EPFL

MASTER THESIS

Deep Learning for Computed Tomography Scans

Author:
Pierre Vincent NAAYEM

Supervisor:
Prof. Alexandre ALAHI

A thesis submitted in fulfillment of the requirements for the degree of Master of Robotics

in the

Visual Intelligence for Transportation Laboratory
Robotics

January 31, 2023

EPFL

Abstract

ENAC & STI
Robotics

Master of Robotics

Deep Learning for Computed Tomography Scans

by Pierre Vincent NAAYEM

Medical image segmentation is an important and challenging task in the field of medical image processing, with numerous applications in areas such as diagnosis, treatment planning, and image-guided surgery. In recent years, deep learning has emerged as a powerful tool for medical image segmentation, with a range of architectures and frameworks being developed to address this problem. This paper presents an overview of the current state-of-the-art in medical image segmentation with a focus on the segmentation of anatomical structures in CT scans. The paper begins by introducing the basics of deep learning and its relevance to medical imaging, before reviewing a selection of open source datasets that are commonly used for medical image segmentation. The paper then explores a range of architectures and frameworks that have been developed for medical image segmentation, including U-Net and its variants, the nnUNet framework, and transformer-based models such as UNETR and Swin UNETR. Additionally, we conduct experiments to compare the performance of the established state-of-the-art nnU-Net with a newer architecture, Swin-UNet, and also to approximate the number of training images required to obtain the best or satisfactory results. Finally, the paper discusses current trends and future directions in the field, including the use of transformer-based models and attention mechanisms, the importance of inter-rater reliability, and the potential of self-supervised learning and domain adaptation. This paper aims to provide a comprehensive overview of the current state-of-the-art in medical image segmentation with CT scans, highlighting key developments and trends in the field, and identifying opportunities for future research and development.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Professor Alexandre Alahi, for his invaluable guidance, support and encouragement throughout the development of this paper. I am extremely grateful for the opportunity to work in his laboratory, the Visual Intelligence for Transportation (VITA), and for the valuable insights and experience I have gained during this process.

Additionally, I would like to thank Dr. Reza and Dr. Saab for their insights and guidance throughout the development of this thesis. Their expertise and knowledge have been invaluable in shaping my research and direction.

I would also like to extend my gratitude to Fadi Saikali from RapidAI for taking the time to listen, comment and evaluate my work.

I would also like to thank Dr. Saeed Saadatnejad, Frano Rajič and the team from the VITA lab for their practical advice and valuable insights.

My gratitude also goes to my family and friends for their continuous support and for their help in reviewing this paper.

I would also like to acknowledge the Medical Imagery Processing community, particularly the researchers from DKFZ and the MONAI project, for providing open source tools and valuable documentations. This work would not have been possible without their contributions.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The task of medical image segmentation has numerous applications in the field of medical imaging, including diagnosis, treatment planning, and image-guided surgery. In this paper, we will focus on the segmentation of anatomical structures in CT scans using deep learning-based approaches. CT scans provide detailed images of many internal structures of the body with higher resolution and contrast compared to traditional X-rays. They can be used to: diagnose conditions – including damage to bones, injuries to internal organs, problems with blood flow, stroke, and cancer.

CT scanning provides medical information that is different from other imaging examinations, such as ultrasound, MRI, SPECT, PET or nuclear medicine. Each imaging technique has advantages and limitations. The principal advantages of CT are its abilities to: Rapidly acquire images; Provide clear and specific information; Image a small portion or all the body during the same examination. No other imaging procedure combines these advantages into a single session. (Radiology (ACR), n.d.)

We aim to provide fellow researchers a clear overview of the state-of-the-art and of the options available to develop a deep learning framework and to collect data in order to train the models and achieve the best results. By presenting a comprehensive review of the current developments in the field, our goal is to provide a valuable resource for researchers and practitioners working in the field of medical image segmentation and deep learning. We will attempt to answer the following research question: *"How can we leverage the latest advancements in medical segmentation to develop an effective framework for delineating anatomical structures on CT scan images?"*

To achieve these objectives, we will review the available open source datasets for CT scan segmentation and compare the performance of various deep learning-based architectures and frameworks. This includes popular models such as U-Net and its variants, the nnUNet framework, and transformer-based models such as UNETR and Swin UNETR. We will also present the key contributors and recent trends in the field, and conduct experiments to evaluate the performance of these models on the datasets. Finally, we will detail noticeable trends and area of improvements and attempt to define a strategy in order to develop a deep learning framework.

1.1 Deep Learning: A Recap

Deep learning is a subfield of artificial intelligence (AI) that involves the use of neural networks, which are computing systems that are inspired by the structure and function of the human brain (Gerven and Bohte, 2017). Artificial neural networks

(ANNs) are a type of computer program that is designed to mimic the way the human brain works. They are used to solve complex problems and can learn from data. ANNs have been studied for more than 70 years (McCulloch and Pitts, 1943) and have recently become popular again as a way to recognize patterns (LeCun, Bengio, and Hinton, 2015) . They are able to learn complex, non-linear functions and can produce good results with minimal human input. They are often used to analyze large amounts of data and can provide insights that would otherwise be difficult to obtain (Lawrence et al., 1997)(Long, Shelhamer, and Darrell, 2015)(Donahue et al., 2017)(Wu, He, and Sun, 2015)(Diba et al., 2017)(Ouyang et al., 2015) (Sengupta et al., 2020).

1.1.1 What is an Artificial Neural Network?

An (ANN) is made up of many interconnected, simple functional units, or neurons, that act in concert to solve classification or regression problems. Classification problems involve separating the input space (the range of all possible values of the inputs) into a discrete number of classes, while regression problems involve approximating the function (the black box) that maps inputs to outputs. ANNs are made up of layers of neurons that interact with the environment, process information, and relay processed information back out to the output. Each neuron is connected to the outputs of a subset of the neurons in the previous layer, and the output of the layer is an m-dimensional vector computed as the input vector pre-multiplied by an m×n matrix of weights. To classify inputs non-linearly or to approximate a non-linear function with a regression, each neuron adds a numerical bias value to the result of its input sum of products and passes that through a nonlinear activation function. The ANN learns an approximation to the function that produced each of the outputs from its corresponding input by optimizing over input-output pairs to minimize the error function. (Cybenko, 1989)(Hornik, 1991)(Lu et al., 2017)(Hanin, 2019)(Sengupta et al., 2020)

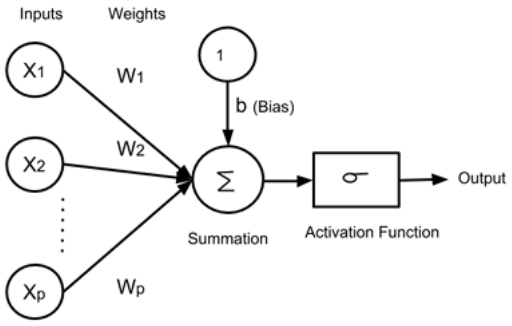


FIGURE 1.1: The Perceptron Learning Model

1.1.2 How do these networks learn?

These neural networks are capable of learning by changing the distribution of weights in order to approximate a function that represents the patterns in the input data. In other words, the neural network is "stimulated" with new data and the weights are adjusted in order to improve the accuracy of the network's predictions. This process continues until the error in approximation is below a certain threshold for example,

at which point the learning process is considered to be complete. The term "deep" in deep learning refers to the complexity of the neural networks, which typically have many hidden layers in order to learn detailed representations of the data. Deep learning has become an important part of AI research due to its ability to handle large amounts of data and its ability to generalize to new problems. (Schmidhuber, 2015)(Marcus, 2018)(Papernot et al., 2015)(Abbe and Sandon, 2019).

1.1.3 Why are deep neural networks garnering so much attention now?

Deep neural networks are a type of artificial intelligence that have been around for a long time, but have recently gained a lot of attention from academics and industry. This is due to a number of factors, such as the availability of large datasets with high quality labels, advances in parallel computing capabilities, and software platforms that make it easier to integrate architectures into a GPU computing framework. Additionally, better regularization techniques have been developed to help avoid overfitting, and robust optimization algorithms have been developed to produce near-optimal solutions. Table 1.1 provides a summary of the main advancements in the field.

TABLE 1.1: Some Key Advances in Neural Networks Research

People Involved	Contribution
McCulloch & Pitts	ANN models with adjustable weights (1943) (McCulloch and Pitts, 1943)
Rosenblatt	The Perceptron Learning Algorithm (1957) (Rosenblatt, 1958)
Widrow & Hoff	Adaline (1960), Madaline Rule I (1961) & Madaline Rule II (1988) (Winter and Widrow, 1988)(Widrow and Lehr, 1990)
Minsky & Papert	The XOR Problem (1969) (Nievergelt, 1969)
Werbos (Doctoral Dissertation)	Backpropagation (1974) (Werbos, 1994)
Hopfield	Hopfield Networks (1982) (Hopfield, 1982)
Rumelhart, Hinton & Williams	Renewed interest in backpropagation: multilayer adaptive backpropagation (1986) (Rumelhart, McClelland, and PDP Research Group, 1986)
Vapnik, Cortes	Support Vector Networks (1995) (Cortes and Vapnik, 1995)
Hochreiter & Schmidhuber	Long Short Term Memory Networks (1997) (Hochreiter and Schmidhuber, 1997)
LeCunn et. al	Convolutional Neural Networks (1998) (Lecun et al., 1998)
Hinton & Ruslan	Hierarchical Feature Learning in Deep Neural Networks (2006) (Hinton, Osindero, and Teh, 2006)

1.1.4 Why is deep learning relevant to medical imaging?

Deep learning algorithms are capable of learning from large amounts of data and can be used to perform a wide range of tasks, such as image and speech recognition, natural language processing, and predictive modeling. In recent years, deep learning has achieved impressive results in many different domains and has been applied to a variety of tasks, including medical image analysis. Semantic segmentation is a specific type of deep learning algorithm that is used to analyze and interpret medical images, such as CT scans and MRI scans, in order to identify and classify different structures and tissues within the body. By using semantic segmentation, medical professionals can gain a better understanding of the structure and function of the human body, which can aid in the diagnosis and treatment of diseases. (Sengupta et al., 2020)

1.2 Semantic Segmentation

Semantic segmentation is the process of assigning a semantic label, or category, to each pixel in an image. For example, in an image of a cityscape, semantic segmentation could be used to assign labels such as "building," "road," "sky," etc. to each pixel in the image (Cordts et al., 2015). This allows a machine learning model to understand the content of the image at a pixel-level granularity, which can be useful for a variety of applications such as object detection, image classification, and scene

understanding.

One variant of semantic segmentation is called instance segmentation, which goes one step further by assigning a unique label to each individual object instance in an image. For example, in an image with multiple cars, instance segmentation would assign a unique label to each car rather than just the "car" category in general. This allows the model to differentiate between individual objects and better understand the relationships between them in the image.

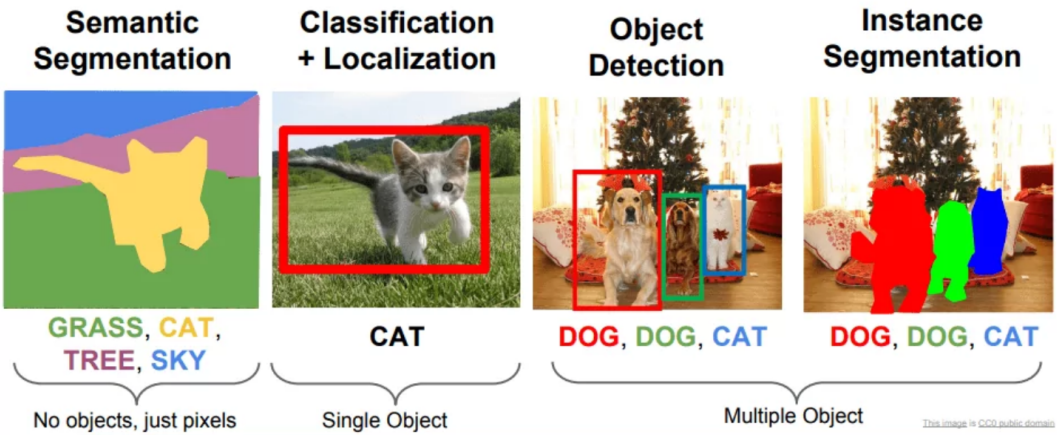


FIGURE 1.2: Illustration of segmentation along other computer vision tasks (**Stanford**)

Semantic segmentation is a fundamental task in medical image processing that involves dividing medical images into multiple regions, each corresponding to a different semantic concept or class of tissue or structure. It has numerous applications in areas such as tumor segmentation (Ranjbarzadeh et al., 2021), organ localization (Xu et al., 2019), and surgical planning. In recent years, there have been significant advances in the field of semantic segmentation, particularly with the use of convolutional neural networks (CNNs) (Wang, Huang, and Xu, 2010),(Wang et al., 2019). Traditional methods such as random forest and conditional random field (CRF) classifiers have been largely replaced by deep learning methods such as fully convolutional networks (FCN) (Long, Shelhamer, and Darrell, 2015), (Shelhamer, Long, and Darrell, 2017) and U-nets (Ronneberger, Fischer, and Brox, 2015), (Zhou et al., 2020), which have demonstrated improved performance and the ability to be trained end-to-end (Zhao, Huang, and Sun, 2004).

However, the high cost and time required for pixel-level annotation can be a limiting factor in the development of these models, leading to the emergence of weakly-supervised learning techniques. Additionally, the generalization ability of these models can be limited by the assumption that the training and test datasets are independent and identically distributed, leading to the development of domain adaptation techniques. The use of data from multiple modalities, such as magnetic resonance imaging (MRI) and computed tomography (CT), has also been explored as a way to enhance the performance of semantic segmentation in medical images through multi-modal data fusion. Finally, the development of real-time semantic segmentation models that can perform the task quickly enough for use in real-time applications, such as during surgery, has also been a focus of research.

1.3 Review Methodology

In order to address our research question, we conducted a comprehensive literature review to gather information on the latest advancements in medical segmentation for CT scan images. Our literature search was conducted using various databases, including (*Nature* 2023), (*PubMed* n.d.), (*ScienceDirect.com | Science, health and medical journals, full text articles and books.* N.d.) and (*arXiv.org e-Print archive* n.d.). We also relied on the benchmark repository (*Papers with Code - The latest in Machine Learning* n.d.) to find relevant and up-to-date performance results. During our review, we compared the models and datasets used in relevant papers and evaluated their performance using metrics and benchmarks. One of the main difficulties we encountered was finding the intersection between machine learning practitioners and radiologists or medical professionals. To overcome this challenge, we engaged in networking and actively sought out specialists in the field to gain their insights. We then approached specialists in radiology and asked for their advice and requirements to produce a framework that would be helpful to them.

In addition to literature review, we also conducted our own experiments with models and data. This allowed us to compare our results to the papers we reviewed and verify the completeness of information or highlight any inconsistencies. This process also allowed us to encounter problems, generate questions, and find solutions, which amplified the creative and research process. We also looked beyond the medical domain and studied models that were state-of-the-art in other applications of segmentation. This allowed us to find transformer-based models and understand their potentials.

For our experimentation, we replicated the results of the swin UNETR (Hatamizadeh et al., 2022) and nnU-Net models (Isensee et al., 2021) on the WORD dataset using the same methodologies as described in the original WORD paper (Luo et al., 2022). We also conducted a learning curve experiment where we replicated the results of the Totalsegmentator paper (Wasserthal, 2022) in a similar fashion and gradually reduced the amount of data to infer a minimum amount of training images required to obtain satisfactory results on typical anatomical structures.

We are aware of our limitations in this study. One limitation is the completeness of information in this paper, which is limited by the completeness of its sources. Additionally, we were limited in computer power and time, so a restrained number of tests had to be realized with trainings taking a week of GPU resource. Another important limitation is the opacity of the private sector, with private companies keeping data, code, architecture details for themselves.

Chapter 2

Open-Source Datasets

In this chapter, we will be introducing various open source datasets that are available for use in the segmentation of anatomical structures in CT scans. These datasets are important resources for developing and evaluating algorithms for automated organ segmentation, which is a crucial step in many medical image analysis tasks. We will present a summary of each dataset that we have collected and briefly summarize their contents and intended uses. Finally, we will provide a summary of all the datasets that we have collected into our database.

2.1 Multi-Atlas Labeling Beyond the Cranial Vault (BTCV)

The BTCV dataset consists of 100 clinically acquired CT images, with 50 images for segmentation of abdominal organs and another 50 images for segmentation of pelvic structures. Each scan has been manually annotated by two experienced undergraduate students and reviewed for label accuracy by a radiologist or radiation oncologist on a volumetric basis using the MIPAV software.

Concerning the abdominal set, 50 abdomen CT scans were randomly acquired from multiple scanners at the Vanderbilt University Medical Center under Institutional Review Board (IRB) supervision. The acquisition was part of an ongoing colorectal cancer chemotherapy trial, and a retrospective ventral hernia study. The scans were captured during portal venous contrast phase with volume sizes ranging from $512 \times 512 \times 85$ to $512 \times 512 \times 198$ and field of views approximately $280 \times 280 \times 280 mm^3$ to $500 \times 500 \times 650 mm^3$. The in-plane resolution ranges from $0.54x0.54 mm^2$ to $0.98x0.98 mm^2$, while the slice thickness ranges from $2.5mm$ to $5.0mm$. The registration data was generated by NiftyReg.

Concerning the cervix set, the dataset consists of planning CT scans of cervical cancer patients that were in varying stages of the disease but that were all eligible for radiotherapy. Images were acquired from scanners at the Erasmus Medical Center Cancer Institute in Rotterdam. The CT scans consist of between 148 and 241 axial slices (depending on body size) of 512×512 voxels. Voxel resolution was either 1.27×1.27 or 0.977×0.977, both with a slice thickness of $2.5mm$. The scans were acquired with a full-bladder drinking protocol. Most patients were scanned in prone position using a belly board, but for practical reasons some patients were scanned in supine position (patient numbers 1577656, 1689606, 2609008, 5463446, 5681868, 6770262, 6798630). The registration data was generated by ElastiX.

Both sets of images were divided into training and testing cohorts pseudo randomly to ensure that data from all scanners was included in both sets. They are intended to be used to evaluate the performance of automated segmentation approaches and provide researchers an opportunity to characterize their methods on a standardized dataset. The BTCV dataset is available on synapse.org. (*Multi-Atlas Labeling Beyond the Cranial Vault - Workshop and Challenge - syn3193805 - Wiki* n.d.)

Labels:

- Abdomen: spleen, right kidney, left kidney, gallbladder, esophagus, liver, stomach, aorta, inferior vena cava, portal vein and splenic vein, pancreas, right adrenal gland, left adrenal gland.
- Cervix: bladder, uterus, rectum, small bowel.

Note: Some patients may not have right kidney or gallbladder, and thus are not labeled.

2.2 The Liver Tumor Segmentation Benchmark (LiTS)

The Liver Tumor Segmentation (LiTS) (Bilic et al., 2023) benchmark is a dataset of CT images of the liver containing both healthy and abnormal (i.e. tumor) tissue. It was created to evaluate and compare the performance of automated segmentation algorithms for liver tumors and consists of 140 CT scans, 101 of which contain liver tumors. The images have been manually annotated by experts to provide ground truth segmentations of the liver and liver tumors. LiTS is intended for research and development of automated liver tumor segmentation algorithms and is available for download at the following website: (*CodaLab - Competition* n.d.).

The annotations for the image datasets were created using a manual protocol in which a radiologist with more than 3 years of experience in oncologic imaging used the ITK-SNAP software to label the datasets slice by slice, assigning either the label "Tumor" or "Healthy Liver" to each slice. The "Tumor" label included any neoplastic lesion, regardless of its origin (i.e. both primary liver tumors and metastatic lesions). Any part of the image that was not assigned one of these labels was considered "Background." The segmentations were then reviewed by three additional readers who were blinded to the initial segmentation, with the most senior reader serving as a tie-breaker in cases of labeling conflicts. Scans with very small and uncertain lesion-like structures were excluded from the annotation process. This annotation protocol was used to create the final set of annotations for the image datasets.

