Step 1 to perform coarse pancreas predictions. The predictions are interpolated back to the original size using the nearest-neighborhood interpolation. The pixel numbers of all connected components are counted in the pancreas prediction, and only the largest connected component is kept. A 3D bounding box of the pancreas region can be calculated through pancreas prediction. The ROI crops need to be larger than $128 \times 128 \times 128$ voxels to match the FCN input size. The details of generating ROI crops are shown in Fig. 4.3. The length,

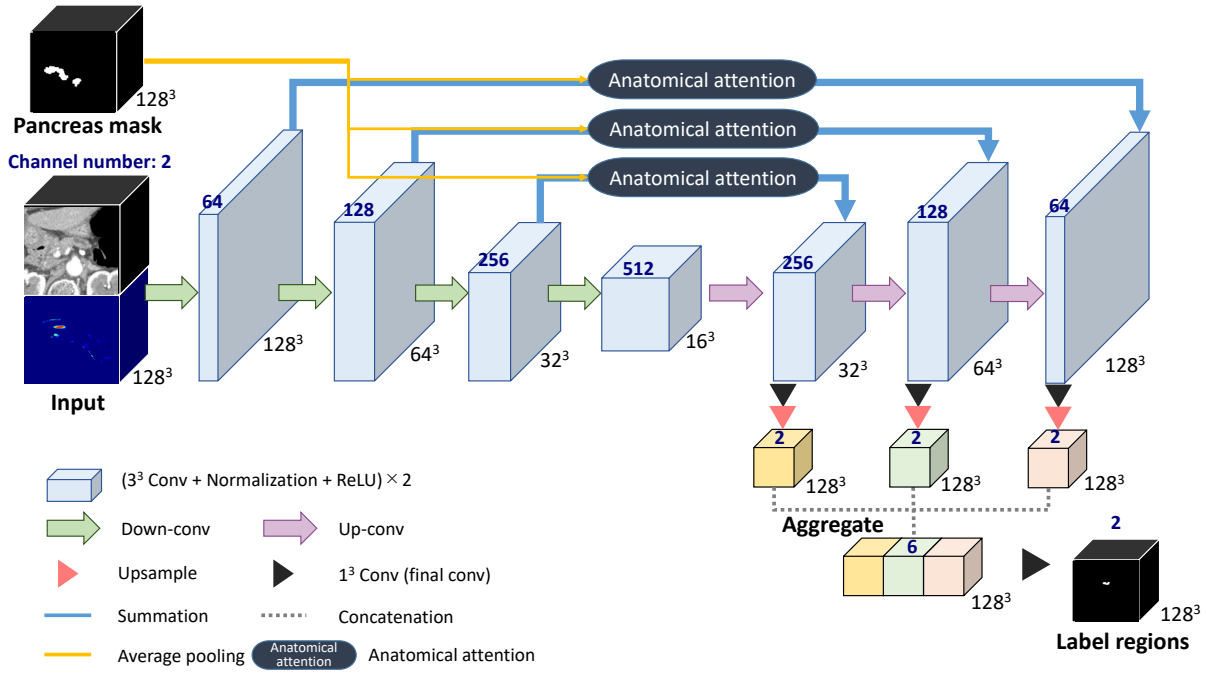


Figure 4.4: Proposed anatomical attention FCN architecture. Channel numbers are listed above the boxes in blue. Convolution blocks are shown by blue boxes, and the feature map sizes are indicated next to the boxes.

width, and height of the bounding box are defined as (x, y, z) assuming $x, z < 128$ and $y \geq 128$. The final length, width, and height of the ROI crop are (x', y', z'). Edges that are shorter than 128 pixels (like x and z) are extended to 128 pixels to match with the FCN input size. For edges longer than 128 pixels, 5 pixels are added on both sides to keep the boundaries of the pancreas inside the ROI crops.

4.4.4 Fine pancreatic duct segmentation

Network architecture

For dilated pancreatic duct segmentation, this work proposes an anatomical attention-based FCN as shown in Fig. 4.4. The FCN structure consists of an encoder part and a de-

coder part with resolution levels of four. A training set $\mathbf{S} = \{\mathbf{I}_n, \mathbf{L}_n, \mathbf{M}_n, \mathbf{P}_n; n = 1, ..., N\}$ is prepared, where $\mathbf{I}_n \in \mathcal{R}^{H \times W \times D}$ indicates the n-th volumetric CT image from the total N training sets. The volumetric image size in height, width, and depth of the n-th training set is $H \times W \times D$ pixels. $\mathbf{L}_n$ represents the corresponding ground-truth volumetric image of the pancreatic duct region, and $\mathbf{M}_n$ represents the mask of the pancreas region, which is obtained from the segmentation result in 4.4.2. $\mathbf{P}_n$ is the features generated from the tubular structure enhancement filter. The input of the FCN is a two-channel union of volumetric CT image $\mathbf{I}_n$ and the corresponding pancreatic duct enhancement $\mathbf{P}_n$. The coarse pancreas prediction mask $\mathbf{M}_n$ is employed to guide the anatomical attention on each level for the decoder. In multi-level FCNs, the high-resolution features are typically more focused on spatial information, whereas low-resolution features usually concentrate on the semantic information from the input. Combining features from multiple scales enables the learning of additional complementary information, which helps boost and refine the final prediction [122, 123]. Therefore, the proposed method aggregates the feature maps from each level to produce the final segmentation similar to the deep supervision [124]. The feature of each scale is upsampled to accommodate the size of the final prediction. To achieve this aggregation, a $1 \times 1 \times 1$ convolution is applied to reduce the channels of the feature map for each level to the class number. The outputs from each level are subsequently scaled using bilinear interpolation to match the resolution of the final prediction. These features are concatenated, and a $1 \times 1 \times 1$ convolution is applied to transform the number of channels to the number of output classes.

Anatomical attention

The attention mechanism is widely utilized in the medical image analysis field [97, 123, 125, 126], to capture useful information and ignore the useless context in FCNs.

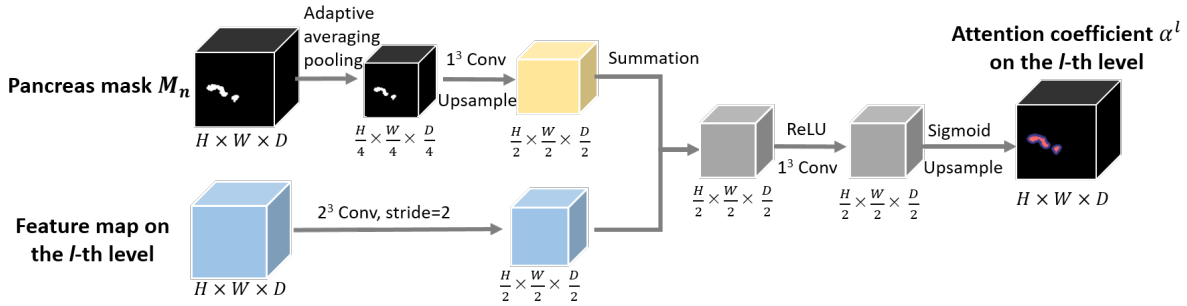


Figure 4.5: Process of computing the anatomical attention coefficient on the l-th level of FCN. Colored boxes represent feature maps with their sizes indicated below them.

To effectively focus on the organ that the target object belongs to, here, a pancreatic attention-guided method is proposed inspired by the grid-attention method [97]. Instead of using the image spatial information obtained in the bottleneck, the proposed method introduces pancreatic attention, which fully exploits the pancreas region's coarse prediction. Since the dilated pancreatic duct only makes up a small fraction of the entire pancreas, it is preferred to focus on the whole pancreas rather than just the target.

An attention coefficient $\mathcal{A}_j^l \in [0, 1]$ can be computed for each level l of the FCN based on grid-attention [97], where j is the j-th voxel of the input image. A coarse pancreas segmentation result is used as a mask for the attention gate to provide spatial information. The detailed procedure of calculating the attention coefficient vector $\boldsymbol{\mathcal{A}}^l$ on the l-th level is shown in Fig. 4.5. The pancreas masks $\mathbf{M}_n$ are downsampled using adaptive averaging pooling [127] to match the size of the bottleneck layer, and is followed by a $1 \times 1 \times 1$ convolution to learn the pixel-wise focus regions $\mathbf{g}_n$. For the input feature map, a $2 \times 2 \times 2$ convolution with stride 2 is applied. The output is summed up with the focus region $\mathbf{g}_n$ over the channel dimension as:

$$\vec{\mathbf{g}}_n = \phi(\vec{W}_g \mathcal{Z}(\vec{P}_n) + b_g). \tag{4.1}$$

An attention vector $\boldsymbol{\mathcal{A}}^l$ with values between 0 and 1 is produced by the sigmoid activation function. Furthermore, upsampling operations are necessary to make the attention vector's size suitable for each level's feature map. The output of anatomical attention is obtained by multiplying the input feature map $\mathbf{x}^l$ and attention vector elementally, as:

$$\hat{\mathbf{x}^l} = \boldsymbol{\mathcal{A}}^l \cdot \mathbf{x}^l. \tag{4.2}$$

Tubular structure enhancement channel

To fully capture the tubular structure of the pancreatic duct, an additional channel is introduced as input to the FCN. For vessel segmentation, vessel enhancement algorithms are often incorporated to increase the robustness of the segmentation performance [128]. Most of these algorithms try to represent the curvature of the vessel-like structure with the second derivatives of the volumetric image intensities. In the medical image analysis field, the Frangi filter is commonly used as a filter to enhance vessels to identify tubular structures and suppress other image features such as noise and non-vessel structures [129]. It is a Hessian-based method proposed to strengthen the differences in intensity in medical volumetric images with eigenvalues $|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$, where λ_1, λ_2, and λ_3 are derived from the Hessian matrix to indicate the principal curvature of the intensity profile at each voxel. An ideal tubular structure in 3D voxel has $|\lambda_1| \approx 0$, $|\lambda_1| \ll |\lambda_2|$, and $\lambda_2 \approx \lambda_3$. The Frangi filter is formulated as:

$$F = \begin{cases} \left(1 - \exp\left(-\frac{R_a^2}{2\alpha^2}\right)\right) \exp\left(-\frac{R_b^2}{2\beta^2}\right) \left(1 - \exp\left(-\frac{S^2}{2\gamma^2}\right)\right) & \text{if } \lambda_2 \leq 0 \text{ and } \lambda_3 \leq 0, \\ 0 & \text{otherwise}, \end{cases} \tag{4.3}$$

where $\alpha = 0.5$ and $\beta = 0.5$ are fixed by experience to control the sensitivity of the filter, and γ uses half of the maximum Hessian norm of the intensity range [130]. R_a is used to distinguish the tubular-like and the plate-like structure, R_b to measure the

blob-like structure, and S indicates the low contrast backgrounds. These patterns can be formulated as:

$$R_a = \frac{|\lambda_2|}{|\lambda_3|}, R_b = \frac{|\lambda_1|}{\sqrt{|\lambda_2 \lambda_3|}}, \; S = \sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}. \tag{4.4}$$

For post-processing, pancreatic mask segmentation is utilized to eliminate the values outside the pancreas areas. Additionally, to better adapt to the DNN, min-max normalization is applied to convert the remaining filter output into the range of 0 to 1.

Multi-scale aggregation

Features from different scales have different influences on the final prediction. Features from low resolution are expected to focus on the semantic information, while those from high resolution focus more on spatial information. Aggregation of features has proven to show remarkable performance in segmentation [122, 123, 131]. Thus, the attention vectors from each level are aggregated for final segmentation. The proposed method follows the attention aggregation introduced in [125]. Feature on each scale is indicated as $\vec{F}_j$, where $\{j = 1, 2, 3\}$ indicates the j-th level of the network, except the bottleneck layer. The final prediction of each scale j can be obtained by using an $1 \times 1 \times 1$ convolution to transform the number of channels to the number of output classes K. Due to the resolution difference among feature maps at each scale, bilinear interpolation is used to upsample the feature map to $\vec{\hat{F}}_j$. The output $\vec{\mathcal{F}}$ of the network is formulated as:

$$\vec{\mathcal{F}} = \vec{W}_f \left(\left[\vec{\hat{F}}_1, \vec{\hat{F}}_2, \vec{\hat{F}}_3 \right] \right), \tag{4.5}$$

where $\vec{W}_f$ is an $1 \times 1 \times 1$ final convolution to reduce the channels of concatenated feature map to the class number K.

4.5 Experiments

The standard Attention U-Net [97] is compared with the proposed FCNs, including the Pancreatic-anatomical Attention Network (PANet), PANet with Multi-scale aggregation (PAMNet), and PAMNet utilizing tubular structure enhancement as multi-channel input. The Multi-Channel input evaluation includes both non-normalized (MCPAMNet) and Normalized (NMCPAMNet) variations.