The LiTS dataset is diverse in terms of the types of liver tumor diseases it covers, including primary tumors such as hepatocellular carcinoma and cholangiocarcinoma, as well as secondary liver tumors such as metastases from colorectal, breast, and lung primary cancers. The tumors have varying lesion-to-background ratios and the images represent a mixture of pre- and post-therapy abdominal CT scans acquired with different CT scanners and protocols, including imaging artifacts commonly found in real-world clinical data. The in-plane image resolution ranges from $0.56mm$ to $1.0mm$ and the slice thickness ranges from $0.45mm$ to $6.0mm$. The number of axial slices also varies, ranging from 42 to 1026. The number of tumors per scan varies between 0 and 12, and the size of the tumors varies between $38mm^3$ and

$1231mm^3$. The test set has a higher number of tumor occurrences compared to the training set, but the liver volumes in the two sets are not significantly different. The LiTS dataset is licensed as CC BY-NC-SA and is available for download on the LiTS Challenge website. Ethics approval was not required for the creation of the dataset as all parties agreed to make the data publicly available and the images have been anonymized and reviewed for personal identifiers.

The CT scans in the Liver Tumor Segmentation (LiTS) benchmark dataset were collected from seven clinical sites worldwide, including the Rechts der Isar Hospital at the Technical University of Munich in Germany, the Radboud University Medical Center in the Netherlands, Polytechnique Montr´eal and the CHUM Research Center in Canada, the Sheba Medical Center in Israel, the Hebrew University of Jerusalem in Israel, the Hadassah University Medical Center in Israel, and IRCAD in France. The distribution of the number of scans per institution is described in Table 2 of the LiTS paper (Bilic et al., 2023).

Labels: Lung, Bones, Liver, Bladder, Kidney, Brain

2.3 AbdomenCT-1K

AbdomenCT-1K is a dataset of abdominal CT scans for use in the development and evaluation of automated organ segmentation algorithms. It consists of 1112 3D CT scans from five existing datasets: LiTS (201 cases), KiTS (300 cases), MSD Spleen (61 cases) and Pancreas (420 cases), NIH Pancreas (80 cases), and a new dataset from Nanjing University (50 cases). The 50 CT scans in the Nanjing University dataset are from 20 patients with pancreas cancer, 20 patients with colon cancer, and 10 patients with liver cancer. The number of plain phase, artery phase, and portal phase scans are 18, 18, and 14 respectively. The CT scans have resolutions of 512×512 pixels with varying pixel sizes and slice thicknesses between $1.25 - 5mm$, acquired on GE multi-detector spiral CT. The AbdomenCT-1K dataset is annotated using a hierarchical strategy to improve label consistency. Specifically, 15 junior annotators (with one to five years of experience) refine the segmentation results using ITK-SNAP 3.6 under the supervision of two board-certified radiologists. A senior radiologist with more than 10 years of experience then verifies and refines the annotations. To further reduce inter-rater variability, the annotators are required to learn the existing organ annotation protocols, and the senior radiologist checks and revises any obvious label errors in the existing datasets. Five-fold cross-validation U-Net models are also trained to identify possible segmentation errors, with cases with low DSC or NSD scores double-checked by the senior radiologist. In addition, two radiologists annotate 50 cases in the Nanjing University dataset, with their inter-rater variability presented in Table 2.1 of (Ma et al., 2021).

Organ	Liver	Kidney	Spleen	Pancreas
DSC (%)	98.4 ± 0.52	98.7 ± 0.53	98.6 ± 0.84	93.8 ± 7.78
NSD (%)	95.7 ± 3.04	98.7 ± 2.05	98.2 ± 4.18	92.5 ± 9.40

FIGURE 2.1: Quantitative analysis of inter-rater variability between two radiologists.

Labels: Spleen, Kidney, Liver, Pancreas

2.4 Whole abdominal ORgan Dataset (WORD)

The WORD dataset (Luo et al., 2022) is a large-scale dataset for whole abdominal organ segmentation. It contains 150 abdominal CT volumes with 30495 slices, and each volume has pixel-level annotations for 16 organs. The dataset also includes scribble-based sparse annotations. The 150 CT scans in the WORD dataset were collected from 150 patients before the radiation therapy in a single center. All of them are scanned by a SIEMENS CT scanner without appearance enhancement. The clinical characteristics of the WORD dataset are listed in Table 2.1. Each CT volume consists of 159 to 330 slices of 512×512 pixels, with an in-plane resolution of $0.976mm \times 0.976mm$ and slice spacing of $2.5mm$ to $3.0mm$, indicating that the WORD dataset is a very high-resolution dataset. All scans of WORD dataset are exhaustively annotated with 16 anatomical organs, an example is shown in Figure 2.2. All images were anonymized and approved by the ethics committee to protect privacy where all clinical treatment details have been deleted. A senior oncologist (with 7 years of experience) uses ITK-SNAP (Yushkevich et al., 2006) to delineate all organs slice-by-slice in axial view. After that, an expert in oncology (more than 20 years of experience) checks and revises these annotations carefully and discusses them in cases of disagreement to produce consensus annotations and further ensure the annotation quality. Finally, these consensus labels are released and used for methods or clinical application development and evaluation. Note that all annotations and consensus discussions obey the radiation therapy delineation guideline published by Radiation Therapy Oncology Group (RTOG).

Labels: liver, spleen, kidney (L), kidney (R), stomach, gallbladder, esophagus, pancreas, duodenum, colon, intestine, adrenal, rectum, bladder, head of the femur (L) and head of the femur (R)

Characteristics	Train (n=100)	Validation (n=20)	Test (n=30)
Age (median)	47 (28-75)	52 (32-78)	49 (26-72)
Male	63	12	13
Female	37	8	17
Prostatic cancer	28	7	10
Cervical cancer	29	6	5
Rectal cancer	26	3	8
Others	17	4	7

TABLE 2.1: Clinical characteristics of WORD. Others include some metastatic tumours, such as bone metastasis and soft tissue metastasis.

2.5 A Large-Scale Abdominal Multi-Organ Benchmark for Versatile Medical Image Segmentation (AMOS)

The A Large-Scale Abdominal Multi-Organ Benchmark for Versatile Medical Image Segmentation (AMOS) dataset (Ji et al., 2022) is a large-scale dataset for abdominal organ segmentation in medical images, with 500 CT and 100 MRI scans collected from multi-center, multi-vendor, multi-modality, multi-phase, and multi-disease patients. These scans are annotated with voxel-level labels for 15 abdominal organs.

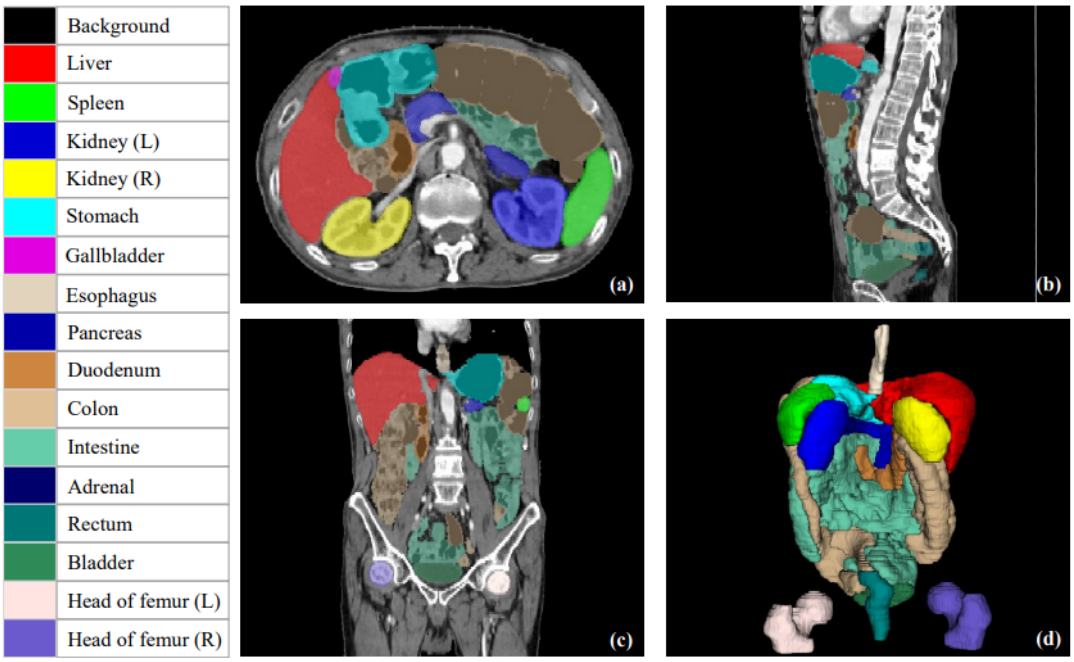


FIGURE 2.2: An example from WORD of 16 annotated abdominal organs in a CT scan. The left table lists the annotated organs' categories. (a), (b), (c) denote the visualization in axial, coronal, and sagittal views, respectively. (d) represents the 3D rendering results of annotated abdomen organs.

The CT/MRI data for AMOS were collected from 600 patients diagnosed with abdominal tumors/abnormalities at Longgang District People's Hospital and Longgang District Central Hospital in China. The data were collected retrospectively between 2018 and 2021 using eight different scanners following standard clinical acquisition protocols. The AMOS dataset includes CT scans from five different scanners: Aquilion ONE, Brilliance 16, Somatom Force, Optima CT660, and Optima CT540. The scans were randomly split into three groups: 200 scans for training, 100 scans for validation, and 78 scans for testing (In Distribution, ID). An additional 122 scans from the Optima CT660 and Optima CT540 scanners were used as unseen test data for testing (Out of Distribution, OOD). This was done in order to study the effect of domain shift, or the differences in imaging protocol, device vendors, and patient populations, on the ability of models to generalize to data from a new, unseen scanner. Overall, the AMOS dataset is split into two sub-datasets: AMOS-CT and AMOS-MRI. The data were split according to the scanner used for data acquisition. The AMOS-CT dataset contains 400 scans for training and 100 for testing.

The AMOS documentation (Ji et al., 2022) includes intensity and spatial statistics for the CT and MRI scans, as well as organ volume distribution data. The intensity and spatial statistics for the CT and MRI scans are presented in Tables 8 and 9 of the paper, respectively. Table 8 provides intensity and spatial statistics for the conventional abdominal organ segmentation datasets, including the AMOS-CT and AMOS-MRI sub-datasets. Table 9 provides intensity and spatial statistics for data generated from different scanners, including the AMOS-CT and AMOS-MRI sub-datasets. The organ volume distribution data is presented in Figure 4, which shows the distribution of organ volumes in the BTCV, Chaos, AbdomentCT-1K, and AMOS datasets.

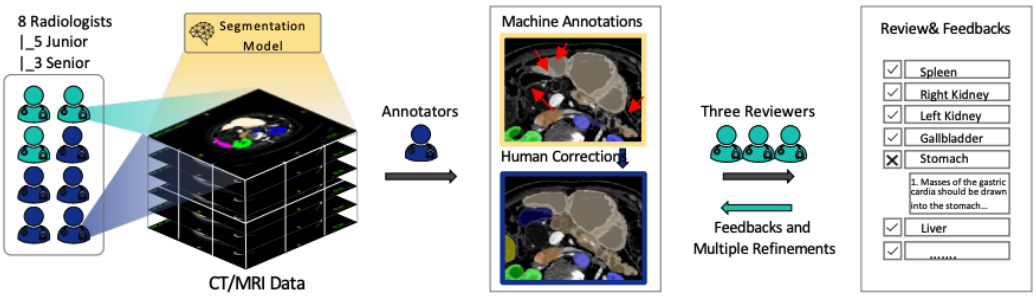


FIGURE 2.3: Annotation workflow of AMOS. The coarse annotations automatically labeled by pretrained segmentors will be further refined by human annotators for multiple times, including 5 junior radiologists for the initial stage and 3 senior specialists for the second checking stage.

The AMOS dataset was annotated using a semi-automatic annotation workflow (Figure 2.3), which involved two stages: a coarse labeling stage and a refinement stage. In the coarse labeling stage, 50 CT and 20 MRI scans were initially annotated by human labors, and then several representative models, such as 3D-UNet and VNet, were trained on these annotated scans to pre-label the remaining scans automatically and coarsely. In the refinement stage, five junior radiologists revisited the segmentation results on a case-by-case basis, and then three senior radiologists with more than 10 years of clinical experience conducted the final validation and annotation review. This process was iterated several times to achieve consensus on the final, high-quality annotations.

The goal of the AMOS dataset is to provide a realistic and diverse dataset that can challenge existing methods and inspire future research in the field of medical image segmentation. To this end, the dataset is diverse in terms of the types of abdominal diseases it covers and the scanners and protocols used for data acquisition. The dataset is intended to be a benchmark for the evaluation of multi-organ segmentation algorithms in practical applications, and the AMOS dataset, benchmark servers, and baselines are publicly available for download and use. (*Multi-Modality Abdominal Multi-Organ Segmentation Challenge 2022 - Grand Challenge* n.d.)

Labels: spleen, left kidney, right kidney, gallbladder, esophagus, liver, stomach, aorta, inferior vena cava, pancreas, adrenal glands, duodenum, bladder, prostate/uterus.

2.6 Combined (CT-MR) Healthy Abdominal Organ Segmentation (CHAOS)

The CHAOS dataset (Kavur et al., 2021) is a multi-modal collection of CT and MRI scans from healthy subjects that was assembled in conjunction with the IEEE International Symposium on Biomedical Imaging in 2019. It is used to evaluate the performance of deep learning models for abdominal organ segmentation. The CT scans in the dataset were acquired at the portal venous phase after contrast agent injection from 40 patients at the PACS of DEU Hospital. These scans are provided as 16-bit DICOM images with a resolution of 512x512 pixels, x-y spacing between

0.7-0.8 mm, and an inter-slice distance of 3 to 3.2 mm. Multiple scanners were used, including Philips SecuraCT with 16 detectors, Philips Mx8000 CT with 64 detectors, and Toshiba AquilionOne with 320 detectors. The patients in the dataset range in age from 18 to 76, with an average age of 44.85 for the CT group. The dataset includes healthy abdomen organs without any pathological abnormalities and was collected from the Department of Radiology at Dokuz Eylul University Hospital in Izmir, Turkey.

The CHAOS dataset includes 20 training cases and 20 testing cases in both the CT and MRI datasets, with T1-Dual and T2 SPIR sequences. The CT scans include annotations for the liver, while the MRI scans also have annotations for the left and right kidneys and spleen. The training data includes both DICOM images and their corresponding ground truth masks, while the testing data only includes DICOM images. The ground truth masks for the testing data are reserved for challenge validation and will not be publicly shared. (*CHAOS Challenge - combined (CT-MR) healthy abdominal organ segmentation - ScienceDirect* n.d.).

Labels: liver

2.7 TCIA Test & Validation Radiotherapy CT Planning Scan Dataset

The TCIA Test & Validation Radiotherapy CT Planning Scan Dataset is a collection of head and neck CT scans that have been segmented by trained radiographers according to a standard segmentation class definition. The validation and test sets were created as part of a collaboration between DeepMind and University College London Hospitals NHS Foundation Trust (UCLH) to apply deep learning to radiotherapy (Nikolov et al., 2021b). It consists of CT planning scans selected from two open source datasets available from The Cancer Imaging Archive (Clark et al., 2013): TCGA-HNSC (Zuley et al., 2016) and Head-Neck Cetuximab (Bosch et al., 2015). The in-plane pixel spacing ranged from 0.94mm to 1.27mm. Non-CT planning scans and those that did not meet the same slice thickness as a standard of 2.5mm were excluded.

The TCIA dataset includes a total of 31 scans, which were split into a validation set (7 scans, 6 patients) and a test set (24 scans, 24 patients). The average patient age in the validation set is 56.5 years and in the test set is 59.9 years. The dataset includes a mix of male and female patients. More dataset characteristics are available in the table 2 of (Nikolov et al., 2021b).

The ground truth labels for the TCIA dataset were produced by manually segmenting the full volumes of all 21 OARs included in the study. This was done initially by a radiographer with at least four years experience in the segmentation of head and neck OARs and then arbitrated by a second radiographer with similar experience. Further arbitration was then performed by a radiation oncologist with at least five years post-certification experience in head and neck radiotherapy. Organs at risk were selected according to the Brouwer Atlas (Brouwer et al., 2015), which defines a set of consensus guidelines for delineating OARs in head and neck radiotherapy. A total of 21 OARs were selected for inclusion in the study. The released test and validation set data are available at (*TCIA Test & Validation Radiotherapy CT*

Planning Scan Dataset 2022).

Labels: Brain, Brainstem, Cochlea (left), Cochlea (right), Lacrimal gland (left), Lacrimal gland (right), Lens (left), Lens (right), Lung (left), Lung (right), Mandible, Optic nerve (left), Optic nerve (right), Orbit (left), Orbit (right), Parotid gland (left), Parotid gland (right), Spinal canal, Spinal cord, Submandibular gland (left), Submandibular gland (right)

The TCGA-HNSC dataset is a collection of CT scans of patients with head and neck squamous cell carcinoma, and is available for download from The Cancer Imaging Archive. The Head-Neck Cetuximab dataset consists of CT scans of patients with head and neck cancer who were treated with the drug cetuximab, and is also available for download from The Cancer Imaging Archive. Both of these datasets contain additional data beyond what was used to create the TCIA Test & Validation Radiotherapy CT Planning Scan Dataset.

In addition to the TCIA dataset, the UCLH and PDDCA datasets are also available for use in this study. they have not been released, but may be requested from the authors of the paper. They have been labeled similarly to the TCIA dataset. The UCLH dataset consists of CT planning scans from patients at the University College London Hospitals NHS Foundation Trust. The scans were acquired between 2008 and 2016 and include both initial and re-scan images from patients who received radical radiotherapy treatment for head and neck cancer between 2008 and 2016. The UCLH dataset includes 663 scans and 389 patients in the training set. It was split into a training set (389 patients, 663 scans), validation set (51 patients, 100 scans) and test set (46 patients, 75 scans). The PDDCA dataset consists of CT scans from the Princess Alexandra Hospital in Brisbane, Australia. The PDDCA dataset includes 15 scans and 15 patients in the test set.

2.8 The Medical Segmentation Decathlon (MSD)

The medical segmentation challenge is a data set containing 2,633 images that have been de-identified and reformatted to the Neuroimaging Informatics Technology Initiative (NIFTI) format. The images are from multiple institutions and cover multiple anatomies and modalities and were acquired during real-world clinical applications. There are ten data sets in total, with each set containing between one and three region-of-interest (ROI) targets for a total of 17 targets. Two-thirds of the data for eight of the data sets were released as training sets (images and labels), while the remaining one-third was released as a test set (images without labels). The other two data sets (brain tumor and liver) were taken from well-known challenges and the original training/test split was preserved. The data sets cover a variety of anatomies and modalities, including brain, heart, hippocampus, liver, lung, prostate, pancreas, and bladder, and use imaging techniques such as magnetic resonance imaging (MRI) and computed tomography (CT). The target ROIs for each data set are described in detail in (Antonelli et al., 2022).

TABLE 2.2: Summary of the ten data sets of the Medical Segmentation Decathlon.