4.5.1 Datasets

Manual data annotation of dilated pancreatic duct

A pancreatic duct dilation dataset is generated from scratch using Pluto [132], a platform designed for CAD of medical images. Contrast-enhanced CT images were taken for patients with over 3 mm pancreatic duct dilation on an abdominal ultrasound scan. It is important to note that none of the patients had been diagnosed with PDAC. A median filter with a factor of 3 voxels was applied to smooth the volumetric CT images. Then, the display window level as well as the window width were manually adjusted, and the Hounsfield Unit (HU) intensity range of the pancreatic duct dilation region was observed. The region growing algorithm [133] was employed and then the starting voxel was chosen from the volumetric CT image. As the HU intensity of pancreatic duct dilation regions was similar to the surrounding tissues, falsely connected components were manually erased slice by slice. After that, connected-component labeling with 26-connected neighborhood connectivity was adopted to set labels for each component. Finally, label numbers for the pancreatic duct were selected and threshold processing was applied to remove unnecessary labels. After these semi-automatic operations, the annotations were manually checked slice by slice.

Specifications

Two datasets were used in this experiment. For coarse pancreas segmentation, a public pancreas annotation dataset [95] from The Cancer Imaging Archive (TCIA) was used which includes 82 contrast-enhanced portal venous phase abdomen 3D volumetric CT images. Each volumetric image contained 181–466 slices with a size of 512×512 pixels per slice. The slice thickness was between 0.5–1.0 mm. The whole set was split into training, validation, and test sets with 46, 16, and 18 volumetric images, respectively. For dilated pancreatic duct segmentation, a dataset was generated from scratch, that consisted of 30 abdominal contrast-enhanced portal venous phase 3D volumetric CT images containing 192–887 slices with a size of 512×512 pixels per slice. The voxel dimensions of the volumetric CT images were [0.59–0.75, 0.59–0.75, 0.30–1.0] mm. The dilated pancreatic duct data were split into five folds equally with each fold containing six volumetric CT images. Each fold was set as the test set and cross-validation was performed for the remaining four folds.

The intensity of all volumetric images was rescaled to $[-200, 200]$ HU and then normalized to $[0, 1]$. The resolutions of each volume CT image were resampled into 1 mm isotropically. Random cropping was applied alternately to generate sub-volumes with a mini-batch size of two. Both the input and output sizes were $128 \times 128 \times 128$ voxels. A combination of the Dice loss function and Focal loss function with $\lambda = 0.7$ were introduced and Adam optimization [81] was used with an initial learning rate of $1e{-4}$. All experiments were performed on an NVIDIA Quadro P6000 GPU with 24GB of memory. The implementation of FCNs was based on PyTorch [134].

4.5.2 Evaluation metrics

The segmentation performance of the dilated pancreatic duct was evaluated using four types of metrics, including Dice Similarity Coefficient (DSC), Sensitivity, Normalized

Surface Distance (NSD), and 95% percentile of the Hausdorff Distance (HD95).

DSC calculates the overlap percentage between the segmentation result and the ground truth which is defined as:

$$\mathrm{DSC} = \frac{2 \times \mathrm{TP}}{2 \times \mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \tag{4.6}$$

where TP, FP, and FN are the voxel numbers of true-positive, false-positive, and false-negative voxels. The value range of DSC is [0, 1], where DSC $= 1$ represents the optimum segmentation.

Sensitivity quantifies TP's contribution to the overall accurate segmentation as:

$$\mathrm{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{4.7}$$

NSD is another distance-based metric that is used to calculate the degree of overlap there is between two borders as:

$$\mathrm{NSD} = \frac{|B_g \cap R_s^\tau| + |B_s \cap R_g^\tau|}{|B_g| + |B_s|}, \tag{4.8}$$

where B_g and B_s indicate the boundary of the ground truth g and the predicted results p. R_g^τ is the border regions of the ground truth, and R_p^τ is the border regions of the prediction result under the maximum tolerated distance τ. The maximum tolerated distance represents the threshold of uncertainties and acceptable deviations between ground truth and prediction result boundaries, which is set to 1 here. The value range of NSD is $[0, 1]$, where NSD $= 1$ representes that the boundaries are fully overlapped.

HD is the most frequently used distance-based metric for segmentation tasks. It measures the distance between two point sets, that is the maximum value of the minimum

Table 4.2: Ablation study to show the effectiveness of scaling the V-Net in the network level, numbers of initial filter, and input size for pancreatic mask segmentation. The best DSC is shown in **bold**.

Level #	Initial filter #	Parameters #	Input size	Normalization	DSC [%]
4	16	9.9M	128^3	BN	53.2
			160^3	BN	59.4
			160^3	IN	74.8
5	16	41.2M	128^3	BN	50.9
			160^3	BN	67.8
			160^3	IN	75.7
6	16	166.2M	128^3	BN	39.9
			160^3	BN	69.3
			160^3	IN	78.3
6	32	664.7M	128^3	BN	44.1
			160^3	BN	63.3
			160^3	IN	**78.8**

distance from a point in the ground truth to the prediction result as:

$$\mathrm{HD} = \max\{\sup_{p_g \in S_g} \inf_{p_s \in S_s} \|p_g, p_s\|, \sup_{p_s \in S_s} \inf_{p_g \in S_g} \|p_g, p_s\|\}, \tag{4.9}$$

where the voxel on the surface of the ground truth S_g and the segmentation S_s are denoted by p_g and p_s, respectively. HD95 ignores the impact of small subsets of the outliers by only computing the 95% percentile values rather than the maximum distances of all points between ground-truth and segmentation results. Here, HD95 was used to evaluate the segmentation performance.

4.5.3 Experimental results on coarse pancreas segmentation

Table 4.2 shows the quantitative evaluation result of the coarse pancreatic mask segmentation. V-Net was used as the baseline and the network was scaled up in depth, width, and resolution, which correspond to the level number, filter number, and input size, respectively. Scaling up the V-Net positively influenced pancreas segmentation. In

this experiment, both BN and IN were used, with IN being more beneficial for pancreas segmentation.

4.5.4 Experimental results on dilated pancreatic duct segmentation

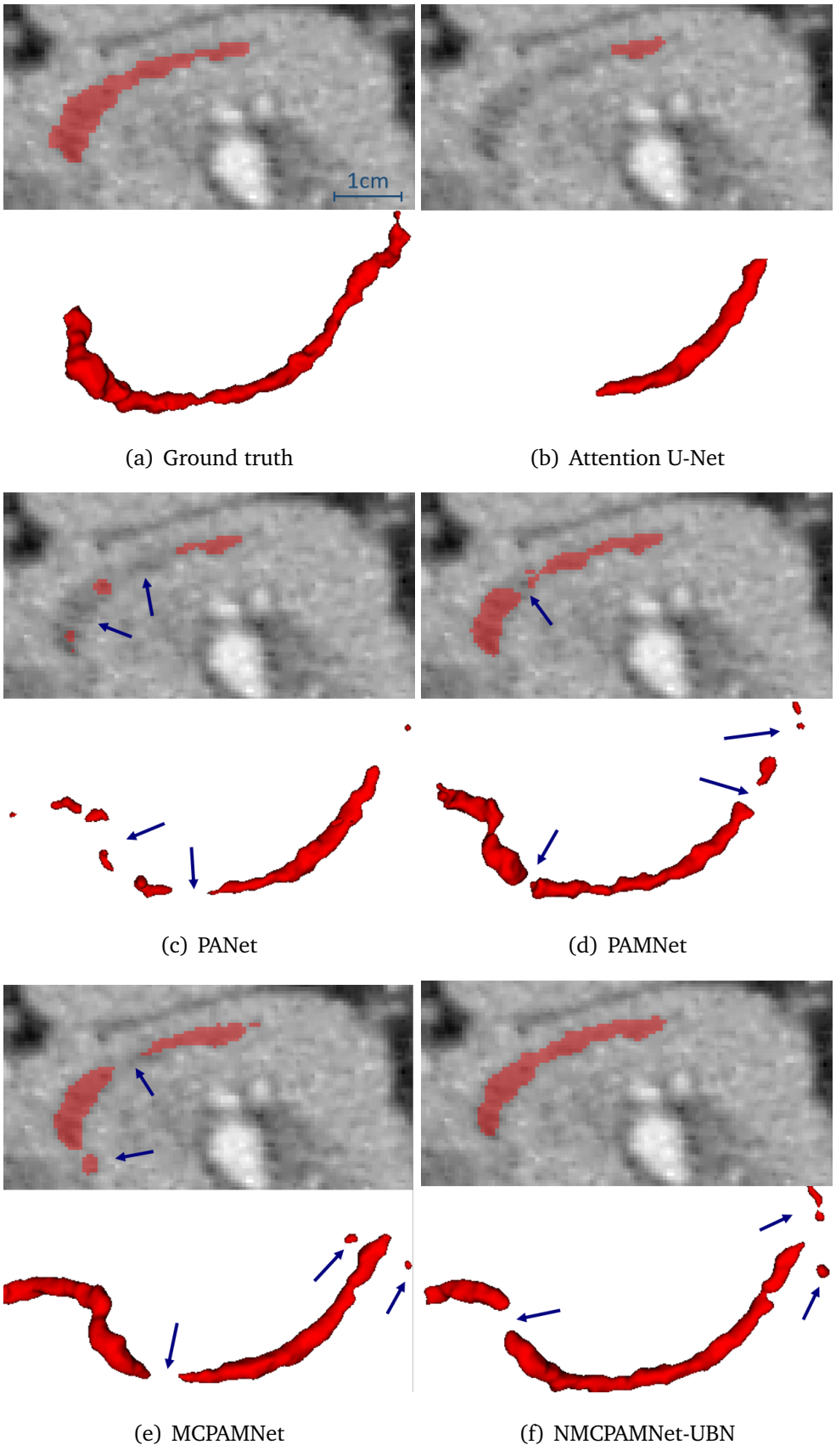
Table 4.3 shows an ablation study of the proposed FCNs on the dilated pancreatic duct segmentation. In the context of pancreatic duct segmentation, the baseline model was the Attention U-Net, with 3D U-Net serving as the backbone FCN. Recognizing V-Net's strong performance in pancreas segmentation, additional experiments were conducted using V-Net as the backbone. Furthermore, evaluations of different FCNs were conducted with both IN and BN to identify the optimal combination. The best results were obtained by NMCPAMNet using the V-Net as the FCN backbone and IN as the normalization operation. Fig. 4.6 shows segmentation examples of coronal slice and 3D rendering in each method with U-Net and BN as the backbone. Fig. 4.7 presents a segmentation comparison using NMCPAMNet with four different backbone combinations: U-Net+BN, U-Net+IN, V-Net+BN, and V-Net+IN. Fig. 4.8 shows a comparison of the heatmap depiction of attention coefficients using Attention U-Net and the proposed anatomical attention.

The proposed PANet outperformed the standard Attention U-Net [97]. The multiple levels of FCN were aggregated for the final output in MPANet, and the tubular structure enhancement as an additional input channel was incorporated in MCPAMNet. To further boost the efficiency of tubular structure enhancement, the min-max normalization was adopted on the features in NMCPAMNet.

Table 4.4 provides additional comparisons between the proposed method and other reported pancreatic duct segmentation strategies. Both studies [116] and [117] were carried out on PDAC patients, whose pancreatic ducts were substantially larger than normal cases. Furthermore, these studies used a dataset of 239 cases, which was much

Table 4.3: Ablation study of using the proposed FCNs on the dilated pancreatic duct segmentation. The best value for each metric is shown in **bold**.

Network	Backbone	DSC [%](↑)	Sensitivity [%](↑)	NSD [%](↑)	HD95(↓)
Attention U-Net	U-Net+BN	50.7 ± 17.0	56.8 ± 22.3	53.1 ± 19.4	89.5 ± 43.6
	U-Net+IN	52.5 ± 16.5	62.9 ±.21.8	59.7 ±.20.2	83.5 ± 45.5
PANet	U-Net+BN	52.9 ± 16.1	58.3 ± 21.8	59.9 ± 20.2	67.8 ± 41.1
	U-Net+IN	54.2 ± 13.2	66.4 ± 17.2	63.1 ± 15.9	72.1 ± 38.7
	V-Net+BN	52.6 ± 17.0	62.3 ± 22.2	60.8 ± 19.9	72.0 ± 42.6
	V-Net+IN	51.1 ± 14.5	64.9 ± 17.9	63.0 ± 14.5	86.6 ± 36.3
PAMNet	U-Net+BN	53.4 ± 11.4	66.1 ± 14.1	63.2 ± 14.5	75.5 ± 39.8
	U-Net+IN	52.1 ± 13.3	**67.5 ± 14.9**	**63.7 ± 15.4**	77.2 ± 46.9
	V-Net+BN	52.7 ± 14.4	62.3 ± 19.9	60.5 ± 17.0	64.1 ± 45.1
	V-Net+IN	53.8 ± 14.1	60.7 ± 17.4	60.8 ± 18.4	66.9 ± 44.9
MCPAMNet	U-Net+BN	53.3 ± 13.1	65.2 ± 17.5	59.9 ± 14.0	63.0 ± 42.4
	U-Net+IN	53.3 ± 10.7	60.3 ± 18.6	61.9 ± 12.1	63.6 ± 45.2
	V-Net+BN	53.5 ± 13.6	65.5 ± 17.7	60.8 ± 16.2	66.9 ± 43.0
	V-Net+IN	55.5 ± 11.3	62.5 ± 18.2	63.0 ± 14.2	60.4 ± 46.4
NMCPAMNet	U-Net+BN	55.2 ± 11.9	64.4 ± 17.2	63.2 ± 12.2	68.4 ± 44.1
	U-Net+IN	55.4 ± 12.0	64.2 ± 15.8	**63.7 ± 13.4**	62.5 ± 45.6
	V-Net+BN	55.4 ± 13.0	62.8 ± 19.4	63.1 ± 14.7	65.0 ± 44.7
	V-Net+IN	**55.7 ± 12.5**	64.3 ± 19.3	63.0 ± 15.7	**56.4 ± 40.3**

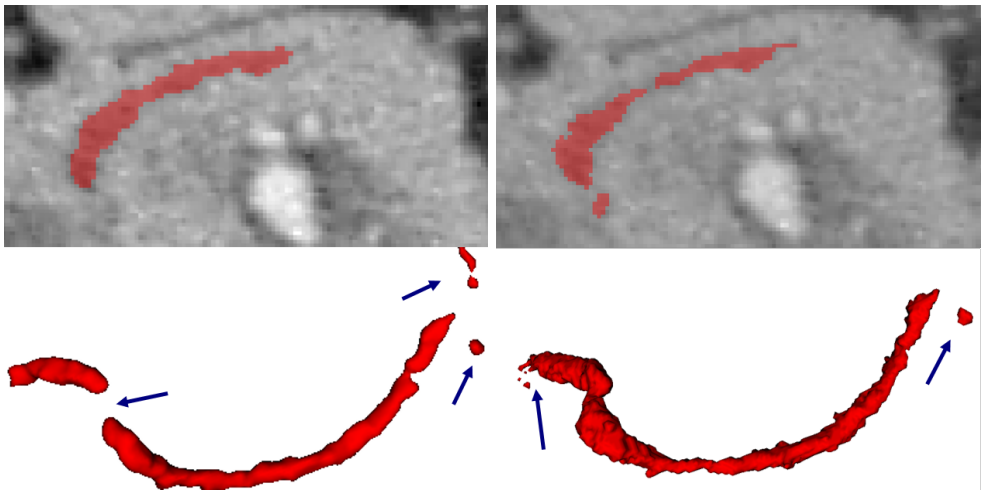


(a) Ground truth (b) Attention U-Net

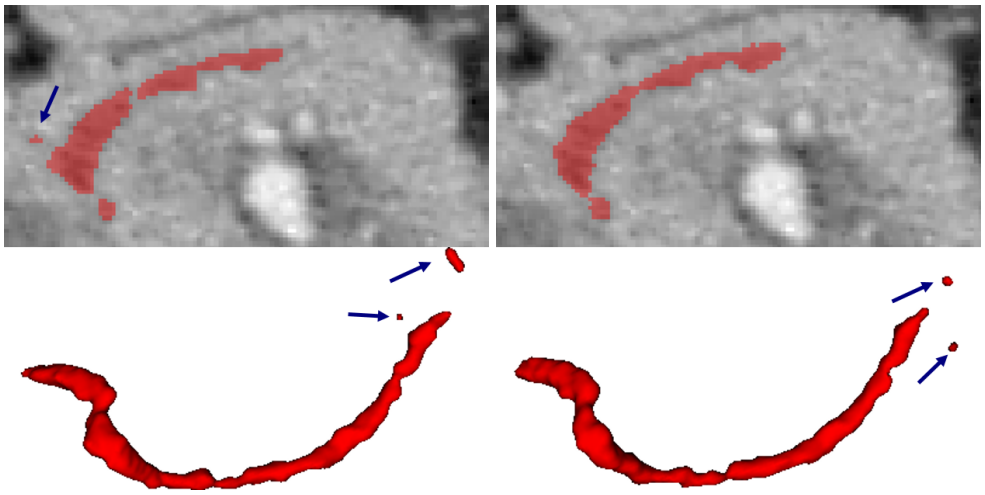
(c) PANet (d) PAMNet

(e) MCPAMNet (f) NMCPAMNet-UBN

Figure 4.6: Comparison of pancreatic duct segmentation results. Segmentation failure is indicated by a blue arrow. The upper image represents the axial slice, while the lower one depicts the 3D rendering.

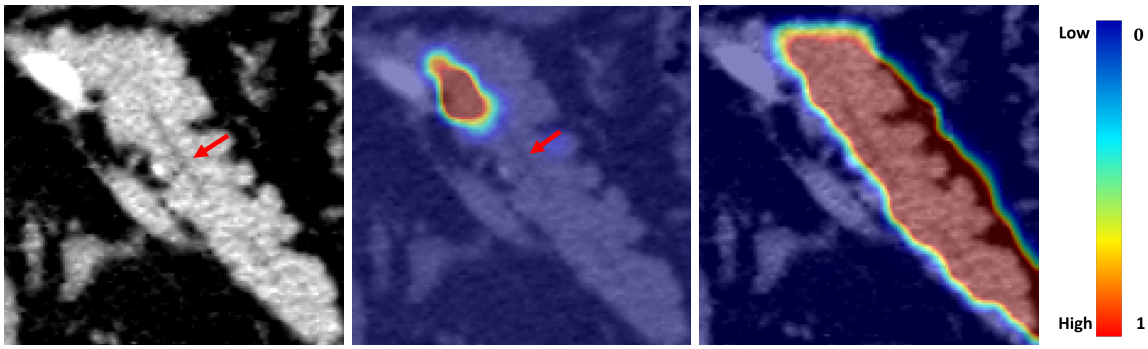


(a) U-Net+BN (b) U-Net+IN



(c) V-Net+BN (d) V-Net+IN

Figure 4.7: Comparison of pancreatic duct segmentation result using NMCPAMNet architecture with four different backbone settings. Segmentation failures are indicated by a blue arrow. The ground truth of this case is shown in Fig. 4.6 (a). The upper image represents the axial slice, while the lower one depicts the 3D rendering.



(a) CT (b) Attention U-Net [97] (c) Proposed anatomical attention

Figure 4.8: Heatmap visualization of attention coefficients. The pancreatic duct is indicated by the red arrow inside the pancreas.

Table 4.4: Comparison to previous pancreatic duct segmentation methods. Other pancreatic duct region segmentation results using single or multiple phase volumetric CT images are included. The proposed methods are highlighted in **bold**. The methods marked with an $\star$ were also proposed by the author.

Methods	Phase	# of data	DSC (%)
3D-UNet-single-phase [116]	Arterial	239 PDAC	38.35 $\pm$ 28.98
3D-UNet-single-phase [116]	Venous	239 PDAC	40.25 $\pm$ 27.89
3D-ResDSN-single-phase[116]	Arterial	239 PDAC	47.04 $\pm$ 26.42
3D-ResDSN-single-phase [116]	Venous	239 PDAC	49.81 $\pm$ 26.23
Cascade SE-Dense U-net [135]$\star$	Venous	30 normal	49.87 $\pm$ 22.54
MPA-Net[136]$\star$	Venous	30 normal	54.16 $\pm$ 12.60
Proposed method	**Venous**	**30 normal**	**55.70 $\pm$ 12.50**
3D-UNet-multi-phase (fusion) [116]		239 PDAC	39.06 $\pm$ 27.33
3D-UNet-multi-phase-HPN [116]		239 PDAC	44.93 $\pm$ 24.88
3D-ResDSN-multi-phase (fusion) [116]	Multiple	239 PDAC	48.49 $\pm$ 26.37
3D-ResDSN-multi-HPN [116]		239 PDAC	56.77 $\pm$ 23.33
Multi-phase Alignment Ensemble [117]		239 PDAC	64.38 $\pm$ 29.67

greater than the 30 examples used here. Although it is difficult to directly compare studies using different datasets, the proposed method yielded the highest DSC on pancreatic duct segmentation on a single-phase volumetric CT image.

4.6 Discussion and Conclusion

4.6.1 Discussion

This work proposed an anatomical attention-based strategy for dilated pancreatic duct segmentation from volumetric CT images. Because the pancreatic duct only takes up a minor percentage of the total volumetric CT image, this work started by localizing the pancreas area, as inspired by clinical experience [137]. A basic yet efficient model for coarse pancreas segmentation was employed using the publicly accessible TCIA pancreas dataset. Then, pancreatic segmentation was concentrated based on the pancreas ROI obtained in the previous step. Anatomical attention was introduced to force the FCN to focus on the pancreas anatomy area and disregard other portions in order to improve the pancreatic duct segmentation performance. Furthermore, tubular structure enhancement was incorporated as an extra FCN input channel to boost the tubular feature of the pancreatic duct.

Table 4.2 showed the ablation study of scaling the V-Net in depth, width and resolution aspects for pancreas segmentation. For the pancreatic mask segmentation, increasing the input size and number of levels of the V-Net were both beneficial and efficient. On the other hand, increasing the initial filter number helped slightly but significantly raised the parameter numbers of FCN. Thus, we can say that it is necessary to strike a balance between segmentation performance and model complexity. For normalization, IN was better suited for pancreatic segmentation using V-Net than BN. When applying IN on V-Net, the DSC for pancreatic segmentation increased considerably at each network scale. This improvement is significant because the performance of BN relies heavily on batch size, which is constrained by computer power. Due to these constraints, 3D segmentation tasks often use a relatively small batch size during training, which was specifically set to 2 in this chapter.

For pancreatic duct segmentation, the segmentation results were evaluated in four different aspects, including DSC, sensitivity, NSD, and HD95. These four metrics measured the overlap, sensitivity, border accuracy, and maximum deviation of the segmentation result. An ablation study using the proposed FCNs on the dilated pancreatic duct segmentation was shown in Table 4.3. For pancreatic duct segmentation, focus on the entire pancreas anatomy improved the segmentation compared to the original Attention U-Net [97]. In medical image analysis, it is not always optimal to focus on a particular target. Narrowing the FCN focus would result in a lower fault tolerance during training when the target region is quite small. This hypothesis was further affirmed by the visualization of attention coefficients in Fig. 4.8. Some pancreatic duct parts were outside of the focus by standard attention U-Net. The segmentation performance was also enhanced by multiscale aggregation of FCN, which made full use of the knowledge acquired at each level. The DSC on pancreatic duct segmentation was significantly enhanced by introducing the normalized tubular structure enhancement as a second input channel. The additional channel helped FCN understand the duct's tubular structure better. The proposed NMCPAMNet with V-Net baseline and IN demonstrated the most favorable performance across all four metrics. It achieved the highest scores in DSC and the lowest in HD95. While it may not have shown the highest accuracy in terms of sensitivity and NSD, its performance remained comparable to other methods. Segmentation examples of 3D rendering and coronal slice segmentation were shown in Fig. 4.6 and Fig. 4.7. It was shown that tubular structure enhancement could improve the connection of the pancreatic duct segmentation. The pancreatic duct segmented by NMCPAMNet using V-Net+IN as the backbone exhibited smoother duct segmentation with less exceeding segmentations.

The proposed method was also compared with existing pancreatic duct segmentation strategies in Table 4.4. When compared to other reported results of pancreatic duct segmentation using single-phase volumetric CT images only, the proposed method

outperformed all existing strategies, despite the fact that only 30 cases were employed instead of the larger dataset's 239 cases. Incorporating multiple-phase volumetric CT images could provide additional features and potentially improve the accuracy of pancreatic duct segmentation. However, obtaining paired annotated multi-phase pancreatic duct dataset is quite difficult.

4.6.2 Conclusion

In this chapter, an anatomical attention-based strategy was investigated for the segmentation of the dilated pancreatic duct from volumetric CT images. The proposed strategy was motivated by clinical experience; when radiologists look for the pancreatic duct from volumetric CT images, they first try to locate the pancreas area. Thus, an attention mechanism that enabled to focus on the entire pancreas anatomy rather than just the target was proposed. To fully capture the vessel-like structure of the pancreatic duct, a tubular structure enhancement was employed as an additional input channel for the FCN. The proposed FCNs were evaluated using four different assessment measures, which demonstrated their effectiveness on dilated pancreatic duct segmentation. Upon comparing the results with other strategies that relied on single-phase volumetric CT images, the proposed method exhibited significant superiority. The proposed technique might be applied to other tube-like structure segmentation tasks for other anatomies in the future. Nevertheless, the duct component still has some exceeding and improper segmentation with the current proposed method. For the use of PDAC diagnosis in real-world settings, the overall accuracy still needs to be improved to capture the full anatomy of the duct. This remains as future work.