Phase	Task	Modality	Protocol	Target	# Cases (Train/Test)
Development phase	Brain	mp-MRI	FLAIR, T1w, T1 \w Gd, T2w	Edema, enhancing and non-enhancing tumor	750 4D volumes (484/266)
	Heart	MRI	–	Left atrium	30 3D volumes (20/10)
	Hippocampus	MRI	T1w	Anterior and posterior of hippocampus	394 3D volumes (263/131)
	Liver	CT	Portal-venous phase	Liver and liver tumor	210 3D volumes (131/70)
	Lung	CT	–	Lung and lung cancer	96 3D volumes (64/32)
	Pancreas	CT	Portal-venous phase	Pancreas and pancreatic tumor mass	420 3D volumes (282/139)
	Prostate	mp-MRI	T2, ADC	Prostate PZ and TZ	48 4D volumes (32/16)
Mystery phase	Colon	CT	Portal-venous phase	Colon cancer primaries	190 3D volumes (126/64)
	Hepatic Vessels	CT	Portal-venous phase	Hepatic vessels and hepatic tumor	443 3D volumes (303/140)
	Spleen	CT	Portal-venous phase	Spleen	61 3D volumes (41/20)

2.9 Totalsegmentator

The Totalsegmentator dataset (Wasserthal et al., 2022) consists of CT images that were randomly sampled from a PACS (Picture Archiving and Communication System) from the University Hospital Basel over the past 10 years. Only images from patients with general research consent were included, and images of legs and hands were excluded. The CT series of each examination was also randomly sampled. The images in the dataset therefore represent a diverse range of pathologies, scanners and sequences (native, arterial, portal venous, late phase, dual energy), with and without contrast agent, with different bulb voltages, slice thicknesses and resolutions, and with different kernels (soft tissue kernel, bone kernel). If the expert annotator was unsure how to segment certain structures due to high ambiguity, the examination was excluded (40 subjects). The anatomical structures that were annotated include 27 organs, 59 bones, 10 muscles, and 8 vessels, and are intended to cover a majority of relevant structures for most use cases. The annotations were created by a combination of manual segmentation by a radiologist and automatic segmentation using existing models, with manual refinement as necessary. The final dataset contains 1204 images, and the distribution of image types, age range, and acquisition sites is shown in Figure 2 of the paper. The dataset and a trained model for segmentation are available on GitHub (Wasserthal, 2022)

To annotate the scans, the authors first identified 104 anatomical structures that needed to be segmented. They then tried to use existing models to create initial segmentations for these structures, which were further refined manually if necessary. For the remaining structures that did not have existing models, manual segmentation was necessary. This manual segmentation was supervised by a board-certified

radiologist. The authors also used an in-house dataset to transfer segmentations between CTs with and without contrast agent and trained a U-Net on these images to accurately segment the heart subparts in all kinds of CT sequences. To speed up the manual segmentation process, the authors used the Nora imaging platform (Anastasopoulos, Reisert, and Kellner, 2017), which allowed them to quickly and easily label the images and create and save the segmentation masks. They also used an active learning approach, where a model was trained on a small number of manually labeled subjects and then used to correct the model predictions on the remaining subjects, reducing the amount of manual corrections needed. This process was repeated as more data was labeled, reducing the amount of manual correction required.

Labels: adrenal gland left , adrenal gland right , aorta , autochthon left , autochthon right , brain , clavicula left , clavicula right , colon , duodenum , esophagus , face , femur left , femur right , gallbladder , gluteus maximus left , gluteus maximus right , gluteus medius left , gluteus medius right , gluteus minimus left , gluteus minimus right , heart atrium left , heart atrium right , heart myocardium , heart ventricle left , heart ventricle right , hip left , hip right , humerus left , humerus right , iliac artery left , iliac artery right , iliac vena left , iliac vena right , iliopsoas left , iliopsoas right , inferior vena cava , kidney left , kidney right , liver , lung lower lobe left , lung lower lobe right , lung middle lobe right , lung upper lobe left , lung upper lobe right , pancreas , portal vein and splenic vein , pulmonary artery , rib left 1 , rib left 10 , rib left 11 , rib left 12 , rib left 2 , rib left 3 , rib left 4 , rib left 5 , rib left 6 , rib left 7 , rib left 8 , rib left 9 , rib right 1 , rib right 10 , rib right 11 , rib right 12 , rib right 2 , rib right 3 , rib right 4 , rib right 5 , rib right 6 , rib right 7 , rib right 8 , rib right 9 , sacrum , scapula left , scapula right , small bowel , spleen , stomach , trachea , urinary bladder , vertebrae C1 , vertebrae C2 , vertebrae C3 , vertebrae C4 , vertebrae C5 , vertebrae C6 , vertebrae C7 , vertebrae L1 , vertebrae L2 , vertebrae L3 , vertebrae L4 , vertebrae L5 , vertebrae T1 , vertebrae T10 , vertebrae T11 , vertebrae T12 , vertebrae T2 , vertebrae T3 , vertebrae T4 , vertebrae T5 , vertebrae T6 , vertebrae T7 , vertebrae T8 , vertebrae T9

2.10 MedSeg Datasets

MedSeg (*Database* n.d.) is a free, browser-based segmentation tool for CT and MRI that allows users to segment images using either manually or with the help of AI models. MedSeg maintains an open-access database of segmentations, including lateral ventricles, liver segments, inferior vena cava, brain vasculature, lung lobes and vessels, lymph node regions in the neck, and more. MedSeg is not intended for direct clinical use, but rather aims to accelerate the development of clinically validated models through its open-access segmentation database and other efforts. Developed by radiologists, MedSeg is designed to be user-friendly and efficient, with a vanilla JavaScript implementation that ensures fast performance and GPU hardware acceleration. In addition to its segmentation tool, MedSeg also offers a segmentation service for CT and MRI datasets, as well as research collaboration and the opportunity for users to contribute data for training AI models.

A table summarizing all the datasets and their contents is available in appendix A. Most of these datasets contain a single scan. The first dataset on brain ventricles was unfortunately not available at the time of writing.

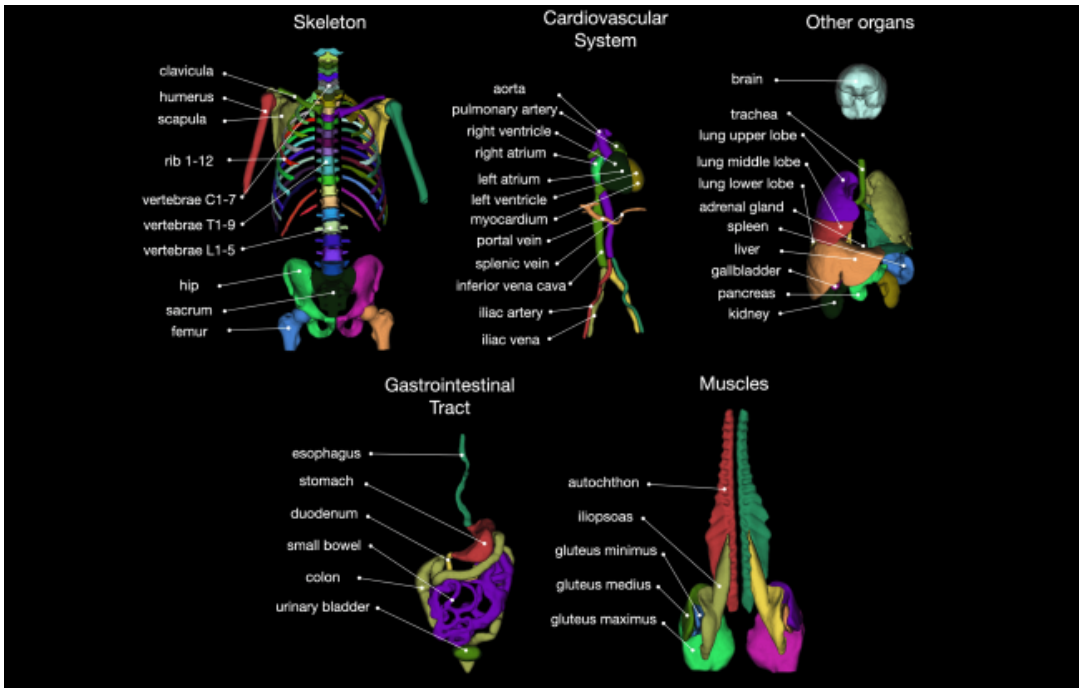


FIGURE 2.4: Overview of all 104 anatomical structures which are segmented in the TotalSegmentator dataset.

2.11 Summary of the Collected Datasets

A total of ten open source datasets have been collected, including BTCV, LiTS, AbdomenCT-1K, WORD, AMOS, CHAOS, TCIA Test & Validation Radiotherapy CT Planning Scan Dataset, MSD, Totalsegmentor, and MedSeg. All of these datasets were found through a thorough search of various online repositories and databases, including synapse.org and paperwithcode.com, for example. The collected datasets are described as follows:

1. The BTCV dataset consists of 100 clinically acquired CT images for the segmentation of abdominal and pelvic structures. Each scan have been manually annotated by trained raters and reviewed for label accuracy by a radiologist or radiation oncologist.

2. The LiTS dataset is a collection of 140 CT images of the liver which 101 contain liver tumor, including both healthy and abnormal tissue.

3. The AbdomenCT-1K dataset is a collection of over 1,000 CT scans of the abdominal region from 12 medical centers.

4. The WORD dataset is a large-scale dataset for whole abdominal organ segmentation, containing 150 abdominal CT volumes with pixel-level annotations for 16 organs.

5. The AMOS dataset is a collection of 500 CT scans of the abdomen and pelvis, with manual annotations for organs, vessels, and other structures.

6. The CHAOS dataset is a collection of 40 CT scans of the abdomen, with manual annotations for organs.

7. The TCIA Test & Validation Radiotherapy CT Planning Scan Dataset is a collection of 31 head and neck CT scans that have been segmented by trained radiographers according to a standard segmentation class definition.

8. The Medical Segmentation Decathlon (MSD) is a dataset containing a total of 1047 CT scans, consisting of 131 liver scans, 64 lung scans, 282 pancreas scans, 126 colon scans, 303 hepatic vessel scans, and 41 spleen scans, all with various annotations.

9. The Totalsegmentator dataset is a collection of 1204 CT images with manual annotations for 104 different anatomical structures, including 27 organs, 59 bones, 10 muscles, and 8 vessels. It was created by a combination of manual segmentation by a radiologist and automatic segmentation with manual refinement.

10. The MedSeg LS and IVC datasets consist of 50 and 20 CT scans, respectively, with annotations for liver segments and the inferior vena cava. Both datasets were collected in 2021 and are available through the MedSeg database.

All of these datasets have been widely used in the medical image analysis community and are intended to be used for research and development of automated segmentation algorithms.

TABLE 2.3: Number of samples per annotated organ in our database

Organ	Value
Adrenal gland (L)	1904
Adrenal gland (R)	1904
Aorta	1754
Autochthon (L)	1204
Autochthon (R)	1204
Bladder	690
Bones	140
Brain	1375
Brainstem	31
Clavicula (L)	1204
Clavicula (R)	1204
Cochlea (L)	31
Cochlea (R)	31
Colon	1354
Colon cancer primaries	126
Duodenum	1854
Esophagus	1904
Face	1204
Femur (L)	1204
Femur (R)	1204
Gallbladder	1904
Gluteus maximus (L)	1204
Gluteus maximus (R)	1204
Gluteus medius (L)	1204
Gluteus medius (R)	1204
Gluteus minimus (L)	1204
Gluteus minimus (R)	1204
Head of Femur (L)	150
Head of Femur (R)	150
Heart atrium (L)	1204
Heart atrium (R)	1204
Heart myocardium	1204
Heart ventricle (L)	1204
Heart ventricle (R)	1204
hepatic tumor	303
hepatic vessel	303
Hip (L)	1204
Hip (R)	1204
Humerus (L)	1204
Humerus (R)	1204
Iliac artery (L)	1204
Iliac artery (R)	1204
Iliac vena (L)	1204
Iliac vena (R)	1204
Iliopsoas (L)	1204
Iliopsoas (R)	1204
Inferior vena cava	1774
Intestine	150
Kidney (L)	3084
Kidney (R)	3084
Lacrimal gland (L)	31
Lacrimal gland (R)	31
Lens (L)	31
Lens (R)	31
Liver	3215
Liver Segments	50
Liver tumor	232
Lung	140
Lung (L)	31
Lung (R)	31
Lung tumor	64
Lung lower lobe (L)	1204
Lung lower lobe (R)	1204
Lung middle lobe (R)	1204
Lung upper lobe (L)	1204
Lung upper lobe (R)	1204
Mandible	31
Optic nerve (R)	31
Optic nerve (L)	31
Orbit (L)	31
Orbit (R)	31
Pancreas	3186
Pancreatic tumor mass	282
Parotid gland (L)	31
Parotid gland (R)	31
Portal vein and splenic vein	1754
Prostate	500
Pulmonary artery	1204
Rectum	200
Rib (L) 1	1204
Rib (L) 2	1204
Rib (L) 3	1204
Rib (L) 4	1204
Rib (L) 5	1204
Rib (L) 6	1204
Rib (L) 7	1204
Rib (L) 8	1204
Rib (L) 9	1204
Rib (L) 10	1204
Rib (L) 11	1204
Rib (L) 12	1204
Rib (R) 1	1204
Rib (R) 2	1204
Rib (R) 3	1204
Rib (R) 4	1204

Organ	Value
Rib (R) 5	1204
Rib (R) 6	1204
Rib (R) 7	1204
Rib (R) 8	1204
Rib (R) 9	1204
Rib (R) 10	1204
Rib (R) 11	1204
Rib (R) 12	1204
Sacrum	1204
Scapula (L)	1204
Scapula (R)	1204
Small bowel	1254
Spinal canal	31
Spinal cord	31
Spleen	2985
Stomach	1904
Submandibular gland (L)	31
Submandibular gland (R)	31
Trachea	1204
Urinary bladder	1204
Uterus	550
Vertebrae C1	1204
Vertebrae C2	1204
Vertebrae C3	1204
Vertebrae C4	1204
Vertebrae C5	1204
Vertebrae C6	1204
Vertebrae C7	1204
Vertebrae L1	1204
Vertebrae L2	1204
Vertebrae L3	1204
Vertebrae L4	1204
Vertebrae L5	1204
Vertebrae T1	1204
Vertebrae T10	1204
Vertebrae T11	1204
Vertebrae T12	1204
Vertebrae T2	1204
Vertebrae T3	1204
Vertebrae T4	1204
Vertebrae T5	1204
Vertebrae T6	1204
Vertebrae T7	1204
Vertebrae T8	1204
Vertebrae T9	1204
End of Table	

TABLE 2.4: Summary of the collected datasets and their contents.

Dataset	Organs	# scans	Year	References
CT-ORG (LiTS)	Lung, Bones, Liver, Bladder, Kidney, Brain	140	2019	Bilic et al., 2023
Abdomen CT-1K	Spleen, Kidney, Liver, Pancreas	1000	2020	Ma et al., 2021
WORD	Liver, Spleen, Kidney (L), Kidney (R), Stomach, Gallbladder, Esophagus, Pancreas, Duodenum, Colon, Intestine, Adrenal gland (L), Adrenal gland (R), Rectum, Head of Femur (L), Head of Femur (R)	150	2021	Luo et al., 2022
BTCV Abdomen	Liver, Spleen, Kidney (L), Kidney (R), Stomach, Gallbladder, Esophagus, Pancreas, Duodenum, Colon, Intestine, Adrenal gland (L), Adrenal gland (R), Rectum, Head of Femur (L), Head of Femur (R)	50	2015	*Multi-Atlas Labeling Beyond the Cranial Vault - Workshop and Challenge - syn3193805 - Wiki* n.d.
BTCV Cervix	Bladder, uterus, Rectum, small bowel, Kidney (L)	50	2015	
AMOS	Spleen, Kidney (R), Kidney (L), Gallbladder, Esophagus, Liver, Stomach, Aorta, Inferior vena cava, Pancreas, Adrenal gland (R), Adrenal gland (L), Duodenum, Bladder, Prostate, Uterus, Portal vein and splenic vein	500	2022	Ji et al., 2022
CHAOS	Liver	20	2021	2021
TCIA T&V	Brain, Lung, Mandible, Optic nerve, Orbit, Parotid gland, Spinal canal, Spinal cord, Submandibular gland, Cochlea, Lacrimal gland, Lens	31	2021	*TCIA Test & Validation Radiotherapy CT Planning Scan Dataset* 2022
MSD Liver	Liver, Liver tumor	131	2018	*The Medical Segmentation Decathlon | Nature Communications* n.d.
MSD Lung	Lung tumor	64	2018	
MSD Pancreas	Pancreas, Tumor mass	282	2018	

Continuation of Table 2.4

Dataset	Organs	# scans	Year	References
MSD Colon	Colon cancer primaries	226	2018	
MSD Hepatic Vessels	Hepatic tumor, Hepatic vessel	303	2018	
MSD Spleen	Spleen	41	2018	
Total- segmentor	adrenal gland left , adrenal gland right , aorta , autochthon left , au- tochthon right , brain , clavicula left , clavicula right , colon , duodenum , esophagus , face , femur left , femur right , gallbladder , gluteus maximus left , gluteus maximus right , gluteus medius left , gluteus medius right , gluteus minimus left , gluteus min- imus right , heart atrium left , heart atrium right , heart myocardium , heart ventricle left , heart ventricle right , hip left , hip right , humerus left , humerus right , iliac artery left , iliac artery right , iliac vena left , iliac vena right , iliopsoas left , iliopsoas right , inferior vena cava , kidney left , kidney right , liver , lung lower lobe left , lung lower lobe right , lung middle lobe right , lung upper lobe left , lung upper lobe right , pancreas , portal vein and splenic vein , pulmonary artery , rib left 1 , rib left 10 , rib left 11 , rib left 12 , rib left 2 , rib left 3 , rib left 4 , rib left 5 , rib left 6 , rib left 7 , rib left 8 , rib left 9 , rib right 1 , rib right 10 , rib right 11 , rib right 12 , rib right 2 , rib right 3 , rib right 4 , rib right 5 , rib right 6 , rib right 7 , rib right 8 , rib right 9 , sacrum , scapula left , scapula right , small bowel , spleen , stomach , trachea , urinary bladder , vertebrae C1 , vertebrae C2 , vertebrae C3 , vertebrae C4 , vertebrae C5 , vertebrae C6 , vertebrae C7 , vertebrae L1 , vertebrae L2 , vertebrae L3 , vertebrae L4 , vertebrae L5 , vertebrae T1 , vertebrae T10 , vertebrae T11 , ver- tebrae T12 , vertebrae T2 , vertebrae T3 , vertebrae T4 , vertebrae T5 , vertebrae T6 , vertebrae T7 , vertebrae T8 , vertebrae T9	1204	2022	(Wasserthal et al., 2022)
MedSeg LS	Liver segment 1, Liver segment 2, Liver segment 3, Liver segment 4, Liver segment 5, Liver segment 6, Liver segment 7, Liver segment 8	50	2021	*Database* n.d.
MedSeg IVC	Inferior vena cava	20	2021	*Database* n.d.

End of Table

Chapter 3

Architectures and Frameworks

In Chapter 3, we will focus on some of the popular architectures and frameworks for medical image segmentation. We will begin with a discussion of UNet and its variants, which have become widely used in the field due to their success in a variety of medical image segmentation tasks. We will then introduce the nnUNet framework, which is designed to be a plug-and-play tool for state-of-the-art biomedical segmentation. The nnUNet framework is capable of adapting to the properties of any given dataset, allowing researchers to easily apply it to their specific segmentation problem without the need for manual intervention. We will also discuss transformer-based variants of UNet, such as UNETR and Swin UNETR, which have gained popularity in recent years due to their ability to capture long-range dependencies in images. Finally, we will introduce the MONAI framework, which is a flexible and modular toolkit for deep learning in medical imaging. MONAI provides a range of predefined architectures and loss functions, as well as tools for data loading, augmentation, and evaluation, making it easy for researchers to quickly build and evaluate their own models

3.1 U-Net and Variants

3.1.1 The U-Net Architecture

U-Net is a deep learning architecture that was introduced by (Ronneberger, Fischer, and Brox, 2015) for medical image segmentation tasks. The UNet architecture is based on a fully convolutional network (FCN) (Long, Shelhamer, and Darrell, 2015) and consists of an encoder and a decoder section, connected by skip connections. The encoder section is responsible for extracting high-level features from the input image, while the decoder section is responsible for upsampling the feature maps and generating a pixel-level segmentation map.

The UNet architecture is particularly well-suited for image segmentation tasks due to its ability to capture both local and global context information. The skip connections between the encoder and decoder sections allow the network to retain information from the input image at different scales, which is essential for accurately segmenting objects with complex shapes. Additionally, the use of small convolutional kernels in the UNet architecture allows the network to capture fine-grained details in the input image, which is crucial for tasks such as medical image segmentation where accurate object boundaries are necessary.

The U-Net architecture is particularly effective at handling large amounts of data and is able to learn features that are relevant to the task of medical image segmentation. It has a relatively simple and straightforward structure, which makes it relatively easy to implement and modify. These characteristics have contributed to its widespread adoption in the field of medical image processing.

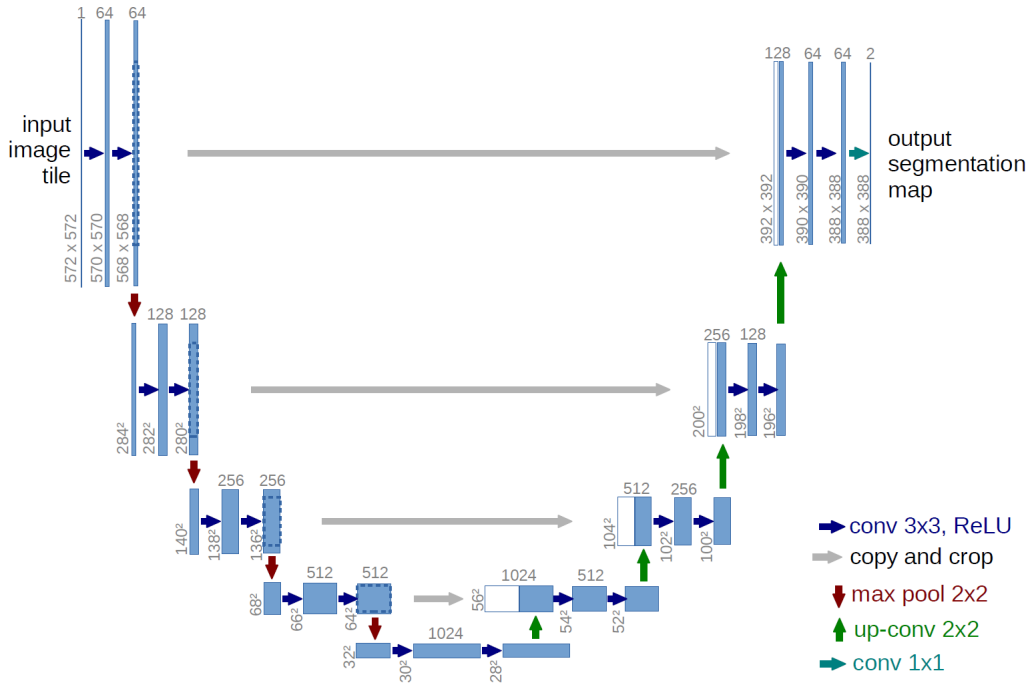


FIGURE 3.1: U-net architecture (example for 32x32 pixels in the lowest resolution). Each bluebox corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations. (Ronneberger, Fischer, and Brox, 2015)

3.1.2 U-Nets Variants

Since its introduction, the U-Net architecture has been widely adopted and modified for a variety of image segmentation tasks. Some common variants of the U-Net architecture include the 3D U-Net, Attention U-Net, CE-Net, U-Net++, U-Net 3+, nnUNet, and U2-Net. (Yin et al., 2022)

- The 3D U-Net (Çiçek et al., 2016) is a variant of the U-Net architecture that is specifically designed for 3D medical image segmentation tasks. The 3D U-Net replaces the 2D convolutional operations in the U-Net architecture with 3D convolutional operations, allowing it to capture spatial information in all three dimensions. This makes the 3D U-Net well-suited for tasks such as volumetric segmentation of CT and MRI scans.

- The Attention U-Net (Oktay et al., 2018)is a variant of the U-Net architecture that incorporates an attention module into the skip connections between the encoder and decoder sections. The attention module allows the network to selectively weight different features in the input image, improving the accuracy of the segmentation map.

- CE-Net (Gu et al., 2019) is a variant of the U-Net architecture that introduces a bottleneck between the encoder and decoder sections. The bottleneck consists of a depthwise separable convolution followed by a channel-wise attention module, which helps to reduce the number of parameters in the network and improve its generalization ability.

- UNet++ (Zhou et al., 2018b) is a variant of the UNet architecture that introduces dense blocks and in-depth supervision into the network. The dense blocks allow the network to capture more contextual information, while the in-depth supervision helps to improve the accuracy of the segmentation map.
- UNet 3+ (Huang et al., 2020) is a variant of the UNet architecture that introduces full-scale skip connections and deep supervision into the network. The full-scale skip connections allow the network to retain more information from the input image, while the deep supervision helps to improve the accuracy of the segmentation map.
- nnUNet (Isensee et al., 2021) is a variant of the UNet architecture that is based on a self-adaptive framework and consists of a pool of three simple UNet models.

3.2 nnUNet Framework

The nnU-Net framework, also known as no new-Net, was proposed by (Isensee et al., 2021) as a robust self-adaptive framework for medical image segmentation. It involves the use of a set of three U-Net models, including 2D and 3D U-Nets and a U-Net cascade. The 2D and 3D U-Nets can generate full-resolution results, while the cascade model is used to overcome the limitations of 3D U-Net on data sets with large image sizes.

nnU-Net is a framework for automated deep learning-based biomedical image segmentation that was designed to address this diversity in datasets within the field. It condenses and automates the key decisions for designing a successful segmentation pipeline for any given dataset, covering the entire pipeline from preprocessing to model configuration, model training, postprocessing, and ensembling. This makes nnU-Net a useful tool for both experienced researchers looking to quickly and effectively develop new segmentation methods and inexperienced users who can use nnU-Net out of the box for their custom 3D segmentation problem without the need for manual intervention. Fig. 3.2 shows the nnU-Net pipeline.

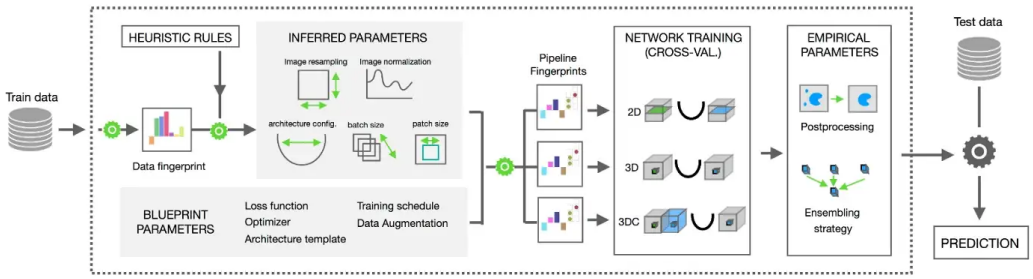
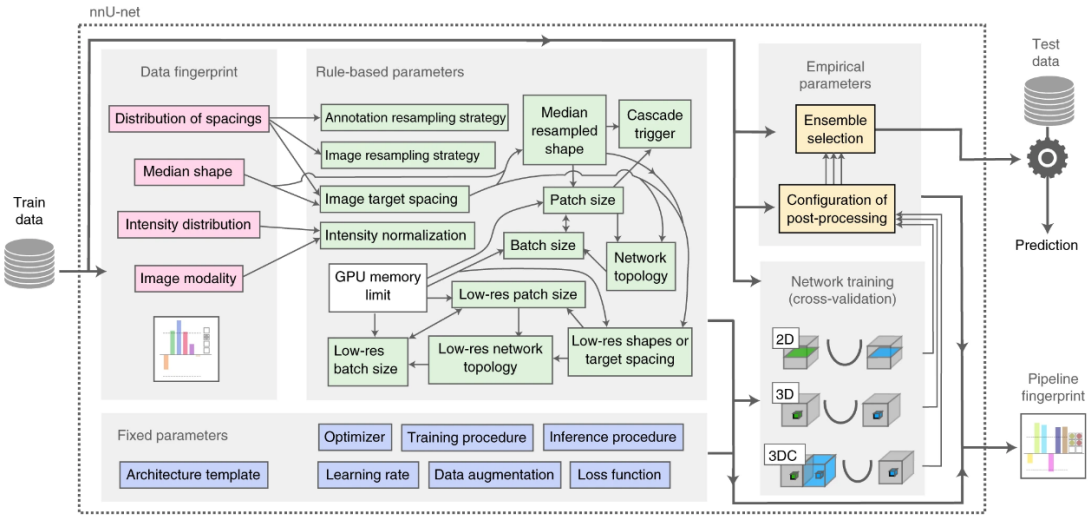


FIGURE 3.2: nnU-net automated workflow for deep learning-based biomedical image segmentation.(Isensee et al., 2021)

One of the key features of nnU-Net is its ability to extract a dataset fingerprint, which includes a set of dataset-specific properties such as image sizes, voxel spacings, and intensity information. This information is used to create three U-Net configurations: a 2D U-Net, a 3D U-Net that operates on full resolution images, and a

3D U-Net cascade where the first U-Net creates a coarse segmentation map in downsampled images which is then refined by the second U-Net. nnU-Net then trains all U-Net configurations in a 5-fold cross-validation to determine the best postprocessing and ensembling for the training dataset. The trained models can then be applied to test cases for inference.

Overall, nnU-Net is a valuable resource for the medical image processing community, providing a standardized benchmark for comparing algorithms and a framework for the rapid development of new segmentation methods. Fig. 3.3 shows how nnU-Net systematically addresses the configuration of entire segmentation pipelines and provides a visualization and description of the most relevant design choices.



Design choice	Required input	Automated (fixed, rule-based or empirical) configuration derived by distilling expert knowledge (more details in online methods)
Learning rate	–	Poly learning rate schedule (initial, 0.01)
Loss function	–	Dice and cross-entropy
Architecture template	–	Encoder–decoder with skip-connection ('U-Net-like') and instance normalization, leaky ReLU, deep supervision (topology-adapted in inferred parameters)
Optimizer	–	SGD with Nesterov momentum ($\mu = 0.99$)
Data augmentation	–	Rotations, scaling, Gaussian noise, Gaussian blur, brightness, contrast, simulation of low resolution, gamma correction and mirroring
Training procedure	–	1,000 epochs × 250 minibatches, foreground oversampling
Inference procedure	–	Sliding window with half-patch size overlap, Gaussian patch center weighting
Intensity normalization	Modality, intensity distribution	If CT, global dataset percentile clipping & z score with global foreground mean and s.d. Otherwise, z score with per image mean and s.d.
Image resampling strategy	Distribution of spacings	If anisotropic, in-plane with third-order spline, out-of-plane with nearest neighbor Otherwise, third-order spline
Annotation resampling strategy	Distribution of spacings	Convert to one-hot encoding → If anisotropic, in-plane with linear interpolation, out-of-plane with nearest neighbor Otherwise, linear interpolation

Image target spacing	Distribution of spacings	If anisotropic, lowest resolution axis tenth percentile, other axes median. Otherwise, median spacing for each axis. (computed based on spacings found in training cases)
Network topology, patch size, batch size	Median resampled shape, target spacing, GPU memory limit	Initialize the patch size to median image shape and iteratively reduce it while adapting the network topology accordingly until the network can be trained with a batch size of at least 2 given GPU memory constraints. for details see online methods.
Trigger of 3D U-Net cascade	Median resampled image size, patch size	Yes, if patch size of the 3D full resolution U-Net covers less than 12.5% of the median resampled image shape
Configuration of low-resolution 3D U-Net	Low-res target spacing or image shapes, GPU memory limit	Iteratively increase target spacing while reconfiguring patch size, network topology and batch size (as described above) until the configured patch size covers 25% of the median image shape. For details, see online methods.
Configuration of post-processing	Full set of training data and annotations	Treating all foreground classes as one; does all-but-largest-component-suppression increase cross-validation performance? Yes, apply; reiterate for individual classes No, do not apply; reiterate for individual foreground classes
Ensemble selection	Full set of training data and annotations	From 2D U-Net, 3D U-Net or 3D cascade, choose the best model (or combination of two) according to cross-validation performance

FIGURE 3.3: nnU-net automated method configuration for deep learning-based biomedical image segmentation.(Isensee et al., 2021)

3.3 Transformer Based Variants - UNETR and Swin UNETR

The swinUNETR porposed by (Hatamizadeh et al., 2022) combines Vision Transformers (ViT)s with self-supervised learning and convolutional layers for the accurate delineation of organs in CT scans.

ViT (Tang et al., 2022) is a novel architecture for computer vision tasks that has shown exceptional performance in self-supervised learning of global and local representations from unlabeled data. By using ViT in the encoder of the swinUNETR model, the authors aim to improve the representation learning ability of the model and achieve better performance on larger and more diverse datasets. They also use self-supervised learning to pre-train the swinUNETR model on publicly available CT images. Self-supervised learning (Hwang and Kim, 2016) is a learning paradigm that uses various pretext tasks, such as masked volume inpainting and rotation, to learn a contextual representation of the training data without the need for manual annotations.

Unlike traditional convolutional neural networks (CNNs), which rely on hand-designed filters and local receptive fields, ViT uses self-attention mechanisms (Vaswani et al., 2017) to compute global interactions between image patches. This enables ViT to capture long-range dependencies and contextual information in the input data, leading to improved performance on various tasks.

The second key component of swinUNETR is the use of self-supervised learning (Tang et al., 2022). In self-supervised learning, the model is trained on a variety of pretext tasks, such as masked volume inpainting and rotation, to learn a contextual representation of the training data without the need for manual annotations. This reduces the burden of data annotation and enables the model to learn from large-scale, unlabeled datasets. In the case of swinUNETR, the self-supervised learning is applied to the 3D Swin Transformer encoder to learn a representation of the anatomy of the organs in the CT scans.

In addition to ViT and self-supervised learning, swinUNETR also uses convolutional layers to capture multiscale feature representations for dense prediction tasks, such as medical image segmentation. The combination of these components enables swinUNETR to achieve high performance and adaptability to diverse clinical scenarios.

To adapt ViT to the task of medical image segmentation, the authors incorporate convolutional layers into the architecture. These layers capture multiscale feature representations and enable dense prediction tasks, such as organ segmentation. In particular, the convolutional layers in swinUNETR are connected to a residual UNet-like decoder at five different resolutions by skip connections. This allows swinUNETR to capture both global and local information in the input data and make more accurate predictions.

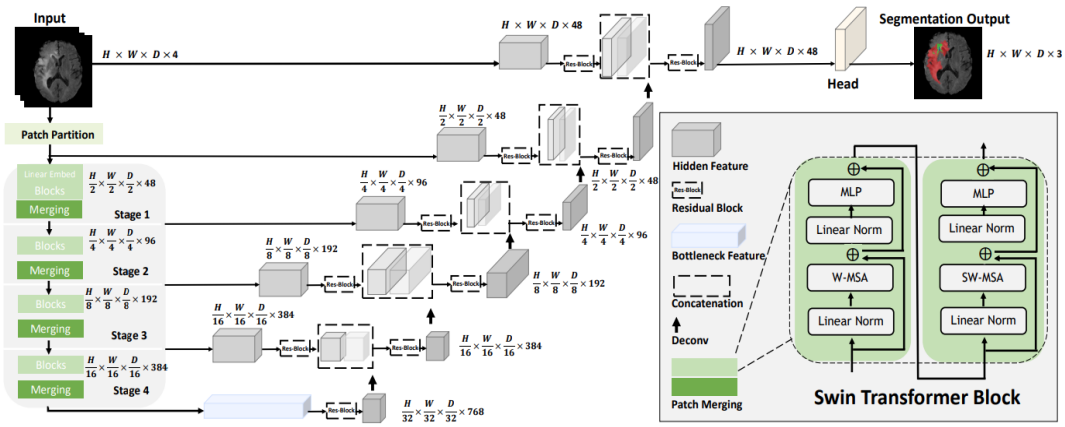


FIGURE 3.4: Overview of the Swin UNETR architecture. The input to the model is 3D multi-modal MRI images with 4 channels. The Swin UNETR creates non-overlapping patches of the input data and uses a patch partition layer to create windows with a desired size for computing the self-attention. The encoded feature representations in the Swin transformer are fed to a CNN-decoder via skip connection at multiple resolutions. Final segmentation output consists of 3 output channels. (Hatamizadeh et al., 2022)

3.4 MONAI Framework

The MONAI (Cardoso et al., 2022) framework offers a range of tools and techniques for performing segmentation tasks in medical imaging. One key feature is its support for extended dataset functionality, including integrated caching and persistence solutions that can reduce the computational overhead of data pre-processing.MONAI also provides a number of specialized transforms and networks specifically designed for medical image segmentation. Its support for flexible pre-processing of multi-dimensional medical imaging data includes techniques such as transform functions for resampling and rescaling, which can be useful for preparing data for use with segmentation algorithms. This is particularly important for medical datasets, which are often large and memory-intensive, and may require significant pre-processing in order to be used effectively for training and evaluation. In addition, MONAI provides compositional and portable APIs that make it easy to integrate its tools and techniques into existing workflows.

MONAI also includes domain-specific implementations of networks, losses, evaluation metrics, and other components that are commonly used in medical imaging segmentation tasks. This includes support for popular networks such as U-Net, as well as specialized loss functions such as Dice loss and Jaccard loss. These components are designed to be highly customizable, allowing users to easily tailor them to their specific needs.

Another key advantage of MONAI is its support for multi-GPU data parallelism. This can greatly accelerate the training and evaluation of segmentation models, particularly when working with large volumetric medical images. MONAI also provides detailed guidance and best practices for achieving the best performance when using its tools and techniques for medical imaging segmentation.

In addition to these specialized tools, MONAI also offers a range of utility functions that can be used to visualize and analyze the results of segmentation tasks. For example, the blend image function allows users to create RGB images by superimposing images and labels, making it easier to visualize the segmentation results. Similarly, the matshow3d function creates a 3D volume figure as a grid of images, making it easier to view and analyze the results of segmentation tasks in three dimensions.

The MONAI framework provides a comprehensive set of tools and features for performing medical image segmentation tasks. Its support for extended dataset functionality, specialized transforms and networks, and visualization tools make it a powerful and effective tool for medical image segmentation tasks.

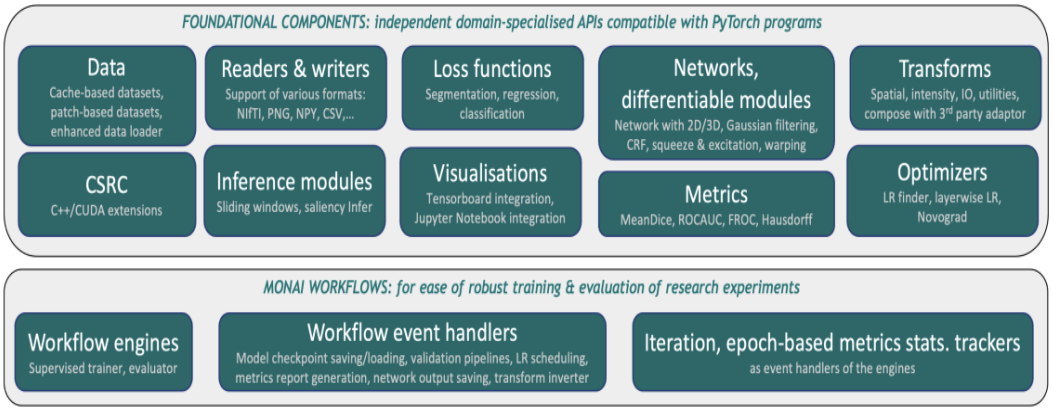


FIGURE 3.5: Pictorial representation of the MONAI Core modules (top) and MONAI workflows components (bottom). (Cardoso et al., 2022)

Chapter 4

Key Contributors in Medical Image Segmentation: Industry and Research Groups

4.1 DKFZ Laboratories

DKFZ Laboratories is a research institute located in Heidelberg, Germany that focuses on cancer research. The institute was founded in 1964 and has since become a leader in the field of medical imaging, particularly in the area of medical image segmentation. DKFZ Laboratories has a strong track record of developing innovative technologies and techniques for the analysis and interpretation of medical images, including techniques for segmentation, visualization, and analysis of various types of medical imagery such as magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound. The institute has a reputation for producing high-quality research and has a strong focus on collaboration and partnerships with industry, academia, and other research organizations. In recent years, DKFZ Laboratories has also become a key contributor to the development of open-source tools and frameworks for medical image analysis, including the MONAI framework for medical image segmentation.

DKFZ Laboratories has made significant contributions to the field of medical image segmentation through the development of the nnU-Net framework. Created by Dr. Fabian Isensee and his team at DKFZ in 2019, nnU-Net has quickly become the state-of-the-art method for medical image segmentation and has won multiple international segmentation competitions. The success of nnU-Net has solidified DKFZ Laboratories as a key contributor in the field of medical image segmentation and has further established the institute as a leader in the development of innovative technologies for the analysis and interpretation of medical images.

4.2 DeepMind

DeepMind is a research organization based in London, United Kingdom that focuses on developing artificial intelligence (AI) technologies and applications. Founded in 2010, DeepMind has become well-known for its work on machine learning and AI, including the development of deep learning algorithms and neural network architectures. DeepMind has also made significant contributions to the field of medical image analysis, particularly in the area of medical image segmentation. One of DeepMind's most notable achievements in this area is the AlphaFold system, which uses machine learning algorithms to predict the 3D structure of proteins from their

amino acid sequences. DeepMind's research has been widely published in top scientific journals and the company has received numerous awards and accolades for its work. In addition to its research activities, DeepMind also has a number of commercial products and partnerships with industry and other organizations.