Chapter 5

Data circulation in CAD for pancreatic cancer

5.1 Relation with real-world data circulation

This section explores how the research presented in this thesis achieves the realization of Real-World Data Circulation (RWDC). A diagram depicting the relationship between this research and RWDC is presented in Fig. 5.1. This thesis includes two types of data circulations. One follows the traditional RWDC structure, similar to the content proposed in [76]. The other introduces a new RWDC called Pancreatic RWDC developed within this thesis. This thesis addressed three segmentation topics focused on the pancreas and its associated structures using volumetric Computed Tomography (CT) images. The promising segmentation results pave the way for future applications in Computer-Aided Diagnosis (CAD) systems for pancreatic cancer. As discussed in Chapter 1, pancreatic cancer stands as one of the deadliest cancers, claiming numerous lives each year. The development of a CAD system capable of aiding in the early detection of pancreatic tumors is a crucial objective. Semantic segmentation of anatomical struc-

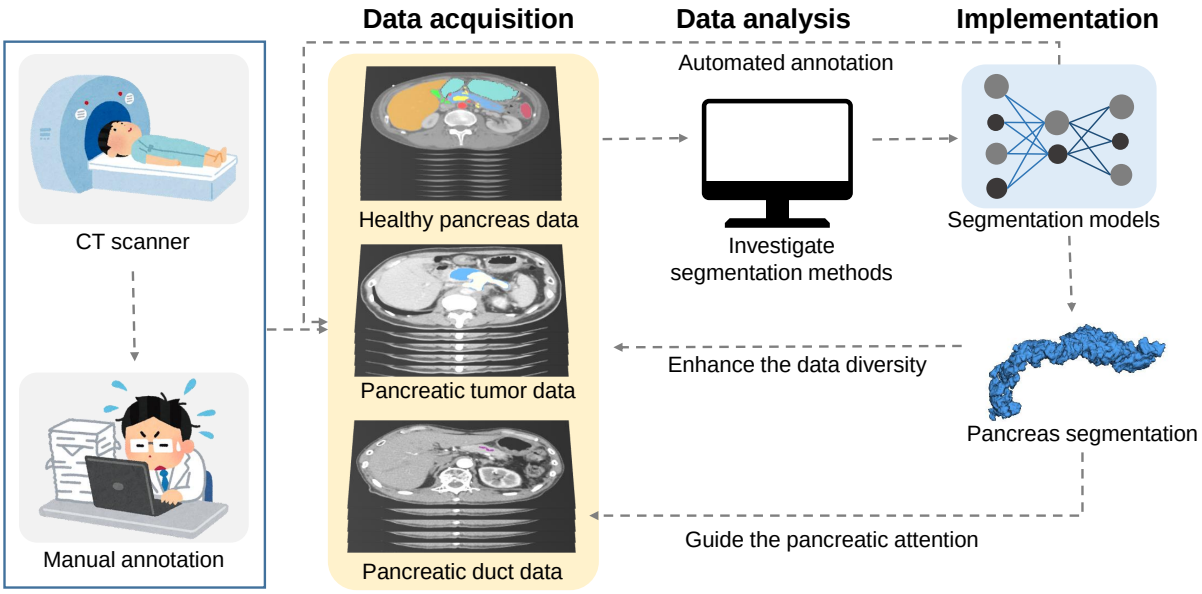


Figure 5.1: Relation of this research and RWDC in CAD for pancreatic cancer.

tures in medical images plays a pivotal role in the development of CAD systems. The pursuit of high-accuracy segmentation results is paramount in advancing the development of CAD systems.

5.1.1 Traditional RWDC

The data circulation process commences with data acquisition. This thesis primarily involves processing volumetric CT images from three patient types: healthy pancreas, pancreas with pancreatic tumor, and pancreas with dilated pancreatic duct. Volumetric CT images of healthy pancreas are obtained from individuals undergoing gastric surgery, where the pancreatic regions exhibit normal conditions. Volumetric CT images of the pancreas with pancreatic tumors are sourced from patients who have already been diagnosed with pancreatic cancer. Additionally, volumetric CT images of the pancreas with pancreatic duct dilation are obtained from patients who exhibit symptoms of pancreatic duct dilation but have not been diagnosed with pancreatic cancer. For segmentation

tasks, the annotation of the segmentation target is indispensable. In the medical image analysis field, the acquisition of data is notably expensive compared to other fields. This is primarily because only well-educated experts can generate annotations accurately. Furthermore, data diversity can introduce an imbalance in the implemented model. This research places a priority on obtaining datasets that are not only large and diverse but also accurate. It is crucial to emphasize that all datasets used in this study have undergone ethical approval, and comprehensive informed consent was obtained from every individual participant involved in the research.

For the following analysis phase, three distinct segmentation topics were introduced, each focusing on different types of volumetric CT images. The primary goal was to explore accurate and efficient segmentation approaches for each specific task. The first task was to perform segmentation of the pancreas and its multiple associated organs on healthy pancreas volumetric CT images. This topic yielded two segmentation methods to improve the accuracy of pancreas segmentation. The outcomes of this task could serve as a fundamental groundwork for the subsequent two tasks. The second task was centered on the segmentation of the pancreas and pancreatic tumors from volumetric CT images obtained from patients diagnosed with pancreatic cancer. Due to the varied appearances of pancreatic tumors in volumetric CT images, a considerable amount of training data is essential to obtain a segmentation model that possesses high generalizability. Addressing this challenge, the Federated Learning (FL) technique was employed for this task. Furthermore, a novel model aggregation method was introduced to enhance the accuracy of pancreatic tumor segmentation. The third task centered on the segmentation of dilated pancreatic ducts from volumetric CT images obtained from patients without pancreatic cancer. Due to the small size of the pancreatic duct in abdominal volumetric CT images, special attention was needed to guide the training of Fully Convolutional Networks (FCNs). The region of interest, crucial for this task, was generated through the implementation of the first task. Furthermore, vesselness

enhancement was incorporated within the FCNs to improve the connectivity conditions of the pancreatic duct. The results of these analyses could be utilized to implement CAD systems, holding the potential for application in real clinical scenarios. The segmentation implementations have the capacity to enhance data acquisition processes. As previously mentioned, annotation is a costly process in the medical image analysis field. A highly accurate implementation model can play a crucial role in automatically generating annotations, offering a more efficient and cost-effective solution. This process constitutes a part of the data circulation.

5.1.2 Pancreatic RWDC

In addition to the traditional RWDC introduced in 5.1.1, this section introduces an additional aspect of data circulation which is related to the pancreas. This thesis analyzes three different types of volumetric CT images, including the healthy pancreas data and pancreatic pathology data. The segmentation of a healthy pancreas and its multiple associated organs could serve as a foundation for analyzing pancreatic pathology data. This proves especially crucial in working on pancreatic pathology, where the initial step is to identify the pancreas region from the abdominal volumetric CT images. In the segmentation of pancreatic tumors, the pancreas itself is also a segmentation target. Utilizing cases with a healthy pancreas can also contribute to refining the segmentation of the pancreas and pancreatic tumors. In the segmentation of dilated pancreatic ducts, targeting the pancreatic duct in abdominal volumetric CT images proves challenging due to its small voxel size. The segmentation of the pancreas serves as a guide for directing attention to the pancreas in this task. This distinction makes this data circulation system different from traditional RWDC approaches [76].

5.2 Societal value

In this thesis, segmentation methods for three distinct topics were introduced, with a focus on the pancreas as the target for CAD system development. The reason behind concentrating on the pancreas lied in the fact that pancreatic cancer has the highest mortality rate among cancers. The low survival rate of pancreatic cancer is attributed to limited symptoms, leading to late-stage detection. Early detection of pancreatic cancer is critical, and CT is widely used for its diagnosis. The accurate segmentation of the pancreas and pancreatic tumor can be leveraged in a CAD system to aid in the detection of pancreatic cancer. However, even with detection, the challenge lies in the difficulty of resection. How can we identify the signs of pancreatic cancer earlier? This thesis focused on the dilated pancreatic duct as an entry point. Clinical studies indicate that dilation of the main pancreatic duct may signal a high risk for pancreatic cancer as demonstrated in Chapter 4. If the dilated regions of the pancreatic duct are obtained easily from volumetric CT images after a comprehensive medical examination, careful follow-up could be taken. This has the potential to lead to the early detection of pancreatic cancer. The segmentation results can also be utilized in virtual endoscopy of the pancreas, which is expected to be a new method for intraductal papillary mucinous neoplasm diagnosis [138]. To the best of my knowledge, the method presented in this thesis is the first one on the segmentation of the dilated pancreatic duct from patients who have not been diagnosed with pancreatic cancer. Additionally, the Pancreatic RWDC could contribute to future data acquisition processes, facilitating the acquisition of more annotated data. This, in turn, has the potential to significantly reduce the cost of pancreas-related segmentation studies.

Chapter 6

Summary and future direction

6.1 Summary

This thesis presented research on pancreas and pancreatic-related segmentation methods from volumetric Computed Tomography (CT) images. Accurate and efficient automated segmentation is essential for the development of Computer-Aided Diagnosis (CAD) systems.

In Chapter 1, the aim of this study was introduced; propose segmentation methods for the pancreas and its related structures from volumetric CT images. CT stands out as the predominant diagnostic modality in clinical practice for non-invasive abdominal diagnosis, and its evolution has led to the widespread adoption of 3D CT, offering more valuable information compared to the previous reliance on 2D imaging. However, the significant reading time required is a notable challenge for radiologists. Thus, the development of CAD systems is desired to assist radiologists and enhance interpretation accuracy simultaneously. Segmentation is a fundamental task in CAD system development, and there is a high demand for accurate segmentation results. Therefore, the main focus of this thesis was on the methodology of semantic segmentation from vol-

umetric CT images. In the abdominal region, the pancreas stands out as one of the most challenging organs to segment, mainly because of its low contrast and significant individual variation. Additionally, considering that pancreatic cancer has the lowest five-year survival rate and is acknowledged as the most lethal cancer compared to others, the pancreas was the primary segmentation target of this work. Thus, the goal of this thesis was to introduce segmentation approaches for the pancreas and its related regions, aiming to aid in pancreatic cancer diagnosis in clinical settings.

The work can be primarily categorized into three main targets: (1) Segmentation of the pancreas within the context of multiple abdominal organs, (2) Segmentation of the pancreas and pancreatic tumors, and (3) Segmentation of the pancreatic duct. While segmenting large objects like abdominal organs is relatively straightforward, dealing with smaller tissues, such as the pancreatic duct, presents a significant challenge. The presented research progressed from addressing simpler tasks to tackling more intricate aspects of segmentation.

In Chapter 2, two distinct methods for segmenting a healthy pancreas and its multiple associated abdominal organs and tissues were proposed. The segmentation of the pancreas presents unique challenges due to its relatively small size and indistinct boundary, making it more difficult than segmenting larger organs. In real clinical practice, identifying and segmenting the pancreas is facilitated by considering the relative positions and anatomical connections between organs. Therefore, this work included its multiple associated abdominal organs as the segmentation targets. The first method was based on improving the traditional Fully Convolutional Network (FCN). A spatial-embedded FCN was introduced to enhance the segmentation accuracy, especially for organs with small voxel sizes. Additionally, improved data augmentation was implemented, including original non-rigid data augmentation and gamma correction, to mitigate overfitting during training. While data augmentation may offer limited benefits for datasets from the same hospital and scanner across most organs, notable improvement

was observed in pancreas segmentation. In the second method, an efficient application of Automated Machine Learning (AutoML) was introduced, simplifying the process of developing multi-organ segmentation from volumetric CT images. To enhance the AutoML process, proxy data were leveraged that effectively represented the entire training dataset for network structure and hyperparameter searching instead of randomly selecting data. Experiments revealed that the use of proxy data had a positive impact on boosting the final AutoML model's performance.

In Chapter 3, a framework for pancreas and pancreatic tumor segmentation utilizing Federated Learning (FL) was introduced. In clinical situations, the development of highly accurate and generalizable pancreas and corresponding tumor segmentation models is strongly encouraged as a prerequisite for CAD systems. While achieving high-performance models in the clinical domain demands extensive, diverse, and high-quality datasets, collecting datasets from multiple resources encounters various restrictions in the medical image analysis field. FL emerged as an innovative solution to address this challenge. When working with training data from various resources, the challenge of heterogeneity can significantly affect the model's performance upon aggregation at the server. Additionally, the imbalance in the number of samples per client results in clients with smaller datasets contributing less to the global model. This chapter investigated two multi-task optimization methods for FL in medical image analysis with heterogeneous datasets: Dynamic Task Prioritization (DTP) and Dynamic Weight Averaging (DWA). In DTP, a higher prioritization was assigned to the client with superior performance. Meanwhile, DTW dynamically adjusted the weight of each client for aggregation. Each method was evaluated within an FL framework and the overall model performance was compared with traditional FL aggregation methods, such as FedAvg and FedProx. DWA showed significant improvement, especially on the Synapse client whose training data was relatively smaller than the other two clients.

In Chapter 4, a framework for the segmentation of the pancreatic duct was intro-

duced, aiming to assist the early detection of pancreatic cancer. The importance of early detection is emphasized by the fact that pancreatic duct dilation often serves as an early symptom of pancreatic cancer. Therefore, the segmentation of dilated pancreatic ducts may offer a crucial starting point for early pancreatic cancer detection, but segmenting the pancreatic duct from volumetric CT images poses a challenge due to its minute structure. This chapter investigated a coarse-to-fine strategy aimed at creating Regions Of Interest (ROI) within the pancreas, facilitating a more refined segmentation of the pancreatic duct. Additionally, an attention mechanism was introduced that focused on the entire pancreatic anatomy rather than just the target area. To ensure a comprehensive representation of the vessel-like structure of the pancreatic duct, tubular structure enhancement was integrated as an extra input channel for the FCN. When comparing the results with other reported outcomes for pancreatic duct segmentation, the proposed method demonstrated significant superiority over strategies that rely on single-phase volumetric CT images. This work also has the potential to be applied to other segmentation tasks involving tubular structures in different anatomies in the future.

In Chapter 5, the connection between the research presented in this thesis and Real-World Data Circulation (RWDC) was introduced. An RWDC process consists of three steps: data acquisition, data analysis, and implementation. The circulation commenced with data acquisition, with all volumetric CT images used in this thesis obtained from real clinical settings. Three distinct segmentation topics were introduced for analysis, each focusing on different types of volumetric CT images. Applying these segmentation methods in the development of a CAD system has the potential for use in real clinical scenarios. Furthermore, the implementation has the potential to contribute to future data acquisition processes, enabling the acquisition of more annotated data. The society value of this RWDC could contribute to the CAD system, assisting in the detection of pancreatic cancer in clinical settings.

6.2 Future research

The future perspective of the research field encompasses two aspects: methodology and the clinical aspects. In terms of methodology, the primary goal is to boost the accuracy of segmentation for individual organs and tissues by employing more efficient techniques. The pursuit of heightened accuracy remains a perpetual need in the medical field. Additionally, exploring data privacy and collaborative learning represents a promising and essential direction for future research in medical image analysis. On the clinical side, the development of CAD systems is vital and holds promise for upcoming clinical scenarios. Integrating computer technology in the medical field has the potential to greatly assist radiologists and doctors in their daily routines within hospital settings. Moreover, CAD systems are anticipated to play a pivotal role in the early detection of various abnormalities and pathologies.

6.2.1 Methodology aspect

Organs and tissues segmentation

This thesis proposed three different types of segmentation tasks related to the pancreas and its surrounding tissues. The segmentation results for larger targets, such as organs, were close to satisfactory. However, for pancreatic pathology components like pancreatic tumors and dilated pancreatic ducts, the accuracy still requires significant improvement, when considering the application to CAD systems. Improving the segmentation accuracy for each target is consistently the primary demand in this field.

Recently, deep learning-based methods have been developing rapidly. Fully automated segmentation of organs and tissues from volumetric CT images is essential for medical image analysis. The aforementioned studies primarily focused on research using FCNs. The effectiveness of FCNs in capturing spatial dependencies has significantly

contributed to the State-Of-The-Art (SOTA) performance in this domain. Recently, a new method called Transformer [139], originally designed for natural language processing, has demonstrated notable success in segmentation tasks. This development opens up exciting opportunities for further exploration in this field. Its extensions have garnered significant attention and demonstrated considerable success in various medical image segmentation tasks [140, 141]. Using Transformer models could be a future direction for improving the segmentation accuracy of organs and tissues.

Federated Learning (FL)

In a real-world clinical scenario, a generalized segmentation model is desired for CAD systems. The primary challenge in achieving such generalized models is the scarcity of substantial datasets for training. Collecting extensive and diverse training data in the medical image analysis field is always challenging due to the special characteristics of medical images, which may pose security and privacy issues. Additionally, annotation is expensive because only experts can perform the annotation. FL is naturally suited for application in the field because it enables distributed learning across several institutions. However, the current FL setting among different institutions is not very effective, as it requires a server to manage the overall process. Recently, a new concept called Swarm Learning has been proposed [142]. It facilitates collaborative learning without the need for a central coordinator. This could be a potential research direction for distributed learning in the future.

6.2.2 Clinical aspect

CAD system for pancreatic cancer

For pancreatic cancer treatment, the initial step is the detection of pancreatic cancer. Subsequent tasks, such as tumor resection planning, require highly detailed segmenta-

tions of the tumor and surrounding anatomical structures [143]. The accurate segmentation of the pancreas and pancreatic tumor can be utilized in a CAD system to assist in the detection of pancreatic cancer. However, in real clinical practice, factors such as reliability, uncertainty, and robustness against variation become increasingly important when considering segmentation methods [143]. Additionally, the trade-off between accuracy and computational cost is a key consideration for CAD development. Therefore, in the future, there is a need for further improvement in developing high-generalized models with high efficiency.

Early detection of pancreatic cancer

This research was pioneering in its focus on the segmentation of the dilated pancreatic duct, representing a novel contribution to the field. The investigation was based on a dataset comprising 30 cases, acknowledging the limitation of its relatively small size. In subsequent research endeavors, expanding the dataset will be imperative to ensure the robustness and generalizability of the findings.

The segmentation of the dilated pancreatic duct holds significant promise as an early detection marker for pancreatic cancer. Building on this potential, the future work will delve into the critical aspect of seamlessly integrating the segmentation results into real-world applications. Addressing the translational gap between research findings and practical implementations poses a crucial challenge that necessitates careful consideration in the ongoing investigations.

Acknowledgement

This thesis compiled the results of my research during the Doctoral Course in Intelligent Systems at the Graduate School of Informatics, Nagoya University. I am deeply grateful to the many individuals who have made significant contributions to the completion of this thesis.

First and foremost, I am deeply grateful to my advisor, Professor Kensaku Mori. He not only supported us with the necessary research equipment but also offered valuable advice and expertise regarding my research. His insights in our research area have been instrumental in solving numerous challenging problems. Furthermore, I extend my heartfelt appreciation for the opportunities he afforded me to participate in various conferences and overseas studies, allowing me to broaden my insights and stay abreast of the current trends in medical image processing.

I express my sincere thanks to Associate Professor Masahiro Oda. His continuous provision of constructive comments and strong support in manuscript writing and presentation. Through his guidance, I have gained valuable insights into showcasing my research more accurately and efficiently.

I extend my sincere thanks to Dr. Yuichiro Hayashi. He consistently provides me with insightful comments and suggestions on my work. Additionally, his encouragement during my Doctoral Course has been invaluable.

I owe a significant debt of gratitude to Dr. Holger R. Roth from NVIDIA Corpora-

tion. He played a pivotal role in guiding me through the intricacies of deep learning from my Master's course, providing me with a wealth of valuable techniques and warm encouragement throughout my research. I am truly indebted to his substantial support, without which I would not have been able to complete my work.

I express my gratitude to Professor Ichiro Ide and Professor Yoshito Mekada, who diligently reviewed my doctoral thesis. Their valuable advice and insightful comments significantly contributed to enhancing the quality of my work.

I would like to thank to Dr. Tadaaki Miyamoto and medical technologist Gen Sato from Chiba Kensei Hospital. Their wealth of knowledge on the pancreas and dilated pancreatic duct, along with their support in collecting the dataset, has been invaluable to my research. Their contributions have served as a significant inspiration for my doctoral research.

I would like to thank Pochuan Wang from National Taiwan University with whom I collaborated on Federated Learning research. It was a rewarding and pleasant collaboration.

I would like to thank Professor Daniel Ruckert from Imperial College London. Thanks to his acceptance of me as a visiting student at ICL, I gained valuable insights and knowledge during my stay there.

I would like to thank to Dr. Kazunari Misawa from Aichi Cancer Center for providing support with the dataset of abdominal CT volumes for this research.

I would like to thank the Graduate Program for Real-World Data Circulation Leaders. This program not only provided financial support but also offered a myriad of activities throughout my study at Nagoya University. These opportunities have significantly contributed to my learning experience, not only in the realm of research but also in gaining insights from an industry perspective.

I would like to thank Japan Society for the Promotion of Science. The financial support for both research and daily life has been crucial during my doctoral course,

allowing me to concentrate fully on my research endeavors.

Last but not least, I would like to extend my gratitude to the members of Mori Laboratory. Thanks to everyone's support and assistance over the past seven years, I have experienced significant personal and professional growth. I eagerly anticipate the future development of Mori Laboratory within the Graduate School of Informatics at Nagoya University.

Finally, I express my sincere gratitude to my parents and my fiancé for their unwavering love and understanding during the challenges of pursuing a Ph.D. Their support has been invaluable throughout this journey.

Publication

Journal paper

- Chen Shen, Holger R Roth, Yuichiro Hayashi, Masahiro Oda, Gen Sato, Tadaaki Miyamoto, Daniel Ruckert, and Kensaku Mori. Anatomical attention can help to segment the dilated pancreatic duct in abdominal CT. *International Journal of Computer Assisted Radiology and Surgery*, 2023. (accepted)
- Chen Shen, Holger R Roth, Yuichiro Hayashi, Masahiro Oda, Tadaaki Miyamoto, Gen Sato, and Kensaku Mori. A cascaded fully convolutional network framework for dilated pancreatic duct segmentation. *International Journal of Computer Assisted Radiology and Surgery*, 17(2):343–354, 2021.

Conference paper

International Conference

- Chen Shen, Pochuan Wang, Dong Yang, Daguang Xu, Masahiro Oda, Po-Ting Chen, Kao-Lang Liu, Wei-Chin Liao, Chiou-Shann Fuh, Kensaku Mori, Weichung Wang, and Holger R Roth. Joint multi organ and tumor segmentation from partial labels using federated learning. In *Proc. Distributed, Collaborative, and Federated Learning, and Affordable AI and Healthcare for Resource Diverse Global Health, 3rd*

Workshop DeCaF 2022 in Conjunction with MICCAI 2022, *Lecture Notes in Computer Science*, volume 13573, pages 58–67, 2022.

- Chen Shen, Holger R Roth, Vishwesh Nath, Yuichiro Hayashi, Masahiro Oda, Kazunari Misawa, and Kensaku Mori. Effective hyperparameter optimization with proxy data for multi-organ segmentation. In *Proc. SPIE Medical Imaging*, volume 12032, no. 120320T, pages 1–7, 2022.

- Chen Shen, Pochuan Wang, Holger R Roth, Dong Yang, Daguang Xu, Masahiro Oda, Weichung Wang, Chiou-Shann Fuh, Po-Ting Chen, Kao-Lang Liu, Wei-Chih Liao, and Kensaku Mori. Multi-task federated learning for heterogeneous pancreas segmentation. In *Proc. Clinical Image-Based Procedures, Distributed and Collaborative Learning, Artificial Intelligence for Combating COVID-19 and Secure and Privacy-Preserving Machine Learning, 2nd Workshop DCL 2021 Held in Conjunction with MICCAI 2021*, *Lecture Notes in Computer Science*, volume 12969, pages 101–110, 2021.

- Chen Shen, Holger R Roth, Yuichiro Hayashi, Masahiro Oda, Takaaki Miyamoto, Gen Sato, and Kensaku Mori. Attention-guided pancreatic duct segmentation from abdominal CT volumes. In *Proc. Clinical Image-Based Procedures, Distributed and Collaborative Learning, Artificial Intelligence for Combating COVID-19 and Secure and Privacy-Preserving Machine Learning, 10th Workshop CLIP 2021 Held in Conjunction with MICCAI 2021*, *Lecture Notes in Computer Science*, volume 12969, pages 46–55, 2021.

- Chen Shen, Yuichiro Hayashi, Masahiro Oda, Kazunari Misawa, and Kensaku Mori. Unpaired medical image translation between portal-venous phase and non-contrast CT volumes for multi-organ segmentation. In *International Forum on Medical Imaging in Asia 2021 Proc. SPIE*, volume 11792, no. 117920Q, pages 1–6, 2021.

- Chen Shen*, Pochuan Wang*, Holger R Roth, Dong Yang, Daguang Xu, Masahiro Oda, Kazunari Misawa, Po-Ting Chen, Wei-Chih Liao, Kao-Lang Liu, Weichung Wang, and Kensaku Mori. Automated Pancreas Segmentation Using Multi-institutional Collaborative Deep Learning. In *Proc. Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning, 1st Workshop DCL 2020 Held in Conjunction with MICCAI 2020*, *Lecture Notes in Computer Science*, volume 12444, pages 192–200, 2020. (*Equal contribution)

- Chen Shen, Chenglong Wang, Holger R Roth, Masahiro Oda, Yuichiro Hayashi, Kazunari Misawa, and Kensaku Mori. Spatial information-embedded fully convolutional networks for multi-organ segmentation with improved data augmentation and instance normalization. *In Proc. SPIE Medical Imaging*, volume 11313, no. 1131316, pages 1–7, 2020.

- Chen Shen, Fausto Milletari, Holger R. Roth, Hirohisa Oda, Masahiro Oda, Yuichiro Hayashi, Kazunari Misawa, and Kensaku Mori. Multi-class abdominal organs segmentation with improved V-Nets. *In Proc. SPIE Medical Imaging*, volume 10949, no. 109490B, pages 1–7, 2019.

- Chen Shen, Holger R Roth, Hirohisa Oda, Masahiro Oda, Yuichiro Hayashi, Kazunari Misawa, Kensaku Mori. Evaluation of 3D fully convolutional networks for multi-class organ segmentation in contrast-enhanced CT. In *Proc. 32nd International Congress and Exhibition of Computer Assisted Radiology and Surgery*, Volume 13, Issue 1 supplement, pages s21–s22, 2018.

Domenstic Conference

- 申 忱, Fausto Milletari, Holger R Roth, 小田 紘久, 小田 昌宏, 林 雄一郎, 三澤 一成, and 森 健策. CT からの腹部多臓器抽出における group normalization の影響に関す

る考察. *IEICE Technical Report,* MI2018-94, 2019.

- 申 忱, Holger R Roth, 林 雄一郎, 小田 紘久, 小田 昌宏, 三澤 一成, and 森 健策. 転移学習を用いた腹部 thick-slice CT 像における多臓器領域の自動抽出の初期検討. *In Proc. 38th the Japanese Society of Medical Imaging Technology*, OP4-15, 2019.

- Chen Shen, Holger R. Roth, Hirohisa Oda, Masahiro Oda, Yuichiro Hayashi, Kazunari Misawa, and Kensaku Mori. On the influence of Dice loss function in multi-class organ segmentation of abdominal CT using 3D fully convolutional networks. *IEICE Technical Report*, MI2017-51, 2017.

Bibliography

[1] Wikipedia contributors. CT scan. https://en.wikipedia.org/wiki/CT_scan/, 2023. (Accessed on 2024/1/10).