One of the main contributions of DeepMind is the development of a 3D U-Net model architecture based on the nnUnet framework that is able to accurately segment the relevant anatomy with high inter-rater reliability. The team also proposed a new metric, called the "surface DSC at tolerance τ," which measures the overlap between the predicted and ground truth surfaces at a specified tolerance, rather than the overlap between the full volumes as in the traditional Dice similarity coefficient (DSC) metric. This new metric is better suited to the task of segmenting sparsely labelled anatomy in CT scans, as it allows for more flexibility in the annotation process and accounts for variations in the thickness of the anatomic structures.

The DeepMind team also conducted experiments on a large dataset of head and neck CT scans, showing that their model achieved high performance in terms of both the traditional DSC metric and the proposed surface DSC metric. Additionally, they compared the performance of their model to that of radiologists with different levels of experience, demonstrating that the model was able to perform comparably to radiologists with 4 years of experience.

4.3 NVIDIA

NVIDIA is a technology company that designs and manufactures computer graphics processing units (GPUs) and other specialized chips for the gaming, professional visualization, data center, and automotive markets. NVIDIA's GPUs are used in a variety of applications, including gaming, professional visualization, data analytics, and machine learning. The company is known for its powerful graphics processing capabilities and has become a leader in the field of artificial intelligence (AI) and machine learning, with its GPUs being used in a variety of AI applications, including natural language processing, image and video recognition, and autonomous vehicles. NVIDIA also develops software tools and platforms to help developers create and deploy AI applications, and it partners with leading companies in various industries to drive the development of AI technologies.

Its researchers have developed innovative technologies such as Swin UNETR, a transformer-based model for 3D medical image analysis that uses the MONAI framework and has set new state-of-the-art benchmarks on various medical image segmentation tasks. NVIDIA has also made contributions to the field through its partnerships with leading healthcare organizations, such as the National Health Service in the UK, to deploy its MONAI-based AI Deployment Engine platform for disease detection. In addition, NVIDIA has supported the development of medical imaging start-ups through its Inception program, which nurtures over 1,800 healthcare start-ups working on GPU-based tools for optimizing operations, improving diagnostics, and enhancing patient care.

4.4 Medical Open Network for Artificial Intelligence (MONAI)

MONAI (Medical Open Network for Artificial Intelligence) (Cardoso et al., 2022) is an open-source framework for healthcare that builds on the best practices from existing tools and is user-friendly, reproducible, and optimized for the demands of healthcare data. It is designed to handle the unique formats, resolutions, and specialized meta-information of medical images, and provides domain-specific data transforms, neural network architectures, and evaluation methods for medical imaging.

MONAI was originally started by NVIDIA and King's College London in collaboration with the Chinese Academy of Sciences, the German Cancer Research Center, Kitware, MGH & BWH Center for Clinical Data Science, Stanford University, and the Technical University of Munich. It is now a collaborative community of AI researchers from academia and industry, with an advisory board and nine working groups led by thought leaders in the medical research field.

Some key achievements of MONAI include the development of the MONAI Label intelligent image labeling tool, the MONAI Core library with its state-of-the-art transformer-based 3D segmentation algorithms, and the MONAI Deploy initiative for developing, testing, deploying, and running medical AI applications in clinical production. MONAI has also contributed to the development of the international biomedical image analysis challenges (BIAS) initiative, and has established itself as a catalyst for scientific progress and real-life impact through research partnerships, publications, and industry collaboration. MONAI has attracted a number of notable researchers, including Olaf Ronnenberg and Fabian Isensee. These researchers have made significant contributions to the field of medical image processing, particularly in the areas of image segmentation and deep learning.

4.5 RapidAI

RapidAI is a healthcare technology company that develops and provides software tools for the diagnosis and treatment of cerebrovascular disorders, such as stroke. The company's software platform, called Rapid, is designed to analyze brain imaging data and provide information to healthcare professionals that can aid in the identification and diagnosis of cerebrovascular disorders. RapidAI technology has been validated in multiple clinical trials and has received FDA clearance for use in hospitals in the United States. In addition to its focus on stroke, the company is also expanding its platform to other vascular conditions. RapidAI was founded by neurologists Dr. Greg Albers and Dr. Roland Bammer, and is headquartered in Silicon Valley, California.

Chapter 5

Evaluation and Experimentation of the State-of-the-Art

The aim of our experiments is to answer two questions: (1) Does the newer architecture, such as Swin-UNet, perform better than the established state-of-the-art nnU-Net, and (2) can we approximate the number of training images to obtain the best or satisfactory results?

We conducted two experiments to answer each of these questions, respectively. The first experiment simply involved replicating results in-house with a dataset where the two models have not been compared. Additionally, we will provide published benchmark results to support or contradict our claim. The second experiment involved using a large dataset with a large number of classes to train a state-of-the-art model with gradually less training data in separate trainings to obtain a plot of dice score versus the number of samples. More details on the experiments will be given in their respective methodology sections.

5.1 Metrics

The Sørensen–Dice coefficient

The Sørensen–Dice coefficient, also known as the Sørensen index or Dice's coefficient, is a measure of the similarity between two samples or sets. It is defined as twice the size of the intersection of the two sets divided by the sum of the sizes of the two sets. The coefficient ranges from 0 to 1, with a value of 1 indicating that the two sets are identical. The Sørensen–Dice coefficient is commonly used in information retrieval and data analysis, and has applications in fields such as biology and image processing. It is similar to the Jaccard index, but differs in that it counts all elements in the intersection rather than just true positives.

$$Dice = \frac{2 * TP}{(TP + FP) + (TP + FN)} \tag{5.1}$$

5.2 Comparing the Performance of swin-UNETR and nnU-Net

5.2.1 Methodology

In this section, we describe the methods used to evaluate the swinUNETR and nnUNet models for organ segmentation in CT scans.

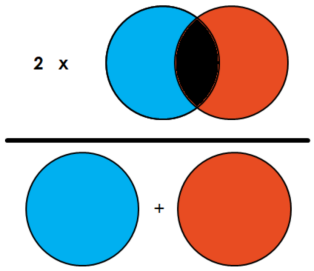


FIGURE 5.1: Illustration of the Dice Coefficient.

To evaluate the performance of swinUNETR and nnUNet, we use two datasets: the WORD dataset and the AMOS dataset. We compare the results of swinUNETR with those of the current state-of-the-art model, nnUnet, using quantitative metrics such as the average Dice score.

To confirm the published results on these datasets, we have also replicated the experiments in-house. Our results on the WORD are consistent with the published results, indicating that nnUNet performs significantly better than swinUNETR in terms of average Dice score.

5.3 Results

5.3.1 Results

Our results (Table 5.2, Table 5.2 and Table 5.3) show that nnUNet outperforms the newest method swinUNTER on the AMOS and WORD datasets, achieving state-of-the-art performance on a range of organ segmentation tasks. The combination of ViT and self-supervised learning in the swinUNETR model may allow it to learn robust representations of the data but did not lead to improved performance on the available datasets.

5.3.2 Discussion

The results of this study demonstrate the need for improvement of new methods such as swinUNETR for organ segmentation in CT scans. The hybrid architecture of the model, combining ViT and CNNs, allows it to capture both global and local representations of the data but did not lead to improved performance on a range of organ segmentation tasks. The use of self-supervised learning for pre-training also allows the model to learn from large amounts of unlabeled data, reducing the need for labor-intensive manual annotations. Overall, our results show that swinUNETR needs wheter improvements or to be integrated in self configuring network alike nnUNet.

5.3.3 Limitations

One limitation of this study is that we only evaluated swinUNETR on the AMOS and WORD dataset. Future work could investigate the performance of the model on other datasets, as well as its ability to generalize to different imaging modalities and clinical scenarios. Additionally, the pre-training of the model was performed on

publicly available CT images, which may not fully capture the diversity and complexity of real-world medical data. Further research is needed to evaluate the performance of swinUNETR on more realistic data.

More thorough analysis of the factors that contribute to the improved performance, such as the use of self-supervised learning and the incorporation of Vision Transformers and the framework encapsulating the model is needed.

Model	**Mean DSC**	**Spleen**	**Right kidney**	**Left kidney**	**Gallbladder**	**Esophagus**	**Liver**	**Stomach**	**Aorta**	**Postcava**	**Pancreas**	**Right adrenal gland**	**Left adrenal gland**	**Duodenum**	**Bladder**	**Prostate /Uterus**
nnUNet	90.0%	97.1%	96.4%	96.2%	83.2%	87.5%	97.6%	92.2%	96.0%	92.5%	88.6%	81.2%	81.7%	85.0%	90.5%	85.0%
Swin-UNTER	86.4%	95.5%	93.8%	94.5%	77.3%	83.0%	95.9%	88.9%	94.7%	89.6%	84.9%	77.2%	78.3%	78.6%	85.8%	77.4%

TABLE 5.1: Official results of Swin-UNTER and nnUNet on the AMOS benchmark Ji et al., 2022

Method	Liver	Spleen	Kidney (L)	Kidney (R)	Stomach	Gallbladder	Esophagus	Pancreas	Duodenum	Colon	Intestine	Adrenal	Rectum	Bladder	Head of Femur(L)	Head of Femur(R)	Mean of all organs
nnUNet(2D)	95.38	93.33	90.05	89.86	89.97	78.43	78.08	82.33	63.47	83.06	85.6	69.9	81.66	90.49	93.28	93.78	84.92
nnUNetV2(2D)	96.19	94.33	91.29	91.20	91.12	83.19	77.79	83.55	64.47	83.92	86.83	70.0	81.49	90.15	93.28	93.93	85.80
ResUNet(2D)	96.55	95.26	95.63	95.84	91.58	82.83	77.17	83.56	66.67	83.57	86.76	70.9	82.16	91.0	93.39	93.88	86.67
DeepLabV3+(2D)	96.21	94.68	92.01	91.84	91.16	80.05	74.88	82.39	62.81	82.72	85.96	66.82	81.85	90.86	92.01	92.29	84.91
UNet++(2D)	96.33	94.64	93.36	93.34	91.33	81.21	78.36	84.43	65.99	83.22	86.37	71.04	81.44	92.09	93.38	93.88	86.28
AttUNet(3D)	96.00	94.90	94.65	94.7	91.15	81.38	76.87	83.55	67.68	85.72	88.19	70.23	80.47	89.71	91.90	92.43	86.21
nnUNet(3D)	96.45	95.98	95.40	95.68	91.69	83.19	78.51	85.04	68.31	87.41	89.30	72.38	82.41	92.59	91.99	92.74	87.44
nnUNetV2(3D)	96.59	96.09	95.63	95.83	91.57	83.72	77.36	85.00	67.73	87.26	89.24	72.22	82.24	92.38	91.99	92.74	87.43
UNETR(3D)	94.67	92.85	91.49	91.72	85.56	65.08	67.71	74.79	57.56	74.62	80.4	60.76	74.06	85.42	89.47	90.17	79.77
CoTr(3D)	95.58	94.9	93.26	93.63	89.99	76.4	74.37	81.02	63.58	84.14	86.39	69.06	80.0	89.27	91.03	91.87	84.66

TABLE 5.2: Results from the WORD Luo et al., 2022 paper

Method	Liver	Spleen	Kidney (L)	Kidney (R)	Stomach	Gallbladder	Esophagus	Pancreas	Duodenum	Colon	Intestine	Adrenal	Rectum	Bladder	Head of Femur(L)	Head of Femur(R)	Mean of all organs
Swin UNETR	96.12	94.75	95.32	94.98	91.24	65.43	74.87	81.02	64.35	83.31	85.54	68.64	76.92	91.44	90.37	89.48	84.41
nnUNetV2(3D)	96.36	96.11	95.23	95.14	93.46	83.18	82.37	84.65	66.13	82.9	83.6	73.99	80.89	90.66	94.42	93.57	87.04

TABLE 5.3: Replicated results of Swin-UNTER and nnUNet on the WORD benchmark

5.4 Investigating the Effect of Training Data Quantity on the Dice Score of nnU-Net

In order to build a deep learning framework for segmenting a large number of anatomical structures, the first step is to acquire a sufficient amount of data. Using the Totalsegmentator dataset, which contains 1204 images and 104 classes, we can measure the performance of nnU-Net on a large number of classes depending on the number of training images.

We aim to investigate the relationship between the number of training samples and the performance of the nnU-Net (3D) model for organ segmentation. We can train the nnU-Net model multiple times, each time with a different number of training samples. We can then plot the resulting performance, measured in terms of Dice score, against the number of training samples. This allow us to see how the performance of the model changes as the number of training samples increased. By analyzing this graph, we can gain a better understanding of the number of training samples needed to achieve a certain level of accuracy on different types of anatomic structures. This information can be useful for designing and optimizing a data collection campaign.

5.4.1 Methodology

In this section, we describe the methods used to evaluate the performance of the nnU-Net model for organ segmentation in CT scans as a function of the quantity of training data.

We use the Totalsegmentator dataset, which contains 1204 images and 104 classes, to measure the performance of the nnU-Net (3D) model for organ segmentation. Our method consisted of replicating the results published by Totalsegmentator to confirm the validity of our approach. The dataset of 1204 subjects was randomly split by the Totalsegmentator team into 1082 training subjects (89.9%), 57 validation subjects (4.7%) and 65 subjects for final testing (5.4%). We use the same split with the same images in each set. Next, we performed pre-training on another dataset, the WORD dataset, with 100 images and 15 classes. Then, we launched 5 trainings with the Totalsegmentator training dataset, each with a different number of training images (10, 50, 100, 500, 1082) and measured the average Dice score on the test set. With 5 data points for each class, we plotted a learning curve of the performance as a function of the number of training images. The first training is done with 1082 images, then the second one with 500 images from the 1082 images of the first, the third training with 100 images from the 500 images of second training, etc... Each of these trainings was performed on 4000 nnU-Net-epochs. "Epochs" are defined by (Isensee et al., 2021) differently, we mention them as "nnU-Net-epochs". The definition is provided in appendix B.

5.4.2 Results

The results of the replication of the Totalsegmentator paper results (Wasserthal, 2022) is displayed in table 5.5 along with the published results and percentage change between our Dice scores and the paper Dice scores. The learning curve for each class is provided in figure 5.2 in logarithmic scale and in figure 5.3 in linear scale.

Our results show that the performance of the nnU-Net model, as measured by the average Dice score, improves as the number of training images increases. By analyzing the learning curve of the performance as a function of the number of training images, we can determine the optimal number of training images needed to achieve a certain level of accuracy for different types of anatomic structures.

Additionally, training plots are provided in appendix C.

5.4.3 Discussion

This study provides valuable insights into the relationship between the number of training samples and the performance of the nnU-Net model for organ segmentation. The results suggest that a sufficient number of training samples is crucial for achieving high accuracy in deep learning frameworks such as nnU-Net for medical image segmentation. For large anatomical structures a training set of 100 images can be tested. For other structures a minimum of 500 images is advised.

The replication results for most classes corresponds to the paper results. A percentage change of less than 5% has been measured for most classes. Classes that have a higher percentage change should be dismissed

5.4.4 Limitations

One limitation of this study is that it only evaluated the nnU-Net model on the Totalsegmentator and WORD datasets. Further research is needed to evaluate the performance of the model on other datasets, as well as its ability to generalize to different imaging modalities and clinical scenarios. Other limitations, follow a quality loss on certain images uploaded by Wasserthal, 2022

5.5 Conclusion

In conclusion, the proposed swinUNETR and nnUNet models shows promising results for the task of abdominal organ segmentation in CT scans. By combining self-supervised learning and Vision Transformers, the model achieved close to state-of-the-art results on the WORD and AMOS datasets. However, the limitations of the WORD and AMOS dataset and the potential for further improvements to the proposed model are also acknowledged. Future work should aim to evaluate the generalizability of the approach on a more diverse range of images and organs, as well as explore additional methods for improving performance such as implementing swinUNETR within the nnU-Net framework.

Our experiment on the Totalsegmentator dataset allowed us build a better intuition of the number of training data required to obtain good results.

5.6 Additional results

The table 5.4 shows the performance of the nnUnet model on the Medical Segmentation Challenge (MSD) Benchmark. The table includes information about the task, modality, protocol, target, number of cases and training samples, and performance measures including Dice and NSD. The nnUnet model demonstrated good performance on tasks such as left atrium segmentation, anterior and posterior hippocampus segmentation, and spleen segmentation, with Dice scores above 0.9 and NSD scores above 0.96. However, the model demonstrated lower performance on tasks such as enhancing brain and pancreatic tumor mass segmentation, with Dice scores below 0.5 and NSD scores below 0.73.

Although the results of the nnUnet model on the Medical Segmentation Challenge (MSD) Benchmark demonstrate a wide range of performance across different tasks and target structures, it is not possible to extract a general rule about the relationship between the number of training samples and the performance of the model as measured by Dice or NSD scores. This is likely because the performance of the model depends on a variety of factors, including the complexity of the task, the quality and diversity of the training data, and the specific design of the model. Therefore, it is important to carefully evaluate the performance of the model on a case-by-case basis, considering the specific characteristics of the task and the available training data.

TABLE 5.4: Performance of nnUnet on Medical Segmentation Challenge (MSD) Benchmark.