[2] Frank H Netter. *Atlas of Human Anatomy, Second Edition*. Novartis, Basel, Switzerland, 1997.

[3] Angela Cantatore and Pavel Müller. *Introduction to computed tomography*. Denmark Technical University, Department of Mechanical Engineering, Copenhagen, Denmark, 2011.

[4] Daniele Panetta. Advances in X-ray detectors for clinical and preclinical computed tomography. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 809:2–12, 2016.

[5] The Japanese Society of Medical Imaging Technology. *Medical Imaging Technology Handbook*. International Academic Publishing Co., Ltd., Tokyo, Japan, 2004.

[6] Hiroshi Fujita, Takayuki Ishida, and Shigehiko Katsuragawa. *Handbook of Practical Image Analysis in Medicine*. Ohmsha, Tokyo, Japan, 2012.

[7] Kunio Doi. Computer-aided diagnosis in medical imaging: Historical review,

current status and future potential. *Computerized Medical Imaging and Graphics*, 31(4–5):198–211, 2007.

[8] Heang-Ping Chan, Kunio Doi, Simranjit Galhotra, Carl J Vyborny, Heber MacMahon, and Peter M Jokich. Image feature analysis and computer-aided diagnosis in digital radiography. I. Automated detection of microcalcifications in mammography. *Medical Physics*, 14(4):538–548, 1987.

[9] Maryellen L Giger, Kunio Doi, and Heber MacMahon. Computerized detection of lung nodules in digital chest radiographs. In *Proc. SPIE Medical Imaging*, volume 767, pages 384–387, 1987.

[10] Maryellen Lissak Giger, Kunio Doi, and Heber MacMahon. Image feature analysis and computer-aided diagnosis in digital radiography. 3. Automated detection of nodules in peripheral lung fields. *Medical Physics*, 15(2):158–166, 1988.

[11] Shigehiko Katsuragawa, Kunio Doi, and Heber MacMahon. Image feature analysis and computer-aided diagnosis in digital radiography: Detection and characterization of interstitial lung disease in digital chest radiographs. *Medical Physics*, 15(3):311–319, 1988.

[12] Phillip H Meyers, Charles M Nice Jr, Hal C Becker, Wilson J Nettleton Jr, James W Sweeney, and George R Meckstroth. Automated computer analysis of radiographic images. *Radiology*, 83(6):1029–1034, 1964.

[13] Richard P Kruger, James R Townes, David Lee Hall, Samuel J Dwyer, and Gwilym S Lodwick. Automated radiographic diagnosis via feature extraction and classification of cardiac size and shape descriptors. *IEEE Transactions on Biomedical Engineering*, BME-19(3):174–186, 1972.

[14] D James Croft. Is computerized diagnosis possible? *Computers and Biomedical Research*, 5(4):351–367, 1972.

[15] Junichiro Toriwaki. Forty years progress in computer aided diagnosis of X-ray images. *Journal of Computer Aided Diagnosis of Medical Images*, 5(6):28–39, 2001.

[16] Jane P Ko and David P Naidich. Computer-aided diagnosis and the evaluation of lung disease. *Journal of Thoracic Imaging*, 19(3):136–155, 2004.

[17] Howard Lee and Yi-Ping Phoebe Chen. Image based computer aided diagnosis system for cancer detection. *Expert Systems with Applications*, 42(12):5356–5365, 2015.

[18] Marco Spadaccini, Andrea Iannone, Roberta Maselli, Matteo Badalamenti, Madhav Desai, Viveksandeep Thoguluva Chandrasekar, Harsh K Patel, Alessandro Fugazza, Gaia Pellegatta, Piera Alessia Galtieri, Gianluca Lollo, Silvia Carrara, Andrea Anderloni, Douglas K Rex, Victor Savevski MEng, Michael B Wallace, Pradeep Bhandari, Thomas Roesch, Ian M Gralnek, Prateek Sharma, Cesare Hassan, and Alessandro Repici. Computer-aided detection versus advanced imaging for detection of colorectal neoplasia: A systematic review and network meta-analysis. *The Lancet Gastroenterology & Hepatology*, 6(10):793–802, 2021.

[19] You Jin Ku, Hak Hee Kim, Joo Hee Cha, Hee Jung Shin, Eun Young Chae, Woo Jung Choi, Hee Jin Lee, and Gyungyub Gong. Predicting the level of tumor-infiltrating lymphocytes in patients with triple-negative breast cancer: Usefulness of breast MRI computer-aided detection and diagnosis. *Journal of Magnetic Resonance Imaging*, 47(3):760–766, 2018.

[20] Seyedehnafiseh Mirniaharikandehei, Alan B Hollingsworth, Bhavika Patel,

Morteza Heidari, Hong Liu, and Bin Zheng. Applying a new computer-aided detection scheme generated imaging marker to predict short-term breast cancer risk. *Physics in Medicine & Biology*, 63(10):105005_1–9, 2018.

[21] Taehee Lee, Eui Jin Hwang, Chang Min Park, and Jin Mo Goo. Deep learning-based computer-aided detection system for preoperative chest radiographs to predict postoperative pneumonia. *Academic Radiology*, 30:2844–2855, 2023.

[22] Kunio Doi. Current status and future potential of computer-aided diagnosis in medical imaging. *The British Journal of Radiology*, 78(suppl_1):s3–s19, 2005.

[23] Mohamed Esmail Karar, Ezz El-Din Hemdan, and Marwa A Shouman. Cascaded deep learning classifiers for computer-aided diagnosis of COVID-19 and pneumonia diseases in X-ray scans. *Complex & Intelligent Systems*, 7:235–247, 2021.

[24] J L Gayathri, Bejoy Abraham, M S Sujarani, and Madhu S Nair. A computer-aided diagnosis system for the classification of COVID-19 and non-COVID-19 pneumonia on chest X-ray images by integrating CNN with sparse autoencoder and feed forward neural network. *Computers in Biology and Medicine*, 141:1–10, 2022.

[25] Paul Suetens, Erwin Bellon, Dirk Vandermeulen, Maria-Helena Smet, Guy Marchal, Johan Nuyts, and Luc Mortelmans. Image segmentation: Methods and applications in diagnostic radiology and nuclear medicine. *European Journal of Radiology*, 17(1):14–21, 1993.

[26] Akinobu Shimizu, Rena Ohno, Takaya Ikegami, Hidefumi Kobatake, Shigeru Nawano, and Daniel Smutek. Segmentation of multiple organs in non-contrast 3D abdominal CT images. *International Journal of Computer Assisted Radiology and Surgery*, 2:135–142, 2007.

[27] Hyunjin Park, Alfred Hero, Peyton Bland, Marc Kessler, Jongbum Seo, and Charles Meyer. Construction of abdominal probabilistic atlases and their value in segmentation of normal organs in abdominal CT scans. *IEICE Transactions on Information and Systems*, E93-D(8):2291–2301, 2010.

[28] Faliu Yi and Inkyu Moon. Image segmentation: A survey of graph-cut methods. In *Proc. 2012 International Conference on Systems and Informatics*, pages 1936–1941, 2012.

[29] Yuri Boykov and Marie-Pierre Jolly. Interactive organ segmentation using graph cuts. In *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention 2000, Part 3, Lecture Notes in Computer Science*, volume 1935, pages 276–286, 2000.

[30] Xinjian Chen, Jayaram K Udupa, Ulas Bagci, Ying Zhuge, and Jianhua Yao. Medical image segmentation by combining graph cuts and oriented active appearance models. *IEEE Transactions on Image Processing*, 21(4):2035–2046, 2012.

[31] Marius George Linguraru, John A Pura, Vivek Pamulapati, and Ronald M Summers. Statistical 4D graphs for multi-organ abdominal segmentation from multiphase CT. *Medical Image Analysis*, 16(4):904–914, 2012.

[32] Chetan Bhole, Nicholas Morsillo, and Christopher Pal. 3D segmentation in CT imagery with conditional random fields and histograms of oriented gradients. In *Proc. Machine Learning in Medical Imaging, 2nd International Workshop MLMI 2011 Held in Conjunction with MICCAI 2011, Part 2, Lecture Notes in Computer Science*, volume 7009, pages 326–334, 2011.

[33] Yukitaka Nimura, Yuichiro Hayashi, Takayuki Kitasaka, Kazunari Misawa, and Kensaku Mori. Automated torso organ segmentation from 3D CT images using

conditional random field. In *Proc. SPIE Medical Imaging*, volume 9785, pages 931–936, 2016.

[34] Ying Yang, Masahiro Oda, Kazunari Misawa, Daniel Rueckert, and Kensaku Mori. Influence of voxel-connection structure in organ segmentation based on conditional random field. *IEICE Technical Report*, MI2014-112, 2017.

[35] José Ignacio Orlando, Elena Prokofyeva, and Matthew B Blaschko. A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. *IEEE Transactions on Biomedical Engineering*, 64(1):16–27, 2016.

[36] Chi-Hoon Lee, Shaojun Wang, Albert Murtha, Matthew RG Brown, and Russell Greiner. Segmenting brain tumors using pseudo–conditional random fields. In *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention 2008, Part I, Lecture Notes in Computer Science*, volume 5241, pages 359–366, 2008.

[37] Yabo Fu, Yang Lei, Tonghe Wang, Walter J Curran, Tian Liu, and Xiaofeng Yang. A review of deep learning based methods for medical image multi-organ segmentation. *Physica Medica*, 85:107–122, 2021.

[38] Yang Lei, Yabo Fu, Tonghe Wang, Richard LJ Qiu, Walter J Curran, Tian Liu, and Xiaofeng Yang. Deep learning in multi-organ segmentation. *Computing Research Repository arXiv Preprints*, arXiv:2001.10619, 2020.

[39] Hina Ajmal, Saad Rehman, Umar Farooq, Qurrat U Ain, Farhan Riaz, and Ali Hassan. Convolutional neural network based image segmentation: A review. In *Proc. SPIE Pattern Recognition and Tracking XXIX*, volume 10649, pages 191–203, 2018.

[40] Farhana Sultana, Abu Sufian, and Paramartha Dutta. Evolution of image segmentation using deep convolutional neural network: A survey. *Knowledge-Based Systems*, 201:1–25, 2020.

[41] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proc. 2015 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

[42] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention 2015, Part III, Lecture Notes in Computer Science*, volume 9351, pages 234–241, 2015.

[43] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention 2016, Part II, Lecture Notes in Computer Science*, volume 9901, pages 424–432, 2016.

[44] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *Proc. 4th International Conference on 3D Vision*, pages 565–571, 2016.

[45] Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, and Klaus H Maier-Hein. nnU-uet: Self-adapting framework for U-Net-based medical image segmentation. *Computing Research Repository arXiv Preprints*, arXiv: 1809.10486, 2018.

[46] Christian I Wade and Matthew J Streitz. *Anatomy, Abdomen and Pelvis: Abdomen*. StatPearls Publishing, St. Petersburg, FL, USA, 2020.

[47] Aaisha Shah and Yusuf S Khan. *Anatomy, Abdomen and Pelvis, Arteries and Veins*. StatPearls Publishing, St. Petersburg, FL, USA, 2021.

[48] Vishy Mahadevan. Anatomy of the liver. *Surgery (Oxford)*, 38(8):427–431, 2020.

[49] Vishy Mahadevan. Anatomy of the stomach. *Surgery (Oxford)*, 32(11):571–574, 2014.

[50] Rickesha L Wilson and Christina E Stevenson. Anatomy and physiology of the stomach. In *Shackelford's Surgery of the Alimentary Tract*, volume 1, pages 634–646. Elsevier, Amsterdam, The Netherlands, 2019.

[51] Reina E Mebius and Georg Kraal. Structure and function of the spleen. *Nature Reviews Immunology*, 5(8):606–616, 2005.

[52] Chantal Housset, Yves Chrétien, Dominique Debray, and Nicolas Chignard. Functions of the gallbladder. *Comprehensive Physiology*, 6(3):1549–1577, 2016.

[53] Daniel S Longnecker, Fred Gorelick, and Elizabeth D Thompson. Anatomy, histology, and fine structure of the pancreas. In *The Pancreas: An Integrated Textbook of Basic Science, Medicine, and Surgery*, chapter 2, page 10. Hoboken, NJ, USA, 2018.

[54] Po Sing Leung. Overview of the pancreas. *The Renin-Angiotensin System: Current Research Progress in The Pancreas, Advances in Experimental Medicine and Biology*, 690:3–12, 2010.

[55] Mark A Atkinson, Martha Campbell-Thompson, Irina Kusmartseva, and Klaus H

Kaestner. Organisation of the human pancreas in health and in diabetes. *Diabetologia*, 63:1966–1973, 2020.

[56] Jeremy L Humphris, Amber L Johns, Skye H Simpson, Mark J Cowley, Marina Pajic, David K Chang, Adnan M Nagrial, Venessa T Chin, Lorraine A Chantrill, Mark Pinese, R Scott Mead, Anthony J Gill, Jaswinder S Samra, James G Kench, Elizabeth A Musgrove, Katherine M Tucker, Allan D Spigelman, Nic Waddell, Sean M Grimmond, and Andrew V Biankin. Clinical and pathologic features of familial pancreatic cancer. *Cancer*, 120(23):3669–3675, 2014.