Task	Modality	Protocol	Target	# Cases (Train/Test)	# Training	Dice	NSD
Brain	mp-MRI	FLAIR, T1w, T1 \w Gd, T2w	Edema	750 4D volumes (484/266)	484	0.68	0.87
Brain	mp-MRI	FLAIR, T1w, T1 \w Gd, T2w	Enhancing	750 4D volumes (484/266)	484	0.48	0.73
Brain	mp-MRI	FLAIR, T1w, T1 \w Gd, T2w	Non-enhancing tumor	750 4D volumes (484/266)	484	0.68	0.91
Heart	MRI	–	Left atrium	30 3D volumes (20/10)	20	0.93	0.96
Hippo-	MRI	T1w	Anterior	394 3D volumes (263/131)	263	0.90	0.98
cam-							
pus							
Hippo-	MRI	T1w	Posterior	394 3D volumes (263/131)	263	0.89	0.98
cam-							
pus							
Liver	CT	Portal-venous phase	Liver	210 3D volumes (131/70)	131	0.95	0.98
Liver	CT	Portal-venous phase	Liver tumor	210 3D volumes (131/70)	131	0.74	0.88
Lung	CT	–	Lung and Lung cancer	96 3D volumes (64/32)	64	0.69	0.69
Pancreas	MRI	Portal-venous phase	Pancreas	420 3D volumes (282/139)	282	0.80	0.95
Pancreas	MRI	Portal-venous phase	Pancreatic tumor mass	420 3D volumes (282/139)	282	0.52	0.73
Prostate	mp-MRI	T2, ADC	Prostate PZ	48 4D volumes (32/16)	32	0.76	0.96
Prostate	mp-MRI	T2, ADC	Prostate TZ	48 4D volumes (32/16)	32	0.90	0.99
Hepatic	CT	Portal-venous phase	Hepatic vessels	443 3D volumes (303/140)	303	0.63	0.83
Ves-							
sels							
Hepatic	CT	Portal-venous phase	Hepatic vessels	444 3D volumes (303/140)	303	0.69	0.79
Ves-							
sels							
Colon	CT	Portal-venous phase	Colon cancer primaries	190 3D volumes (126/64)	126	0.56	0.68
Spleen	CT	Portal-venous phase	Spleen	61 3D volumes (41/20)	41	0.96	0.99

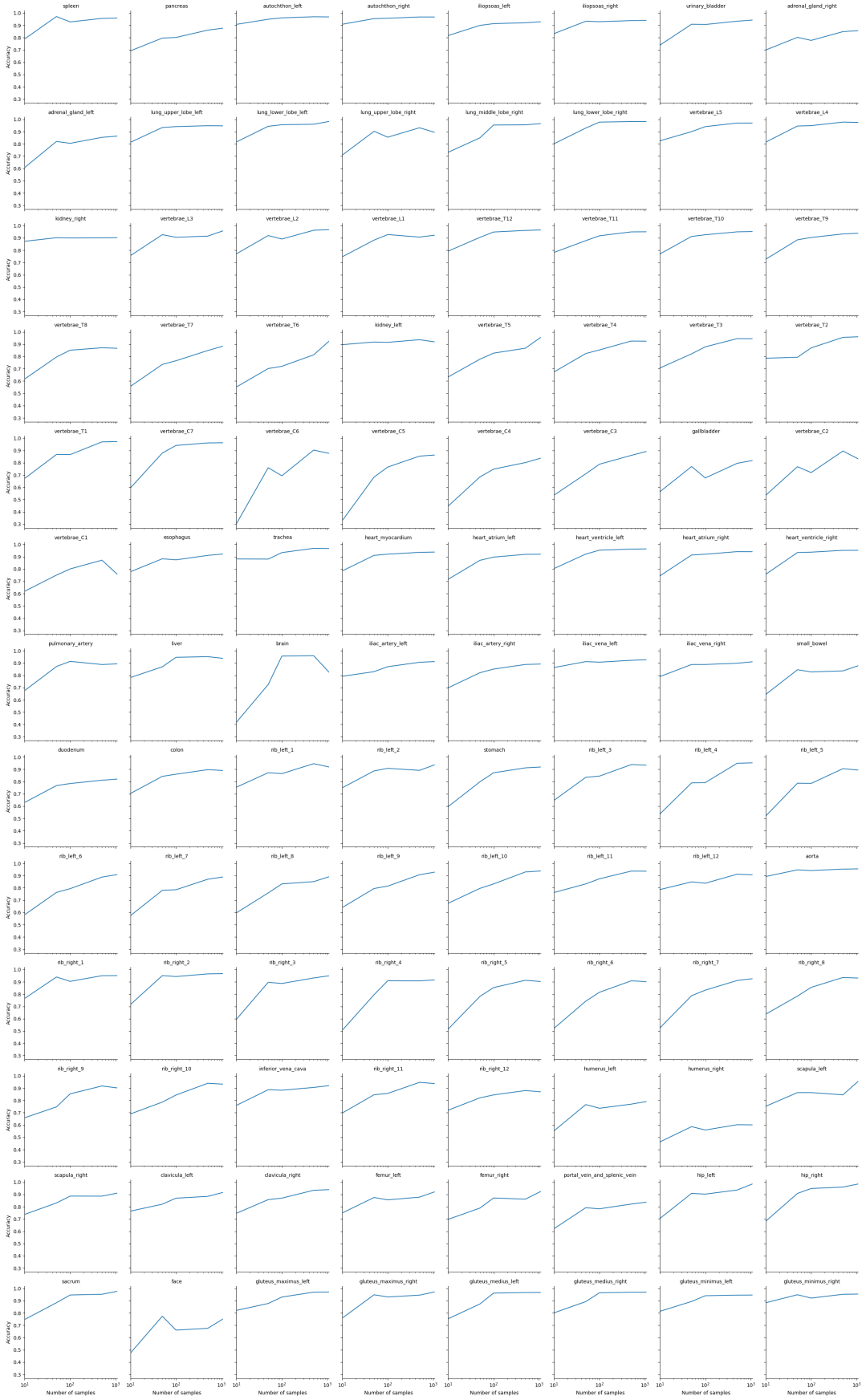


FIGURE 5.2: Logarithmic Plot of the Relationship between Number of Training Images and Dice Score for each Anatomical Structure

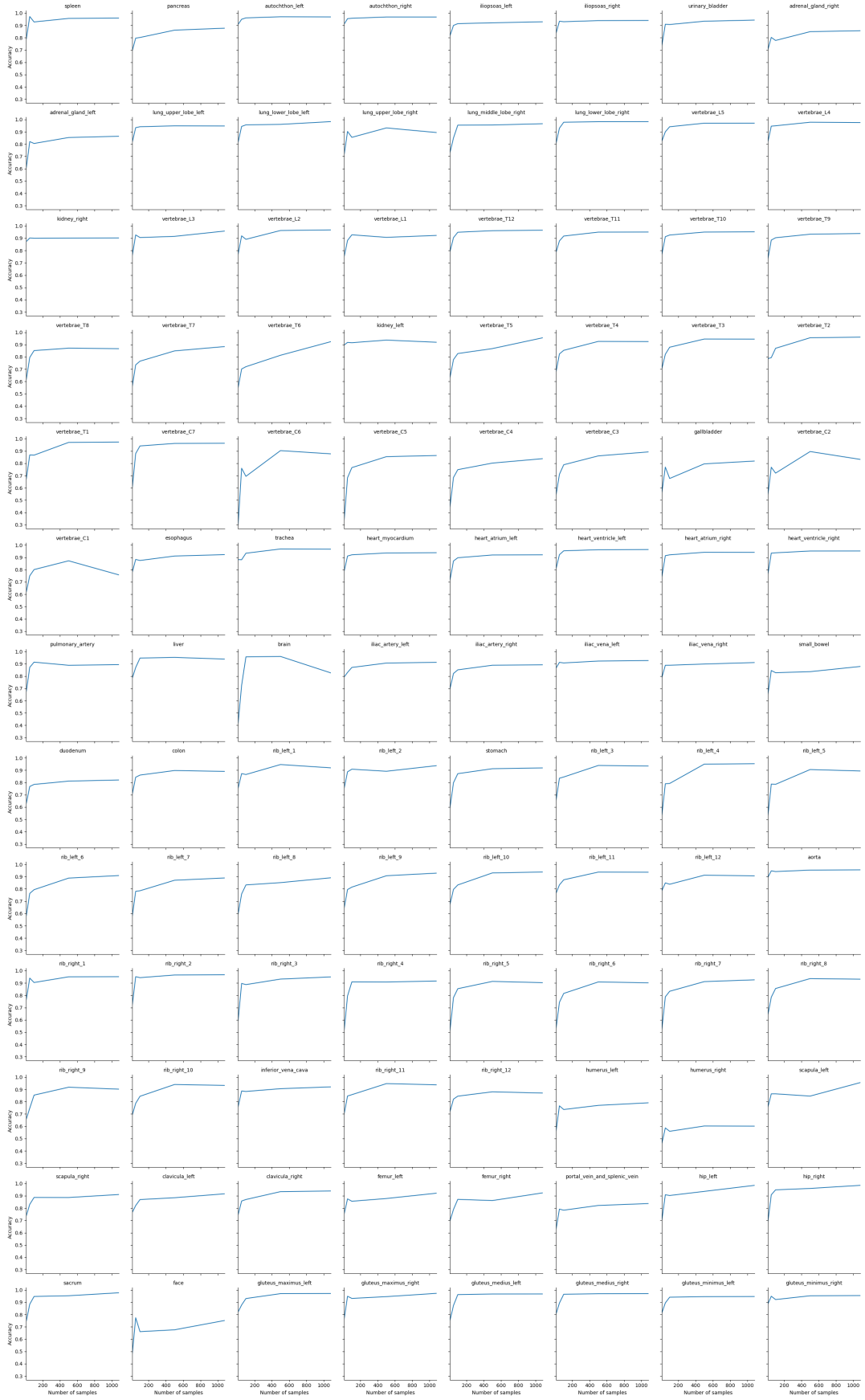


FIGURE 5.3: Linear Plot of the Relationship between Number of Training Images and Dice Score for each Anatomical Structure

TABLE 5.5: Comparison of Replicated Results with Published Results for the Totalsegmentator Study on Organ Segmentation in CT Scans

Organ	Dice (wasserthal)	Dice (ours)	% change
spleen	98.3	95.9	-2.4
kidney(r)	93.9	90.2	-3.9
kidney(l)	95.3	91.9	-3.6
gallbladder	87.5	81.8	-6.5
liver	96.5	93.9	-2.7
stomach	94.7	91.9	-3.0
aorta	98.1	95.5	-2.7
inferior vena cava	93.5	92.0	-1.6
portal vein and splenic vein	88.1	83.7	-5.0
pancreas	88.7	87.7	-1.1
adrenal gland(r)	90.9	85.7	-5.7
adrenal gland(l)	89.8	86.4	-3.8
lung upper lobe(l)	97.1	94.7	-2.4
lung lower lobe(l)	98.5	98.3	-0.2
lung upper lobe(r)	94.4	89.4	-5.3
lung middle lobe(r)	97.5	96.6	-1.0
lung lower lobe(r)	98.8	98.3	-0.6
vertebrae L5	96.1	97.0	1.0
vertebrae L4	96.6	97.5	1.0
vertebrae L3	95.9	95.8	-0.1
vertebrae L2	96.0	96.7	0.8
vertebrae L1	94.8	92.3	-2.7
vertebrae T12	94.4	96.5	2.2
vertebrae T11	95.3	95.0	-0.3
vertebrae T10	95.0	95.2	0.2
vertebrae T9	95.9	93.8	-2.2
vertebrae T8	92.9	86.7	-6.7
vertebrae T7	93.5	88.4	-5.4
vertebrae T6	96.1	92.4	-3.9
vertebrae T5	96.5	95.5	-1.0
vertebrae T4	94.5	92.5	-2.2
vertebrae T3	95.8	94.5	-1.4
vertebrae T2	95.8	96.1	0.3
vertebrae T1	98.4	97.3	-1.1
vertebrae C7	97.5	96.3	-1.2
vertebrae C6	87.1	87.7	0.7
vertebrae C5	88.3	86.3	-2.2
vertebrae C4	89.0	83.8	-5.9
vertebrae C3	96.8	89.3	-7.8
vertebrae C2	98.4	83.2	-15.5
vertebrae C1	89.4	75.8	-15.2
heart esophagus	94.4	92.3	-2.3
heart trachea	97.7	96.7	-1.0
heart myocardium	93.7	93.8	0.1
heart atrium (l)	94.1	92.1	-2.1
heart ventricle (l)	95.5	96.4	0.9
heart atrium (r)	93.9	94.1	0.2
heart ventricle (r)	94.9	95.2	0.4
heart pulmunary artery	92.7	89.3	-3.6
brain	96.9	82.6	-14.7
iliac artery (l)	92.8	91.2	-1.7
iliac artery (r)	91.0	89.2	-2.0
iliac vena (l)	92.9	92.6	-0.3
iliac vena (r)	91.6	91.0	-0.6
small bowel	89.9	87.8	-2.3
duodenum	83.7	82.0	-2.0
colon	89.6	89.0	-0.6
rib (l) 1	97.0	91.9	-5.2
rib (l) 2	95.9	93.7	-2.3
rib (l) 3	96.0	93.4	-2.7
rib (l) 4	96.8	95.3	-1.6
rib (l) 5	92.5	89.4	-3.4
rib (l) 6	94.3	90.9	-3.6
rib (l) 7	93.9	88.9	-5.3
rib (l) 8	92.0	89.0	-3.2
rib (l) 9	93.2	92.8	-0.4
rib (l) 10	91.5	93.7	2.5
rib (l) 11	92.6	93.6	1.0
rib (l) 12	90.4	90.6	0.3
rib (r) 1	96.9	95.1	-1.8
rib (r) 2	97.9	96.7	-1.2
rib (r) 3	96.9	94.9	-2.0

Organ	Dice (wasserthal)	Dice (ours)	% change
rib (r) 4	97.3	91.6	-5.9
rib (r) 5	92.8	90.3	-2.7
rib (r) 6	94.5	90.2	-4.6
rib (r) 7	95.0	92.6	-2.6
rib (r) 8	95.6	93.1	-2.6
rib (r) 9	96.0	90.2	-6.0
rib (r) 10	96.0	93.3	-2.9
rib (r) 11	95.5	93.7	-1.9
rib (r) 12	93.7	87.0	-7.1
humerus (l)	89.9	79.1	-12.0
humerus (r)	82.6	60.1	-27.2
scapula (l)	98.0	95.6	-2.5
scapula (r)	95.4	91.1	-4.6
clavicula (l)	96.0	91.6	-4.5
clavicula (r)	98.1	93.9	-4.3
femur (l)	91.4	92.1	0.8
femur (r)	94.2	92.4	-1.9
hip (l)	96.8	98.5	1.7
hip (r)	96.9	98.5	1.7
sacrum	96.3	97.7	1.4
face	79.7	75.2	-5.7
gluteus maximus (l)	95.8	97.1	1.3
gluteus maximus (r)	96.1	97.2	1.2
gluteus medius (l)	95.8	96.7	1.0
gluteus medius (r)	95.9	97.0	1.1
gluteus minimus (l)	94.8	94.6	-0.2
gluteus minimus (r)	95.2	95.4	0.2
autochthon (l)	96.5	96.9	0.4
autochthon (r)	96.4	96.8	0.4
iliopsoas (l)	94.2	92.9	-1.4
iliopsoas (r)	94.7	94.0	-0.7
urinary bladder	93.4	94.3	1.0
End of Table			

Chapter 6

Summary of Current Trends and Future Directions

6.1 Trends

6.1.1 Models and frameworks

In the field of medical image segmentation, the nnUNet framework has emerged as a state-of-the-art model and framework, with its success attributed in part to the effectiveness of the framework. However, in other benchmarks for semantic segmentation, transformer-based models such as BEiT (Bao et al., 2022) and Swin (Liu et al., 2021) have shown promising results and are in close competition with each other. These models, including ViT-Adapter-L (Chen et al., 2022) and MasK DINO (Li et al., 2022), have achieved strong performance on benchmarks such as ADE20K (Zhou et al., 2017) (Zhou et al., 2018a) and Cityscapes (Cordts et al., 2016), suggesting that they may be game changers in the field of medical image processing. It is important to keep track of these developments and compare them to the nnUNet framework to determine their potential impact on the field. The use of transformer-based models and attention mechanisms (Vaswani et al., 2017) looks promising for future developments in medical image segmentation.

The current trends in medical image segmentation are focused on the use of transformer-based models. One notable example is the Swin UNETR, which represents an attempt to integrate transformers into traditional segmentation models such as the U-Net. While the Swin UNETR has demonstrated promising results and is able to achieve performance that is close to the nnUNet on the AMOS benchmark, it has yet to surpass the nnUNet as the state-of-the-art model for medical image segmentation. This highlights the continued importance of the nnUNet framework as a reliable and effective tool for biomedical segmentation tasks, as well as the need for further research and development in this area to advance the state-of-the-art in medical image segmentation.

In addition, the MONAI framework has emerged as a flexible and extensible toolkit for medical image analysis, providing support for a range of segmentation algorithms and architectures. As this framework continues to evolve, it has the potential to play a critical role in the standardization and comparison of different models, particularly in terms of generalization performance.

Self-supervised learning has also gained significant attention in recent years as a promising approach to building AI systems that can learn from vast amounts of

data without the need for extensive labeling. As noted in a recent paper by Facebook AI researchers (*Self-supervised learning* n.d.), self-supervised learning has had a particularly profound impact on natural language processing, but has yet to achieve the same success in computer vision. Despite this, there have been promising early results in the use of self-supervised learning for vision tasks, and ongoing research in this area has the potential to greatly advance the field of medical image segmentation. By leveraging the structure of the data itself to provide supervisory signals, self-supervised learning can allow for the learning of more subtle and nuanced representations of the world, ultimately bringing us closer to human-level intelligence in AI systems.

6.1.2 Importance of intra and interannotator variability:

In order to evaluate the performance of semantic segmentation models, it is necessary to have reliable and consistent annotations as ground truth. This is where inter-rater reliability (IRR) and inter-annotator agreement (IRA) come into play. (Gisev, Bell, and Chen, 2013)

Inter-rater reliability refers to the consistency of annotations provided by different annotators, while inter-annotator agreement refers to the extent to which different annotators agree on the same annotation. Both of these concepts are important for ensuring the quality and reliability of annotations, as well as for comparing the performance of different models to human annotators.

Although often used interchangeably, there is a technical distinction between the terms agreement and reliability and therefore IRA and IRR. Fundamentally in the context of research studies, agreement is defined as the degree to which scores/ratings are identical, whereas reliability relates to the extent of variability and error inherent in a measurement (Fig. 6.1)

Equation 1 $$\text{Reliability} = \frac{\text{Subject variability}}{\text{Subject variability} + \text{Measurement error}}$$

Equation 2 $$\text{Percent agreement} = \frac{\text{Number of concordant responses}}{\text{Total number of responses}} \times 100\%$$

Equation 3 $$\text{Kappa } (\kappa) = \frac{\text{Proportion observed agreement} - \text{Proportion expected chance agreement}}{1 - \text{Proportion expected chance agreement}}$$

Equation 4 $$\text{Intraclass correlation coefficent (ICC)} = \frac{\text{Between subjects variance}}{\text{Between subjects variance} + \text{Within subjects variance}}$$

FIGURE 6.1: Equations relating to IRA and IRR calculations.

IRA indices, therefore, relate to the extent to which different raters assign the same precise value for each item being rated. In contrast, IRR indices relate to the extent to which raters can consistently distinguish between different items on a measurement scale. The general trend in ratings is important, not the absolute value assigned by each of the raters, and the variation between ratings and measurement error is accounted for in IRR.

One study that demonstrated the importance of intraannotator variability and interannotator variablity as well as intersubject variability is (Schoppe et al., 2020), which used multiple annotators to label organs in whole-body mouse scans. The authors found that using multiple annotators improved the reliability of the annotations, and they were able to demonstrate that their deep learning model performed similarly to the average performance of the human annotators.

Following (Schoppe et al., 2020) study for the assessment of variability in organ volumetry, we can quantify the mean volume per organ across subject for each dataset and quantify three kinds of variability. In the following equations, m will denote the index of a given subject and M the total number of subjects of the dataset; t will denote the index of a time point of a scan of a subject and T the total number of scans per subject; a will denote the index of a given annotator and A the total number of annotators; and o will denote the index of a given organ.

The intersubject variability quantifies the standard deviation of organ volumes across subjects:

$$\text{Intersubject variability}\,(o) = \sqrt{\sum_{m=1}^{M}\left(\overline{v_{m,o}} - \frac{1}{M}\sum_{m=1}^{M}\overline{v_{m,o}}\right)^2}$$
$$\text{with} \quad \overline{v_{m,o}} = \frac{1}{T}\sum_{t=1}^{T} v_{m,t,o}$$

The interannotator variability quantifies the mean standard deviation of independently created annotations across all scans:

$$\text{Interannotator variability}\,(o) = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{T}\sum_{t=1}^{T}\sqrt{\sum_{a=1}^{A}\left(v_{m,t,o,a} - \frac{1}{A}\sum_{a=1}^{A} v_{m,t,o,a}\right)^2}$$

The intraannotator variability quantifies the mean standard deviation of several annotations by the same annotator for the same subject:

$$\text{Intraannotator variability}\,(o) = \frac{1}{M}\sum_{m=1}^{M}\sqrt{\sum_{t=1}^{T}\left(v_{m,t,o,a=1} - \frac{1}{T}\sum_{t=1}^{T} v_{m,t,o,a=1}\right)^2}.$$

Similarly, Nikolov et al. (2018)(Nikolov et al., 2021b) (Nikolov et al., 2021a) used multiple annotators to label head and neck anatomy in CT scans for radiotherapy.

To provide a human clinical comparison for the algorithm, each case was manually segmented by a single radiographer with arbitration by a second radiographer. This was compared to their study's 'gold standard' ground truth graded by two further radiographers and arbitrated by one of two independent specialist oncologists, each with a minimum of four years specialist experience in radiotherapy treatment planning for head and neck patients. They compare their performance (model vs oncologist) to radiographer performance (radiographer vs oncologist).

Model performance was evaluated alongside that of therapeutic radiographers (each with at least 4 years of experience) segmenting the test set of UCLH images (section 2.7) independently of the oncologist-reviewed scans (which they used as their ground truth). They found that the model performed similarly to humans: on all OARs studied there was no clinically meaningful difference between the deep

learning model's segmentations and those of the radiographers.

In conclusion, these variabilities, agreement and reliability concepts are important considerations when evaluating the performance of semantic segmentation models in medical image processing. By using multiple annotators and carefully addressing any discrepancies between their annotations, it is possible to obtain reliable and consistent ground truth data that can be used to accurately compare the performance of different models to human annotators.

6.2 Future work and roadmap

In this work, efforts were made to provide extensive information on main blocs constituting the research for segmentation of organs in CT scans. In future work we aim to develop a dataset and deep learning framework for a specific clinical application. Therefore, we attempt to draft a roadmap in order to achieve best results thanks to a deep learning framework and with focus on a data collection campaign which is essential in order to obtain and evaluate state-of-the-art results on any new anatomical structure or pathology. We will also provide a methodologies and tools to pursue the collection and annotation of images.

6.2.1 Roadmap

To obtain optimal results for a specific application we have outlined the following steps:

1. Identify Prioritized Anatomical Structures: The first step in building a deep learning framework for medical segmentation is to identify the anatomical structures that are the most important to annotate. This will involve consulting with specialists in the field, such as radiologists and oncologists, to determine which structures are of the highest priority. For example, Dr. Ines Saab, a radiologist at the HUG, has emphasized the importance of annotating lymph nodes.

2. Determine the Required Number of Annotated Images: Based on our experience, we have found that a certain number of annotated images is required to achieve a certain range of performance. In our case, this number is around 100 images.

3. Data Collection and Annotation Campaign: The third step is to launch a data collection and annotation campaign in hospitals. This will involve recruiting experienced radiologists and oncologists to supervise the annotation process. Tools such as 3D slicer and NORA will be used to provide intuitive and comprehensive visualization and annotation platforms. Modules can be installed to deploy models, perform active learning, and automate annotation tasks using tools such as Deepgrow.

4. Simplifying the Annotation Process: To simplify the annotation process, we will use a pre-trained model trained on over 4000 images in our database. With active learning, annotators will only have to correct the model's predictions, reducing the amount of work required.

5. Measure Similarity between Annotators: Measuring similarity between annotators is important to address difficult cases with a consensus method. This improves the quality of the data and makes predictions more robust. It also allows us to compare the performance of the final model to that of human annotators.

6. Developing a Deep Learning Framework: To maintain state-of-the-art performance, we will develop a deep learning framework using the MONAI library. Our framework will be inspired by nnU-Net and will feature modularity that allows for updating the architecture, processing, metrics, and automatic configuration.

7. Keeping Up with Trends: Finally, we will keep up with trends in the field, such as the use of transformer-based models and new metrics for edge cases. We will test state-of-the-art models in other segmentation applications on our benchmarks when the codes are published online. This will allow us to incorporate any significant advancements into our framework.

6.3 Data collection campaign

In this section, to provide a detailed workflow for a data collection campaign and annotations we refer to a recent work done by Wasserthal, 2022. which details the experimentations and results of proposed workflow. More details on how they trained their model and the evalutation and results are available in their paper. The totalsegmentator team segmented 104 anatomical structures in 1204 CT images (27 organs, 59 bones, 10 muscles, 8 vessels) covering a majority of relevant classes for most use cases. They show an improved workflow for the creation of ground truth segmentations which speeds up the process by over 10x. The CT images were randomly sampled from clinical routine, thus representing a real world dataset which generalizes to clinical application. The dataset contains a wide range of different pathologies, scanners, sequences and sites. Finally, they train a segmentation algorithm called Totalsegmentator (a trained nnU-net) on this new dataset. They have proven to solve important problems such as obtaining a model that segments a wide range of anatomical structures and that robustly works on clinical data since it was trained on a large clinical dataset.

6.3.1 Data selection

Some steps to collect CT images to populate a new dataset are described as follows: Over a timespan of about 10 years, we randomly sample CT images from a PACS (picture archiving and communication system). Only images from patients with a general research consent can be used. CT images of legs and hands may be excluded for example if they are not in the focus of the project's application. We also sample the CT series of each examination randomly. This enable the dataset to contain CT images with different sequences (native, arterial, portal venous, late phase, dual energy), with and without contrast agent, with different bulb voltages, with different slice thicknesses and resolution and with different kernels (soft tissue kernel, bone kernel). If the human expert annotator is unsure how to segment certain structures because of high ambiguity (e.g. structures highly distorted by pathologies), the examination can be excluded (40 subjects in the Totalsegmentator collection campaign).

6.3.2 Data annotation

Complete manual segmentation for one class is estimated to take roughly 10 min. Based on this time for manual segmentation of 104 classes in 1204 subjects, like for the Totalsegmentator dataset, it would take one person over 10 years to complete the segmentation (assuming the person is working 8 h per day and 5 days a week). This is not feasible. We can use the following approach to speed up the process by several orders of magnitude:

Using existing models

We can use the Totalsegmentator model to create a first segmentation which can be then further refined manually if necessary. Given the potential high variability of such a collected dataset a further refinement may be often necessary. For any remaining classes, manual segmentation is necessary. Supervision of the refinement of segmentations as well as manual segmentation is advised (for example supervised by a board-certified radiologist).

Transfer of segmentations between CTs with and without contrast agent

A procedure described by Wasserthal, 2022 is useful for particular structures to segment (for example the different heart subparts). On all different CT series (also without contrast agent), the following approach is described: An inhouse dataset was available with ground truth segmentations of the heart subparts on CT images with contrast agent. For each image with contrast agent also a native CT image without contrast agent was available. Since this image was from the same examination it was well aligned with the contrast agent image. So, they transferred the segmentation to the CT images without contrast agent and trained a U-Net (Ronneberger et al., 2015) on images with and without contrast agent. This model was then used to predict the heart subparts in the totalsegmentator dataset, resulting in very accurate segmentations of the heart subparts in all kinds of CT sequences.

Improved segmentation workflow

To speedup manual segmentation a proposed platform is the Nora imaging platform (Anastasopoulos, Reisert, and Kellner, 2017). Compared to traditional tools like MITK (Wolf et al., 2005), Nora provides the following advantages: First, subjects do not have to be loaded individually, instead Nora provides a list of all subjects. Second, segmentation masks do not have to be created and saved individually for each image, but whenever you change the subject in Nora, the correct masks are created, named and saved automatically. Third, the autoloading as well as the view (e.g. the intensity window) can be configured for each project individually. Thus, it took Wasserthal, 2022 only 17 s in Nora to load the case, create a segmentation mask and save the segmentation mask, whereas with MITK this procedure took them 45 s. Forth, an even higher speedup is achieved by doing rough segmentations instead of pixel perfect segmentations. With Wasserthal, 2022, those rough segmentations were preliminary and not sufficient for the final result. However, the U-Net is good at learning good segmentations from rough segmentations. The errors in the rough segmentations are reported rather random across subjects (no systematic error) and therefore the U-Net learns to ignore the random part of the segmentation and keeps the consistent part which aligns pretty well with the anatomical structure. Figure 6.2 shows how the U- Net produces a smooth segmentation from the rough input. Fifth,

since rough segmentations are sufficient they used a pencil for segmentation which draws through 3 slices at the same time, providing another speedup. Exemplary, detailed segmentation of the urinary bladder takes 10 min whereas rough segmentation only takes 1 min 25 s and the U-Net produces accurate results also from these rough segmentations. Overall, using a traditional segmentation approach takes 45 s for loading/saving and 10 min for segmentation whereas using the proposed approach it only takes 17 s for loading/saving and 1 min 25 s for segmentation. Thus, this process could be made 7x faster.

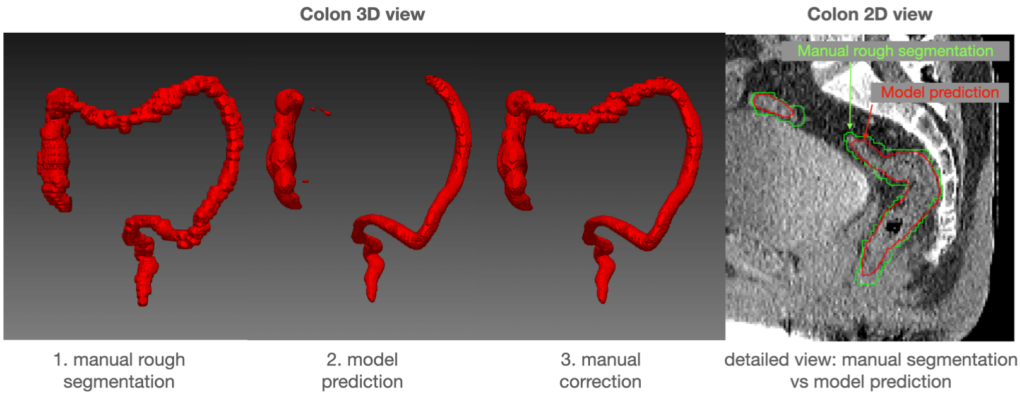


FIGURE 6.2: Example of steps from ground truth creation: It can be seen how the model generates smooth predictions from rough ground truth segmentations and how errors in the model predictions are corrected.

Another alternative is to use 3D slicer and (Deepgrow and Monailabel) (Section 6.3.3)

Active learning

For further speedup, we can use an active learning approach: After finishing manual segmentation of the first 5 subjects, a U-Net can be trained and only its predictions are to be corrected corrected. With totalsegmentator, the U-Net was then retrained again after 20 subjects, 100 subjects and 1000 subjects. After each iteration, less manual correction was necessary.

3D preview for quality control

Another great speedup can be achieved by using 3D renderings using the python fury library (Garyfallidis et al., 2021) of the masks to spot errors. Wasserthal, 2022 relates that after 1 or 2 iterations of the active learning the predictions of most segmentations did not need further refinement. However, it takes a lot of time to load all masks and scroll through all slices to check if they are fine or need correction. This task can be greatly accelerated by generating one PNG image with a 3D rendering of all the 104 masks for the Totalsegemntator. They tested this approach on many subjects and it turns out that if a mask has errors you can typically spot it in the 3D rendering (see figure 6.3). Moreover, the loading of the PNG image is instant, instead of several minutes for loading 104 masks. With this approach, 104 masks

can be checked for errors within several seconds instead of several minutes with a traditional approach. When evaluating segmentations only based on 3D renderings minor errors can be missed. However, for evaluation of major anatomical structures (in contrast to e.g. small lung nodules) errors which can not be seen in the 3D rendering are very rare. And also if evaluating segmentations slice by slice errors can easily be missed since expert raters tend to become negligent when looking at thousands of slices. In several cases they spotted errors in the 3D rendering which were missed by a previous slice by slice evaluation.

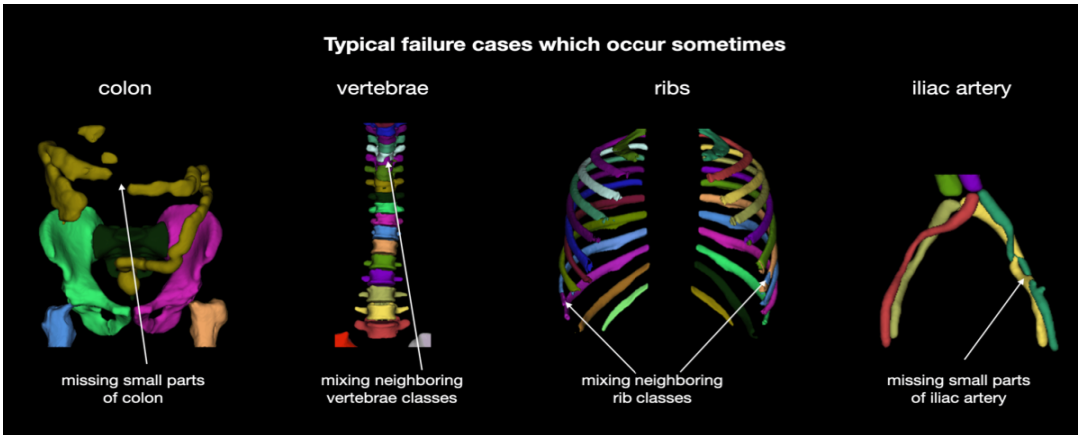


FIGURE 6.3: Overview of typical failure cases of the Totalsegmentator model. Users should be aware that these problems can occur in roughly 15% of the subjects.

6.3.3 Additional deep learning framework, visualization and annotation tools

3D Slicer open source software

One challenge that researchers face when using deep learning frameworks for medical image segmentation is the need to effectively visualize and interact with the results of their segmentation models. 3D Slicer (*3D Slicer image computing platform | 3D Slicer* n.d.) is an open source software platform for medical image computing and visualization. It can be used to develop an interface to a deep learning framework for the segmentation of medical images. The platform provides a number of tools and features that can be used to pre-process, visualize, and analyze medical imaging data, as well as to develop, test, and deploy deep learning models for segmentation tasks.

One key advantage of 3D Slicer is its ability to visualize 3D medical images in a variety of ways, including surface rendering, volume rendering, and multi-planar reformatting. This allows researchers to easily visualize the output of their segmentation models and compare it to the original images. In addition to visualization, 3D Slicer offers a range of tools for interacting with medical images, including tools for manual segmentation, landmark identification, and measurement. These tools can be useful for verifying the accuracy of the output of a deep learning model, or for fine-tuning the model by providing additional training data.

One of the key benefits of using 3D Slicer in combination with a deep learning framework is the ability to easily and efficiently label large quantities of medical images. The platform includes a number of interactive tools that can be used to manually annotate images and create training datasets for deep learning models. In addition, 3D Slicer provides a number of semi-automatic and automatic registration tools that can be used to align and register images, models, and other objects in 3D space.

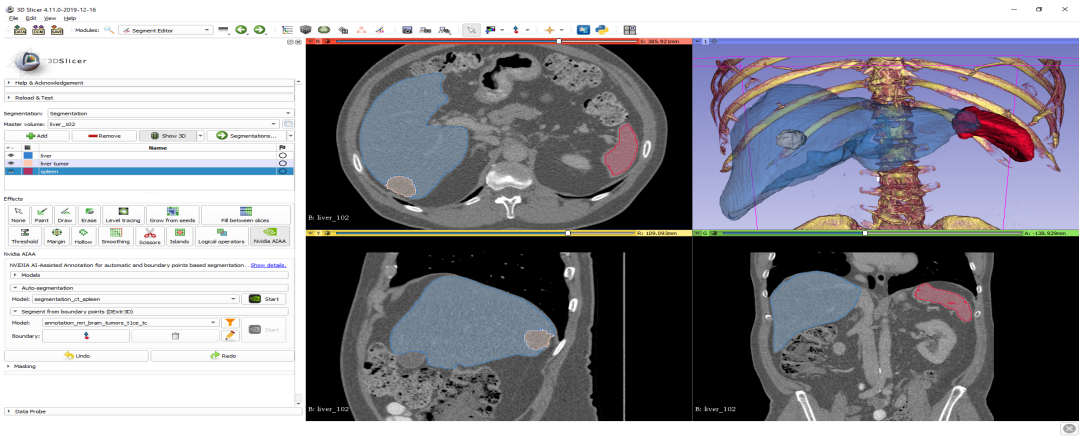


FIGURE 6.4: Illustration of the 3D slicer software

Annotation campaign and Active Learning

MONAI Label (*MONAI Label — MONAI Label 0.6.0 Documentation* n.d.) can be used for a data collection campaign of anatomical and pathological structures in a number of ways. First, MONAI Label provides a range of tools for interactive and automated segmentation of images, including both deep learning models and manual annotation tools such as DeepEdit and DeepGrow. This enables experienced annotators, such as doctors, radiologists, and oncologists, to quickly and accurately label a large quantity of scans with minimal effort.

In addition to its segmentation capabilities, MONAI Label also provides tools for managing and organizing large datasets, including the ability to track and store annotation progress, assign tasks to specific annotators, and manage permissions and access to the data. This makes it easy for annotators to collaborate and work efficiently on a data collection campaign, even if they are located in different locations.

Finally, MONAI Label's active learning capabilities allow the deep learning models to continually improve as more data is collected and annotated. This means that the models can become more accurate and efficient over time, further reducing the time and effort required for annotation.

MONAI Label is a powerful tool for data collection campaigns of anatomical and pathological structures, providing a range of tools and features that enable experienced annotators to efficiently label large quantities of medical imagery.

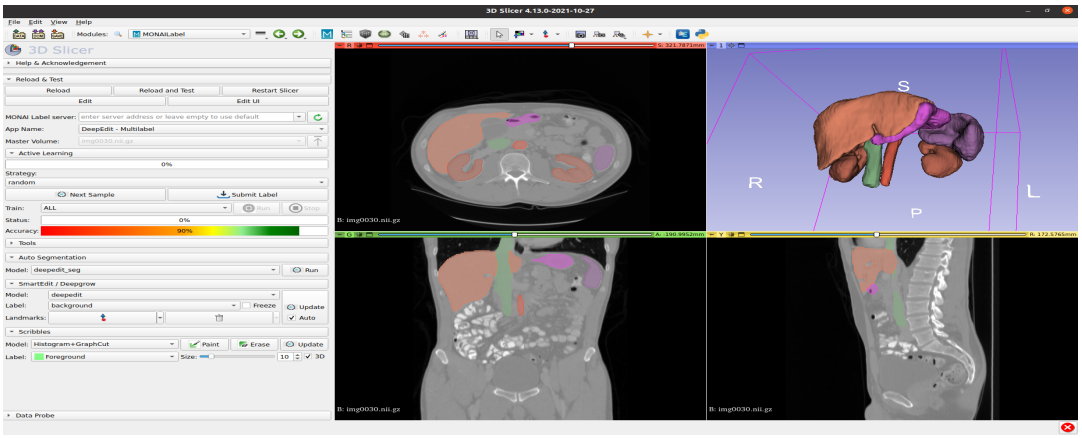


FIGURE 6.5: Illustration of the MONAI label module in the 3D slicer software

6.4 Collaboration with MONAI

Some of the key advantages of collaborating with MONAI on this project include:

1. Access to the latest research and developments in AI for healthcare imaging: MONAI is a highly regarded and well-established project in the field of AI in healthcare imaging, and collaborating with the project would provide the team with access to the latest research and developments in this rapidly-evolving field.

2. Access to a wide range of expertise and experience in medical imaging and AI: a collaboration with MONAI would provide EPFL with access to a wide range of expertise and experience in the field. MONAI includes researchers and experts from a variety of institutions and organizations. This would be beneficial for our own research and development efforts, and would help to ensure that our work is informed by the latest best practices and developments in the field.

3. Opportunity to contribute to the development of new and innovative AI-based solutions for healthcare imaging: One of the goals of MONAI is to accelerate the pace of innovation in the field of AI in healthcare imaging, and collaborating with the project would give EPFL and the proposed platform the opportunity to be part of this important work. The EPFL team could contribute its own expertise and experience to the development of new algorithms, methods, and applications that could help to improve the accuracy and effectiveness of AI in healthcare imaging.

More specifically in terms of developing tools several specific advantages for building the platform, including for visualization, data management, and deep learning:

- Building a visualization tool: One of the key challenges for any platform that processes and analyzes medical imaging data is to provide clear and useful visualizations of the results. Collaborating with MONAI could be useful for building a visualization tool for the platform, as the project includes experts and working groups that are focused on data representation and visualization. This expertise could be valuable for developing effective and user-friendly visualizations for the platform.

- Building a data base: Another important aspect of the proposed platform is the need to manage and store the large volumes of data that will be processed by the platform, including both the input imaging data and the results of the deep learning analysis. Collaborating with MONAI could be useful for building a data base for the platform, as the project includes experts and working groups that are focused on data management and storage, including support for bioinformatics, biomarkers, and metadata. This expertise could be valuable for developing effective and efficient data management solutions for the platform.

- Building a deep learning framework: As mentioned, the proposed platform will use deep learning to process and analyze the imaging data, and developing a robust and effective deep learning framework is a key challenge for the project. Collaborating with MONAI could be useful for building a deep learning framework for the platform, as the project includes the MONAI Core framework, which is a powerful and versatile toolkit for deep learning in medical imaging. This framework could be useful for developing and training the deep learning model that will be used by the platform, and could also be used to implement other functionality such as data preprocessing and augmentation.

- Building a useful interface for radiologists: Another important aspect of the proposed platform is the need to provide a user-friendly and effective interface for radiologists to use when uploading and analyzing imaging data. Collaborating with MONAI could be useful for building a useful interface for radiologists, as the project includes experts and working groups that are focused on user experience and usability. This expertise could be valuable for developing an interface that is clear, intuitive, and easy to use for radiologists.

Chapter 7

Conclusion

In conclusion, the field of medical image segmentation has seen significant progress in recent years, with deep learning and neural network-based approaches emerging as the state-of-the-art for many tasks. The nnUNet framework and U-Net architecture have proven to be particularly effective, with strong performance on a range of benchmarks and real-world applications. However, the field is rapidly evolving, and new models and approaches, such as transformer-based models and self-supervised learning, are showing great promise and are likely to play a significant role in the future of medical image segmentation. It will be important for researchers to continue to track these developments and evaluate their potential impact on the field. In addition, the availability of open source datasets and frameworks such as MONAI and 3D Slicer provide valuable resources for researchers and practitioners to develop and test new approaches to medical image segmentation. With continued research and innovation, the field is well positioned to make significant advances in the analysis and interpretation of medical images, ultimately leading to improved patient care and outcomes.