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

[58] Jonathan D Mizrahi, Rishi Surana, Juan W Valle, and Rachna T Shroff. Pancreatic cancer. *The Lancet*, 395(10242):2008–2020, 2020.

[59] Michael Orth, Philipp Metzger, Sabine Gerum, Julia Mayerle, Günter Schneider, Claus Belka, Maximilian Schnurr, and Kirsten Lauber. Pancreatic ductal adenocarcinoma: Biological hallmarks, current status, and future perspectives of combined modality treatment approaches. *Radiation Oncology*, 14(1):1–20, 2019.

[60] Sachiko Tanaka, Akihiko Nakaizumi, Tatsuya Ioka, Osamu Oshikawa, Hiroyuki Uehara, Miho Nakao, Kiyomi Yamamoto, Osamu Ishikawa, Hiroaki Ohigashi, and Tsugio Kitamra. Main pancreatic duct dilatation: A sign of high risk for pancreatic cancer. *Japanese Journal of Clinical Oncology*, 32(10):407–411, 2002.

[61] Mark D Edge, Maarouf Hoteit, Amil P Patel, Xiaoping Wang, Deborah A Baumgarten, and Qiang Cai. Clinical significance of main pancreatic duct dilation on computed tomography: Single and double duct dilation. *World Journal of Gastroenterology*, 13(11):1701–1705, 2007.

[62] Sachiko Tanaka, Miho Nakao, Tatsuya Ioka, Rena Takakura, Yasuna Takano, Hideaki Tsukuma, Hiroyuki Uehara, Reiko Suzuki, and Junko Fukuda. Slight dilatation of the main pancreatic duct and presence of pancreatic cysts as predictive signs of pancreatic cancer: A prospective study. *Radiology*, 254(3):965–972, 2010.

[63] Tara C Noone, Richard C Semelka, Deneise M Chaney, and Caroline Reinhold. Abdominal imaging studies: Comparison of diagnostic accuracies resulting from ultrasound, computed tomography, and magnetic resonance imaging in the same individual. *Magnetic Resonance Imaging*, 22(1):19–24, 2004.

[64] Carlos Pineda, Rolando Espinosa, and Angelica Pena. Radiographic imaging in osteomyelitis: The role of plain radiography, computed tomography, ultrasonography, magnetic resonance imaging, and scintigraphy. *Seminars in Plastic Surgery*, 23(02):80–89, 2009.

[65] Shandra Bipat, Saffire S K S Phoa, Otto M van Delden, Patrick M M Bossuyt, Dirk J Gouma, Johan S Laméris, and Jaap Stoker. Ultrasonography, computed tomography and magnetic resonance imaging for diagnosis and determining resectability of pancreatic adenocarcinoma: A meta-analysis. *Journal of Computer Assisted Tomography*, 29(4):438–445, 2005.

[66] Julian Panes, Rosa Bouzas, María Chaparro, Valle García-Sánchez, Javier-Pérez Gisbert, Blanca Martínez de Guereñu, Juan Luis Mendoza, José María Paredes, Sergi Quiroga, Tomás Ripollés, and Jordi Rimola. Systematic review: The use of ultrasonography, computed tomography and magnetic resonance imaging for the diagnosis, assessment of activity and abdominal complications of Crohn's disease. *Alimentary Pharmacology & Therapeutics*, 34(2):125–145, 2011.

[67] Rathan M Subramaniam, David A Kurth, Christine A Waldrip, and Frank J

Rybicki. American college of radiology appropriateness criteria: Advancing evidence-based imaging practice. *Seminars in Nuclear Medicine*, 49(2):161–165, 2019.

[68] Akshat Gotra, Lojan Sivakumaran, Gabriel Chartrand, Kim-Nhien Vu, Franck Vandenbroucke-Menu, Claude Kauffmann, Samuel Kadoury, Benoît Gallix, Jacques A de Guise, and An Tang. Liver segmentation: Indications, techniques and future directions. *Insights into Imaging*, 8(4):377–392, 2017.

[69] Sidra Gul, Muhammad Salman Khan, Asima Bibi, Amith Khandakar, Mohamed Arselene Ayari, and Muhammad E H Chowdhury. Deep learning techniques for liver and liver tumor segmentation: A review. *Computers in Biology and Medicine*, 147:1–16, 2022.

[70] Nicola Altini, Berardino Prencipe, Giacomo Donato Cascarano, Antonio Brunetti, Gioacchino Brunetti, Vito Triggiani, Leonarda Carnimeo, Francescomaria Marino, Andrea Guerriero, Laura Villani, Arnaldo Scardapane, and Vitoantonio Bevilacqua. Liver, kidney and spleen segmentation from CT scans and MRI with deep learning: A survey. *Neurocomputing*, 490:30–53, 2022.

[71] S M Reza Soroushmehr, Pavani Davuluri, Somayeh Molaei, Rosalyn Hobson Hargraves, Yang Tang, Charles H Cockrell, Kevin Ward, and Kayvan Najarian. Spleen segmentation and assessment in CT images for traumatic abdominal injuries. *Journal of Medical Systems*, 39:1–11, 2015.

[72] Lei Xu, Junjie Hu, Ying Song, Sen Bai, and Zhang Yi. Clinical target volume segmentation for stomach cancer by stochastic width deep neural network. *Medical Physics*, 48(4):1720–1730, 2021.

[73] Javeria Naz, Muhammad Attique Khan, Majed Alhaisoni, Oh-Young Song, Usman

Tariq, and Seifedine Kadry. Segmentation and classification of stomach abnormalities using deep learning. *Computers, Materials & Continua*, 69(1):607–625, 2021.

[74] Eli Gibson, Francesco Giganti, Yipeng Hu, Ester Bonmati, Steve Bandula, Kurinchi Gurusamy, Brian Davidson, Stephen P Pereira, Matthew J Clarkson, and Dean C Barratt. Automatic multi-organ segmentation on abdominal CT with dense V-networks. *IEEE Transactions on Medical Imaging*, 37(8):1822–1834, 2018.

[75] Jun Ma, Yao Zhang, Song Gu, Cheng Zhu, Cheng Ge, Yichi Zhang, Xingle An, Congcong Wang, Qiyuan Wang, Xin Liu, Shucheng Cao, Qi Zhang, Shangqing Liu, Yunpeng Wang, Yuhui Li, Jian He, and Xiaoping Yang. AbdomenCT-1K: Is abdominal organ segmentation a solved problem? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6695–6714, 2021.

[76] Kazuya Takeda. Introduction to the real-world data circulation paradigm. In *Frontiers of Digital Transformation: Applications of the Real-World Data Circulation Paradigm*, pages 3–10. Springer, Singapore, 2021.

[77] Eli Gibson, Francesco Giganti, Yipeng Hu, Ester Bonmati, Steve Bandula, Kurinchi Gurusamy, Brian R Davidson, Stephen P Pereira, Matthew J Clarkson, and Dean C Barratt. Towards image-guided pancreas and biliary endoscopy: Automatic multi-organ segmentation on abdominal CT with dense dilated networks. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention 2017, Part I, Lecture Notes in Computer Science*, volume 10433, pages 728–736, 2017.

[78] Holger R Roth, Chen Shen, Hirohisa Oda, Masahiro Oda, Yuichiro Hayashi, Kazu-

nari Misawa, and Kensaku Mori. Deep learning and its application to medical image segmentation. *Medical Imaging Technology*, 36(2):63–71, 2018.

[79] Li Chen, Yanjun Xie, Jie Sun, Niranjan Balu, Mahmud Mossa-Basha, Kristi Pimentel, Thomas S Hatsukami, Jenq-Neng Hwang, and Chun Yuan. 3D intracranial artery segmentation using a convolutional autoencoder. In *Proc. 2017 IEEE International Conference on Bioinformatics and Biomedicine*, pages 714–717, 2017.

[80] Chenglong Wang, Holger R Roth, Takayuki Kitasaka, Masahiro Oda, Yuichiro Hayashi, Yasushi Yoshino, Tokunori Yamamoto, Naoto Sassa, Momokazu Goto, and Kensaku Mori. Precise estimation of renal vascular dominant regions using spatially aware fully convolutional networks, tensor-cut and Voronoi diagrams. *Computerized Medical Imaging and Graphics*, 77:1–13, 2019.

[81] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proc. 3rd International Conference for Learning Representations*, 15 pages, 2015.

[82] Jonathan Waring, Charlotta Lindvall, and Renato Umeton. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. *Artificial Intelligence in Medicine*, 104:1–12, 2020.

[83] Dong Yang, Holger R Roth, Ziyue Xu, Fausto Milletari, Ling Zhang, and Daguang Xu. Searching learning strategy with reinforcement learning for 3D medical image segmentation. In *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention 2019, Part II, Lecture Notes in Computer Science*, volume 11765, pages 3–11, 2019.

[84] Vishwesh Nath, Dong Yang, Ali Hatamizadeh, Anas A Abidin, Andriy Myronenko, Holger Roth, and Daguang Xu. The power of proxy data and proxy networks for

hyper-parameter optimization in medical image segmentation. In *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention 2021, Part III, Lecture Notes in Computer Science*, volume 12903, pages 456–465, 2021.

[85] Sam Shleifer and Eric Prokop. Using small proxy datasets to accelerate hyperparameter search. *Computing Research Repository arXiv Preprints*, arXiv: 1906.04887, 2019.

[86] Andriy Myronenko. 3D MRI brain tumor segmentation using autoencoder regularization. In *Proc. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, 4th International Workshop Brainlesion 2018 Held in Conjunction with MICCAI 2018, Part II, Lecture Notes in Computer Science*, volume 11384, pages 311–320, 2018.

[87] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proc. 32nd International Conference on Machine Learning*, pages 448–456, 2015.

[88] Yuxin Wu and Kaiming He. Group normalization. *International Journal of Computer Vision*, 128:742–755, 2020.

[89] Eric P Tamm, Paul M Silverman, Chusilp Charnsangavej, and Douglas B Evans. Diagnosis, staging, and surveillance of pancreatic cancer. *American Journal of Roentgenology*, 180(5):1311–1323, 2003.

[90] Hoo-Chang Shin, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. Deep convolutional neural networks for computer-aided detection: CNN architectures,

dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35(5):1285–1298, 2016.

[91] Nima Tajbakhsh, Jae Y Shin, Suryakanth R Gurudu, R Todd Hurst, Christopher B Kendall, Michael B Gotway, and Jianming Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging*, 35(5):1299–1312, 2016.

[92] Ken Chang, Niranjan Balachandar, Carson Lam, Darvin Yi, James Brown, Andrew Beers, Bruce Rosen, Daniel L Rubin, and Jayashree Kalpathy-Cramer. Distributed deep learning networks among institutions for medical imaging. *Journal of the American Medical Informatics Association*, 25(8):945–954, 2018.

[93] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(12):2935–2947, 2017.

[94] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Proc. 20th International Conference on Artificial Intelligence and Statistics*, pages 1273–1282, 2017.

[95] Holger R Roth, Le Lu, Amal Farag, Hoo-Chang Shin, Jiamin Liu, Evrim B Turkbey, and Ronald M Summers. DeepOrgan: Multi-level deep convolutional networks for automated pancreas segmentation. In *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention 2015, Part I, Lecture Notes in Computer Science*, volume 9349, pages 556–564, 2015.

[96] Yuyin Zhou, Lingxi Xie, Wei Shen, Yan Wang, Elliot K Fishman, and Alan L Yuille. A fixed-point model for pancreas segmentation in abdominal CT scans. In *Proc. International Conference on Medical Image Computing and Computer-Assisted In-*

tervention 2017, Part I, Lecture Notes in Computer Science, volume 10433, pages 693–701, 2017.

[97] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention U-Net: Learning where to look for the pancreas. In *Proc. Medical Imaging with Deep Learning 2018*, 10 pages, 2018.

[98] Yunze Man, Yangsibo Huang, Junyi Feng, Xi Li, and Fei Wu. Deep learning driven CT pancreas segmentation with geometry-aware U-net. *IEEE Transactions on Medical Imaging*, 38(8):1971–1980, 2019.

[99] Pochuan Wang, Chen Shen, Holger R Roth, Dong Yang, Daguang Xu, Masahiro Oda, Kazunari Misawa, Po-Ting Chen, Kao-Lang Liu, Wei-Chih Liao, Weichung Wang, and Kensaku Mori. Automated pancreas segmentation using multi-institutional collaborative deep learning. In *Proc. Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning, 1st International Workshop DCL 2020 Held in Conjunction with MICCAI 2020, Lecture Notes in Computer Science*, volume 12444, pages 192–200, 2020.

[100] Elena Czeizler, Wolfgang Wiessler, Thorben Koester, Mikko Hakala, Shahab Basiri, Petr Jordan, and Esa Kuusela. Using federated data sources and varian learning portal framework to train a neural network model for automatic organ segmentation. *Physica Medica*, 72:39–45, 2020.

[101] Micah J Sheller, G Anthony Reina, Brandon Edwards, Jason Martin, and Spyridon Bakas. Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation. In *Proc. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, 4th International Workshop*

Brainles 2018 Held in Conjunction with MICCAI 2018, Part I, Lecture Notes in Computer Science, volume 11383, pages 92–104, 2018.

[102] Wenqi Li, Fausto Milletarì, Daguang Xu, Nicola Rieke, Jonny Hancox, Wentao Zhu, Maximilian Baust, Yan Cheng, Sébastien Ourselin, M Jorge Cardoso, and Andrew Feng. Privacy-preserving federated brain tumour segmentation. In *Proc. Machine Learning in Medical Imaging, 10th International Workshop MLMI 2019 Held in Conjunction with MICCAI 2019, Lecture Notes in Computer Science*, volume 11861, pages 133–141, 2019.

[103] Qi Dou, Tiffany Y So, Meirui Jiang, Quande Liu, Varut Vardhanabhuti, Georgios Kaissis, Zeju Li, Weixin Si, Heather H C Lee, Kevin Yu, Zuxin Feng, Li Dong, Egon Burian, Friederike Jungmann, Rickmer Braren, Marcus Makowski, Bernhard Kainz, Daniel Rueckert, Ben Glocker, Simon C H Yu, and Pheng Ann Heng. Federated deep learning for detecting COVID-19 lung abnormalities in CT: A privacy-preserving multinational validation study. *npj Digital Medicine*, 4(60):1–11, 2021.

[104] Ittai Dayan, Holger R Roth, Aoxiao Zhong, Ahmed Harouni, Amilcare Gentili, Anas Z Abidin, Andrew Liu, Anthony Beardsworth Costa, Bradford J Wood, Chien-Sung Tsai, Chih-Hung Wang, Chun-Nan Hsu, Cheng-Kuang Lee, Peiying Ruan, Daguang Xu, Dufan Wu, Eddie Huang, Felipe Campos Kitamura, Griffin Lacey, Gustavo César de Antônio Corradi, et al. Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature Medicine*, 27:1735–1743, 2021.

[105] Yingda Xia, Dong Yang, Wenqi Li, Andriy Myronenko, Daguang Xu, Hirofumi Obinata, Hitoshi Mori, Peng An, Stephanie Harmon, Evrim Turkbey, Baris Turkbey, Bradford Wood, Francesca Patella, Elvira Stellato, Gianpaolo Carrafiello, Anna

Ierardi, Alan Yuille, and Holger Roth. Auto-FedAvg: Learnable federated averaging for multi-institutional medical image segmentation. *Computing Research Repository arXiv Preprints*, arXiv: 2104.10195, 2021.

[106] Dong Yang, Ziyue Xu, Wenqi Li, Andriy Myronenko, Holger R Roth, Stephanie Harmon, Sheng Xu, Baris Turkbey, Evrim Turkbey, Xiaosong Wang, Wentao Zhu, Gianpaolo Carrafiello, Francesca Patella, Maurizio Cariati, Hirofumi Obinata, Hitoshi Mori, Kaku Tamura, Peng An, Bradford J Wood, and Daguang Xu. Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan. *Medical Image Analysis*, 70:1–12, 2021.

[107] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. In *Proc. 3rd Conference on Machine Learning and Systems*, volume 2, pages 429–450, 2020.

[108] Virginia Smith, Chao-Kai Chiang, Maziar Sanjabi, and Ameet Talwalkar. Federated multi-task learning. In *Proc. 31st International Conference on Neural Information Processing Systems*, pages 4427–4437, 2017.

[109] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology*, 10(2):1–19, 2019.

[110] Michelle Guo, Albert Haque, De-An Huang, Serena Yeung, and Li Fei-Fei. Dynamic task prioritization for multitask learning. In *Proc. 15th European Conference on Computer Vision, Part XVI, Lecture Notes in Computer Science*, volume 11220, pages 282–299, 2018.

[111] Shikun Liu, Edward Johns, and Andrew J Davison. End-to-end multi-task learn-

ing with attention. In *Proc. 2019 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1871–1880, 2019.

[112] Amber L Simpson, Michela Antonelli, Spyridon Bakas, Michel Bilello, Keyvan Farahani, Bram Van Ginneken, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M Summers, Patrick Bilic, Patrick F Christ, Richard K G Do, Marc Gollub, Jennifer Golia-Pernicka, Heckers, Stephan H, William R Jarnagin, Maureen K McHugo, Sandy Napel, Eugene Vorontsov, Lena Maier-Hein, and M Jorge Cardoso. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *Computing Research Repository arXiv Preprints*, arXiv: 1902.09063, 2019.

[113] Bennett Landmanm, Zhoubing Xu, Eugenio Igelsias, Juan, Martin Styner, Robin Langerak, Thomas, and Arno Klein. 2015 MICCAI multi-atlas labeling beyond the cranial vault – Workshop and Challenge. `https://www.synapse.org/#!Synapse:syn3193805/wiki/89480/`, 2015. (Accessed on 2024/1/10).

[114] Qihang Yu, Dong Yang, Holger Roth, Yutong Bai, Yixiao Zhang, Alan L Yuille, and Daguang Xu. C2FNAS: Coarse-to-fine neural architecture search for 3D medical image segmentation. In *Proc. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4126–4135, 2020.

[115] Zhuotun Zhu, Yingda Xia, Lingxi Xie, Elliot K Fishman, and Alan L Yuille. Multi-scale coarse-to-fine segmentation for screening pancreatic ductal adenocarcinoma. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention 2019, Part VI, Lecture Notes in Computer Science*, volume 11769, pages 3–12, 2019.

[116] Yuyin Zhou, Yingwei Li, Zhishuai Zhang, Yan Wang, Angtian Wang, Elliot K Fishman, Alan L Yuille, and Seyoun Park. Hyper-pairing network for multi-phase

pancreatic ductal adenocarcinoma segmentation. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention 2019, Part 2, Lecture Notes in Computer Science*, volume 11765, pages 155–163, 2019.

[117] Yingda Xia, Qihang Yu, Wei Shen, Yuyin Zhou, Elliot K Fishman, and Alan L Yuille. Detecting pancreatic ductal adenocarcinoma in multi-phase CT scans via alignment ensemble. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention 2020, Part III, Lecture Notes in Computer Science*, volume 12263, pages 285–295, 2020.

[118] Ling Zhang, Yu Shi, Jiawen Yao, Yun Bian, Kai Cao, Dakai Jin, Jing Xiao, and Le Lu. Robust pancreatic ductal adenocarcinoma segmentation with multi-institutional multi-phase partially-annotated CT scans. In *Proc. International Conference of Medical Image Computing and Computer-Assisted Intervention 2020, Part IV, Lecture Notes in Computer Science*, volume 12264, pages 491–500, 2020.

[119] Yan Wang, Peng Tang, Yuyin Zhou, Wei Shen, Elliot K Fishman, and Alan L Yuille. Learning inductive attention guidance for partially supervised pancreatic ductal adenocarcinoma prediction. *IEEE Transactions on Medical Imaging*, 40(10):2723–2735, 2021.

[120] Tahereh Mahmoudi, Zahra Mousavi Kouzahkanan, Amir Reza Radmard, Raheleh Kafieh, Aneseh Salehnia, Amir H Davarpanah, Hossein Arabalibeik, and Alireza Ahmadian. Segmentation of Pancreatic Ductal AdenoCarcinoma (PDAC) and surrounding vessels in CT images using deep convolutional neural networks and texture descriptors. *Scientific Reports*, 12(3092):1–14, 2022.

[121] Mingxing Tan and Quoc Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proc. 36th International Conference on Machine Learning*, pages 6105–6114, 2019.

[122] Holger R Roth, Chen Shen, Hirohisa Oda, Takaaki Sugino, Masahiro Oda, Yuichiro Hayashi, Kazunari Misawa, and Kensaku Mori. A multi-scale pyramid of 3D fully convolutional networks for abdominal multi-organ segmentation. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention 2018, Part IV, Lecture Notes in Computer Science*, volume 11073, pages 417–425, 2018.

[123] Ashish Sinha and Jose Dolz. Multi-scale self-guided attention for medical image segmentation. *IEEE Journal of Biomedical and Health Informatics*, 25(1):121–130, 2021.

[124] Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, and Zhuowen Tu. Deeply-Supervised Nets. In *Proc. 18th International Conference on Artificial Intelligence and Statistics*, volume 38, pages 562–570, 2015.

[125] Jo Schlemper, Ozan Oktay, Michiel Schaap, Mattias P Heinrich, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention gated networks: Learning to leverage salient regions in medical images. *Medical Image Analysis*, 53:197–207, 2019.

[126] Abhijit Guha Roy, Nassir Navab, and Christian Wachinger. Concurrent spatial and channel squeeze & excitation in fully convolutional networks. In *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention 2018, PartI, Lecture Notes in Computer Science*, volume 11070, pages 421–429, 2018.

[127] Brian McFee, Justin Salamon, and Juan Pablo Bello. Adaptive pooling operators for weakly labeled sound event detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(11):2180–2193, 2018.

[128] Jonas Lamy, Odyssée Merveille, Bertrand Kerautret, Nicolas Passat, and Antoine Vacavant. Vesselness filters: A survey with benchmarks applied to liver imaging. In *Proc. 25th International Conference on Pattern Recognition*, pages 3528–3535.

[129] Alejandro F Frangi, Wiro J Niessen, Koen L Vincken, and Max A Viergever. Multiscale vessel enhancement filtering. In *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention 1998, Lecture Notes in Computer Science*, volume 1496, pages 130–137, 1998.

[130] Daniel Jimenez-Carretero, Andres Santos, Sjoerd Kerkstra, Rina Dewi Rudyanto, and Maria J Ledesma-Carbayo. 3D Frangi-based lung vessel enhancement filter penalizing airways. In *Proc. IEEE 10th International Symposium on Biomedical Imaging*, pages 926–929, 2013.

[131] Liang-Chieh Chen, Yi Yang, Jiang Wang, Wei Xu, and Alan L Yuille. Attention to scale: Scale-aware semantic image segmentation. In *Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3640–3649, 2016.

[132] Yukitaka Nimura, Daisuke Deguchi, Takayuki Kitasaka, Kensaku Mori, and Yasuhito Suenaga. Pluto: A common platform for computer-aided diagnosis. *Medical Imaging Technology*, 26(3):187–191, 2008.

[133] S A Hojjatoleslami and Josef Kittler. Region growing: A new approach. *IEEE Transactions on Image Processing*, 7(7):1079–1084, 1998.

[134] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith

Chintala. Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32:8024–8035, 2019.

[135] Chen Shen, Holger R Roth, Yuichiro Hayashi, Masahiro Oda, Tadaaki Miyamoto, Gen Sato, and Kensaku Mori. A cascaded fully convolutional network framework for dilated pancreatic duct segmentation. *International Journal of Computer Assisted Radiology and Surgery*, 17(2):343–354, 2022.

[136] Chen Shen, Holger R Roth, Yuichiro Hayashi, Masahiro Oda, Tadaaki Miyamoto, Gen Sato, and Kensaku Mori. Attention-guided pancreatic duct segmentation from abdominal CT volumes. In *Clinical Image-Based Procedures, Distributed and Collaborative Learning, Artificial Intelligence for Combating COVID-19 and Secure and Privacy-Preserving Machine Learning, 10th Workshop CLIP 2021 Held in Conjunction with MICCAI 2021, Lecture Notes in Computer Science*, volume 12969, pages 46–55, 2021.

[137] Se Woo Kim, Se Hyung Kim, Dong Ho Lee, Sang Min Lee, Yeon Soo Kim, Jin Young Jang, and Joon Koo Han. Isolated main pancreatic duct dilatation: CT differentiation between benign and malignant causes. *American Journal of Roentgenology*, 209(5):1046–1055, 2017.

[138] Naohiro Sata, Katsumi Kurihara, Masaru Koizumi, Munetoshi Tsukahara, Koji Yoshizawa, and Hideo Nagai. CT virtual pancreatoscopy: A new method for diagnosing Intraductal Papillary Mucinous Neoplasm (IPMN) of the pancreas. *Abdominal Imaging*, 31(3):326–331, 2006.

[139] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30:6000–6010, 2017.

[140] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R Roth, and Daguang Xu. UNETR: Transformers for 3D medical image segmentation. In *Proc. 2022 IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 574–584, 2022.

[141] Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger R Roth, and Daguang Xu. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In *Proc. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, 7th International Workshop BrainLes 2021 Held in Conjunction with MICCAI 2021, Part I, Lecture Notes in Computer Science*, volume 12962, pages 272–284, 2021.

[142] Stefanie Warnat-Herresthal, Hartmut Schultze, Krishnaprasad Lingadahalli Shastry, Sathyanarayanan Manamohan, Saikat Mukherjee, Vishesh Garg, Ravi Sarveswara, Kristian Händler, Peter Pickkers, N Ahmad Aziz, Sofia Ktena, Florian Tran, Michael Bitzer, Stephan Ossowski, Nicolas Casadei, Christian Herr, Daniel Petersheim, Uta Behrends, Fabian Kern, Tobias Fehlmann, et al. Swarm learning for decentralized and confidential clinical machine learning. *Nature*, 594(7862):265–270, 2021.

[143] Mark Ramaekers, Christiaan GA Viviers, Boris V Janssen, Terese AE Hellström, Lotte Ewals, Kasper van der Wulp, Joost Nederend, Igor Jacobs, Jon R Pluyter, Dimitrios Mavroeidis, Fons van der Sommen, Marc G Besselink, and Misha D P Luyer. Computer-aided detection for pancreatic cancer diagnosis: Radiological challenges and future directions. *Journal of Clinical Medicine*, 12(13):4209–4224, 2023.