In addition, building a deep learning framework for semantic segmentation applied to medical imagery processing requires careful consideration of various factors. These include the goals and requirements of the framework, the availability and quality of datasets, the choice of architecture and framework, and the need for visualization and interaction tools. By carefully reviewing the literature and considering these factors, it is possible to develop a framework that is modular, quickly adaptable to different datasets, and capable of producing high-quality segmentation results. The MONAI framework and 3D Slicer software offer promising options for achieving these goals, and launching a data collection campaign with experienced annotators can help to ensure that sufficient data is available for training and evaluating the framework. Overall, this overview paper provides a comprehensive overview of the challenges and opportunities involved in building a deep learning framework for semantic segmentation applied to medical imagery processing.

Appendix A

Appendix A

TABLE A.1: Datasets from MedSeg

Name	Description	Structures Identified
1. Lateral ventricules	50 MRI Segmentation of the lateral ventricles in 50 MRI Brain from 50 different datasets to ensure maximum heterogeneity. Images obtained from openneuro.org.	Lateral Ventricles
2. liver segments	50 N/A Case collection of 50 livers with their segments. Images obtained from Decathlon Medical Segmentation competition.	Liver Segments
3. Inferior vena cava	20 N/A Dataset derived from Declathon with the inferior vena cava segmented on 20 cases.	Inferior Vena Cava

Continued on next page

Table A.1 – continued from previous page

Name	Description	Structures Identified
4. Brain vasculature	1 MRI Single dataset showing a large array of arteries (and some veins) in the brain. Right hand side is highly detailed, left hand side more simple segmentation showing divisions of the major arteries. 7T MRI TOF.	1. Paraopthalmic part of right internal carotid artery 2. Posterior communicating part of right internal carotid artery 3. Anterior choroidal part of the right internal carotid artery 4. Right carotid terminus 5. Right M1 6. Right posterior communicating artery 7. Right P1 8. Right P2 9. Left P1 10. Left posterior communicating artery 11. Left P2 12. Right A1 13. Anterior communicating artery (partially fenestrated, variant) 14. Right A2 15. Left paraopthalmic part of the internal carotid 16. Left opthalmic artery 17. Left posterior communicating part of the internal carotid 18. Left anterior choroidal artery part of the internal carotid 19. Left carotid terminus 20. Left A1 21. Left A2 22. Right opthalmic artery 23. Left M1 24. Right anterior choroidal artery 25. Right posteromedial central arteries 26. Right superior cerebellar artery 27. Right paramedian artery branch 28. Left anterior choroidal artery 29. Right medial frontobasal artery 30. Left medial orbitofrontal artery 31. Left polar frontal artery 32. Right polar frontal artery 33. Left pericallosal artery 34. Left callosomarginal artery (A3/4) 35. Right callosomarginal artery (A3/4) 36. Right anterior internal frontal artery 37. Left anterior internal frontal artery 38. Right intermediate internal frontal artery 39. Left intermediate internal frontal artery 40. Basilar artery 41. Insular segment of the right middle cerebral artery 42. Insular branches of the left middle cerebral artery (M3/M4) 43. Right lenticolstriate arteries 44. Left posteromedial central arteries 45. Right anterior insular part of middle cerebral artery (M2/M3/M4) 46. Right middle meningeal artery 47. Right prefrontal artery 48. Right artery of precentral gyrus 49. Right artery of central sulcus 50. Right polar temporal artery 51. Right temporal branch 52. Right anterior temporal artery branch 53. Right middle temporal artery branch 54. Right posterior parietal artery 55. Right posterior temporal artery branch of the MCA 56. Right parietal branches of the middle cerebral artery 57. Right posterior insular part of the middle cerebral artery 58. Right superior terminal branch of the middle cerebral artery 59. Superior sagittal sinus 60. Right posterior temporal branches of the PCA 61. Right medial occipital artery 62. Right precuneal branches 63. Right calcarine branches 64. Right parietooccipital branches 65. Left P3 66. Left P4/P5 67. Left M2-branches 68. Left M3/M4 69. Right parietal cortical branches of the insular part of the middle cerebral artery 70. Right superior cerebellar artery (duplication, variant) 71. Left superior cerebellar artery 72. Left lenticulostriate arteries 73. Right lateral frontobasal artery 74. Left middle meningeal artery

Continued on next page

Table A.1 – continued from previous page

Name	Description	Structures Identified
5. Lung lobes, vessels and airways	N/A N/A The data contains the following masks: Right upper lobe, Right middle lobe, Right lower lobe, Left upper lobe, Left lower lobe, Airways, Pulmonary arteries, Pulmonary veins. Data from TCIA.	Right upper lobe, Right middle lobe, Right lower lobe, Left upper lobe, Left lower lobe, Airways, Pulmonary arteries, Pulmonary veins
6. Lymph node regions in the neck	N/A N/A Dataset showing the lymph node regions in the neck (1-7) on the right side. Dataset obtained from TCIA. Following the guidelines published here.	Lymph node regions in the neck
7. Urinary system	N/A N/A Single dataset showing the kidney, blood vessels, ureter and bladder. For educational purposes. The dataset is free to download and use.	1. Kidney, 2. Renal vein, 3. Renal pelvis, 4. Renal artery, 5. Ureter, 6. Bladder
8. Deep spaces of the neck	1 N/A Single dataset showing the most commonly used deep spaces of the neck. Obtained from TCIA.	1. Carotid 2. Masticator 3. Buccal 4. Parotid 5. Submandibular 6. Submental 7. Posterior Cervical 8. Mucosal Pharyngeal 9. Perivertebral 10. Larynx 11. Sublingual 12. Paraspinal 13. Parapharyngeal
9. Vasculature of the neck	1 N/A Single dataset showing major arteries and veins in the neck. For educational purposes. Obtained from TCIA.	1. Brachiocephalic trunk 2. Common carotid artery 3. Internal carotid artery 4. External carotid artery 5. Subclavian artery 6. Vertebral artery 7. Basilar artery 8. Internal jugular vein 9. Brachiocephalic vein 10. Subclavian vein 11. External jugular vein 12. Posterior auricular vein 13. Retromandibular vein (posterior branch) 14. Facial vein 15. Transverse cervical veins 16. Superficial temporal veins, 17. Middle temporal vein,18. Deep superficial veins, 19. Retromandibular vein (anterior branch)

Continued on next page

Table A.1 – continued from previous page

Name	Description	Structures Identified
10. Esophageal Cancer	1 MRI Single case showing the outline of a tumor in the esophagus and surrounding structures in the mediastinum. For educational purposes. The dataset is free to download and use. Source: TCIA.	1. Esophageal cancer, 2. Lymph nodes, 3. Trachea, 4. Right main bronchus, 5. Left main bronchus, 6. Pulmonary trunk, 7. Arch of Azygos, 8. Aorta, 9. Supreme intercostal vein, 10. Fat stranding, 11. Left pulmonary artery, 12. Right pulmonary trunk, 13. Superior vena cava, 14. Azygos vein
11. Vasculature of the Abdomen	1 3D CT Single case with near-total segmentation of important vascular structures in the abdomen, 43 labels in total. This case contains a few but not uncommon variants. Great teaching example.	1. Inferior vena cava, 2. Left common iliac artery, 3. Right common iliac artery, 4. Aorta, 5. Coeliac trunk, 6. Common hepatic artery, 7. Gastroduodenal artery, 8. Hepatic artery proper, 9. Right hepatic artery, 10. Left hepatic artery, 11. Left gastric artery, 12. Right gastric artery, 13. Splenic artery, 14. Portal vein, 15. Right portal vein, 16. Left portal vein, 17. Superior mesenteric vein, 18. Left gonadal vein, 19. Right gonadal vein, 20. Inferior mesenteric vein, 21. Splenic vein, 22. Short gastric veins, 23. Jejunal branches of the superior mesenteric vein, 24. Left renal vein, 25. Right renal vein, 26. Superior mesenteric artery, 27. Right renal artery, 28. Left renal artery, 29. Left ascending lumbar vein, 30. Hemiazygos vein, 31. Azygos vein, 32. Separate segment 5/6 vein (variant), 33. Right hepatic vein (segment 7 only here, variant), 34. Middle hepatic vein, 35. Left hepatic vein, 36. Ileal branches of the superior mesenteric vein, 37. Superior posterior pancreaticoduodenal vein, 38. Right colic vein, 39. Right gastroepiploic vein, 40. Middle colic vein, 41. Inferior mesenteric artery, 42. Pancreaticoduodenal veins (plexus), 43. Right common iliac vein

Continued on next page

Table A.1 – continued from previous page

Name	Description	Structures Identified
12. Vasculature of the Pelvis	1 3D CT Case from Medical Segmentation Decathlon. Nearly all vascular structures of the pelvis. 74 labels. Not all structures are shown on both sides. A few interesting and not uncommon anatomical variants.	1. Aorta, 2. Right common iliac artery, 3. Left common iliac artery, 4. Left external iliac artery, 5. Right external iliac artery, 6. Inferior mesenteric artery, 7. Inferior vena cava, 8. Right common iliac vein, 9. Left common iliac vein, 10. Right external iliac vein, 11. Left external iliac vein, 12. Right inferior epigastric artery, 13. Left inferior epigastric artery, 14. Right internal iliac artery, 15. Left internal iliac artery, 16. Middle sacral artery, 17. Right deep circumflex iliac artery, 18. Left deep circumflex iliac artery, 19. Right superficial epigastric vein(s), 20. Left superficial epigastric vein(s), 21. Right common femoral artery, 22. Left common femoral artery, 23. Right common femoral vein, 24. Left common femoral vein, 25. Right gonadal vein (double/fenestration, variant), 26. Left gonadal vein, 27. Right great saphenous vein, 28. Left great saphenous vein, 29. Right deep femoral vein, 30. Right internal iliac vein, 31. Right superior gluteal vein, 32. Right inferior gluteal vein, 33. Right inferior vesical veins, 34. Right obturator veins, 35. Right femoral artery, 36. Right deep femoral artery, 37. Left femoral artery, 38. Left deep femoral artery, 39. Right femoral vein, 40. Left femoral vein, 41. Right pubic veins, 42. Right internal pudendal artery and vein, 43. Right obturator artery, 44. Right inferior gluteal artery, 45. Left internal iliac vein, 46. Left superior gluteal vein, 47. Left ureter, 48. Right ureter, 49. Left inferior gluteal vein, 50. Left internal pudendal artery and vein, 51. Left inferior vesical veins, 52. Left obturator artery, 53. Left obturator veins, 54. Left superior gluteal artery, 55. Left inferior gluteal artery, 56. Right epigastric vein, 57. Left epigastric vein, 58. Superior rectal artery, 59. Left colic artery, 60. Superior sigmoid artery (Part of the sigmoid branches), 61. Superior rectal vein (with branches), 62. Sigmoid veins, 63. Inferior mesenteric vein, 64. Vesical plexus vein branch (only left side shown), 65. Inferior rectal veins, 66. Right inferior rectal veins, 67. Left inferior rectal vein, 68. Left colic vein, 69. Middle sacral vein, 70. Left sacral veins, 71. Left lateral circumflex vein, 72. Left lateral circumflex artery, 73. Right lateral circumflex artery, 74. Descending branch of the right lateral circumflex artery

Continued on next page

Table A.1 – continued from previous page

Name	Description	Structures Identified
13. Musculature of the Pelvis	1 3D CT Case from Medical Segmentation Decathlon. Nearly all muscles of the pelvis. 67 labels. A few muscles were difficult to separate accurately.	1. Right psoas, 2. Left psoas, 3. Right iliacus, 4. Left iliacus, 5. Right sartorius, 6. Left sartorius, 7. Right quadratus lumborum, 8. Left quadratus lumborum, 9. Right rectus abdominis, 10. Left rectus abdominis, 11. Right gluteus maximus, 12. Left gluteus maximus, 13. Right gluteus medius, 14. Right gluteus minimus, 15. Left gluteus medius, 16. Left gluteus minimus, 17. Right tensor fascia lata, 18. Left tensor fascia lata, 19. Right transversus abdominis, 20. Right external oblique, 21. Right internal oblique, 22. Right rectus femoris, 23. Left rectus femoris, 24. Left transversus abdominis, 25. Left external oblique, 26. Left internal oblique, 27. Right spinalis thoracis, 28. Right ilicostalis lumborum, 29. Right longissimus thoracis, 30. Left spinalis thoracis, 31. Left iliocostalis lumborum, 32. Left longissimus thoracis, 33. Right piriformis, 34. Left piriformis, 35. Right superior gemellus, 36. Right inferior gemellus, 37. Right obturator internus, 38. Right obturator externus, 39. Right pectineus, 40. Right gracilis, 41. Right adductor longus, 42. Right adductor brevis, 43. Right adductor minimus, 44. Right adductor magnus, 45. Right biceps femoris (long head), 46. Right semitendinosus, 47. Right semimembranosus, 48. Right quadratus femoris, 49. Right vastus intermedius, 50. Right vastus lateralis, 51. Left obturator externus, 52. Left obturator internus, 53. Left adductor magnus, 54. Left semitendinosus, 55. Left biceps femoris (long head), 56. Left semimembranosus, 57. Left gracilis, 58. Left adductor longus, 59. Left pectineus, 60. Left adductor brevis, 61. Left adductor minimus, 62. Left vastus intermedius, 63. Left vastus lateralis, 64. Left gemellus superior, 65. Left gemellus inferior, 66. Right ischiocavernosus, 67. Left ischiocavernosus

Appendix B

Appendix B

Here are elements from the paper (Isensee et al., 2021) defining what is an epoch in their framework:

1. Distilling this knowledge into successful method design results in the following heuristic rule: "initialize the patch size to median image shape and iteratively reduce it while adapting the network topology accordingly (including network depth, number and position of pooling operations along each axis, feature map sizes and convolutional kernel sizes) until the network can be trained with a batch size of at least two given GPU memory constraints."

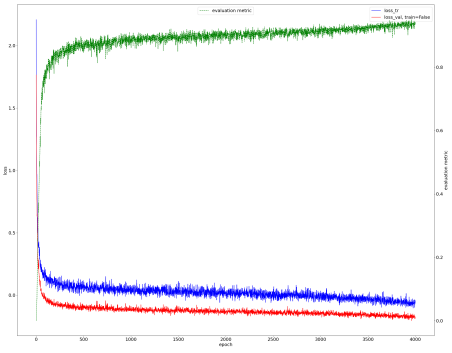
2. To enable large patch sizes, the batch size of the networks in nnU-Net is small. In fact, most 3D U-Net confgurations were trained with a batch size of only two (Fig. SN5.1a in Supplementary Note 5).

3. Training schedule. Based on experience and as a trade-off between runtime and reward, all networks are trained for 1,000 epochs, with one epoch being defined as iteration over 250 mini-batches.

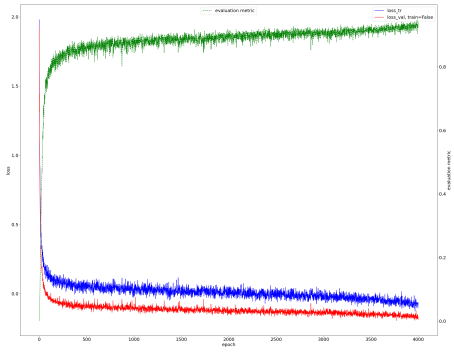
4. Samples for the mini-batches are chosen from random training cases. Oversampling is implemented to ensure robust handling of class imbalances; 66.7% of samples are from random locations within the selected training case, while 33.3% of patches are guaranteed to contain one of the foreground classes that are present in the selected training sample (randomly selected). The number of foreground patches is rounded with a forced minimum of 1 (resulting in one random and one foreground patch with a batch size of two). A variety of data augmentation techniques are applied on the fly during training: rotations, scaling, Gaussian noise, Gaussian blur, brightness, contrast, simulation of low resolution, gamma correction and mirroring. Details are provided in Supplementary Note 4.

Appendix C

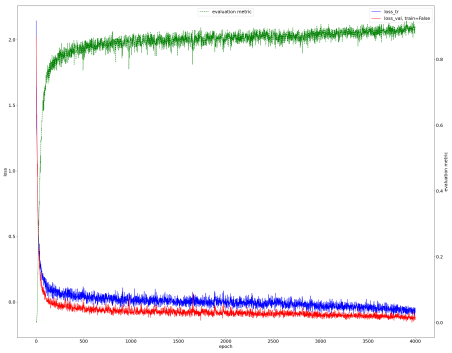
Appendix C



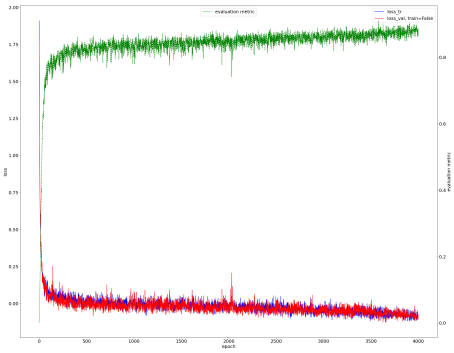
(A) 1082 training images



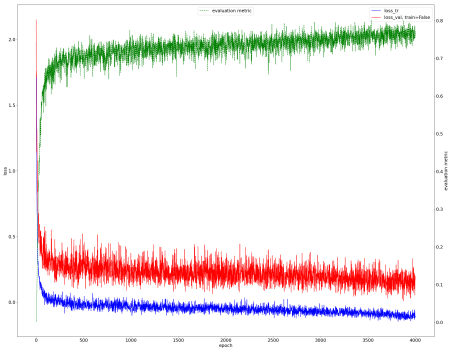
(B) 500 training images



(C) 100 training images



(D) 50 training images



(E) 10 training images

FIGURE C.1: training and validation loss along with Dice score progression curves for a nnU-Net pretrained on the WORD dataset and training on different numbers of images from the training set og the Totalsegmentator dataset

Bibliography

3D Slicer image computing platform | 3D Slicer (n.d.). URL: https://www.slicer.org/ (visited on 01/31/2023).

Abbe, Emmanuel and Colin Sandon (Apr. 29, 2019). *Provable limitations of deep learning*. DOI: 10.48550/arXiv.1812.06369. arXiv: 1812.06369[cs,math,stat]. URL: http://arxiv.org/abs/1812.06369 (visited on 12/16/2022).

Anastasopoulos, C., M. Reisert, and E. Kellner (Apr. 2017). ""Nora Imaging": A Web-Based Platform for Medical Imaging". In: *Neuropediatrics*. Abstracts of the 43rd Annual Meeting of the Society for Neuropediatrics. Vol. 48. ISSN: 0174-304X, 1439-1899 Issue: S 01 Journal Abbreviation: Neuropediatrics. Georg Thieme Verlag KG, P26. DOI: 10.1055/s-0037-1602977. URL: http://www.thieme-connect.de/DOI/DOI?10.1055/s-0037-1602977 (visited on 01/09/2023).

Antonelli, Michela et al. (July 15, 2022). "The Medical Segmentation Decathlon". In: *Nature Communications* 13.1. Number: 1 Publisher: Nature Publishing Group, p. 4128. ISSN: 2041-1723. DOI: 10.1038/s41467-022-30695-9. URL: https://www.nature.com/articles/s41467-022-30695-9 (visited on 12/27/2022).

arXiv.org e-Print archive (n.d.). URL: https://arxiv.org/ (visited on 01/30/2023).

Bao, Hangbo et al. (Sept. 3, 2022). *BEiT: BERT Pre-Training of Image Transformers*. DOI: 10.48550/arXiv.2106.08254. arXiv: 2106.08254[cs]. URL: http://arxiv.org/abs/2106.08254 (visited on 12/30/2022).

Bilic, Patrick et al. (Feb. 2023). "The Liver Tumor Segmentation Benchmark (LiTS)". In: *Medical Image Analysis* 84, p. 102680. ISSN: 13618415. DOI: 10.1016/j.media.2022.102680. arXiv: 1901.04056[cs]. URL: http://arxiv.org/abs/1901.04056 (visited on 12/18/2022).

Bosch, Walter R. et al. (2015). *Data From Head-Neck_Cetuximab*. In collab. with TCIA Team. Version Number: 1 Type: dataset. DOI: 10.7937/K9/TCIA.2015.7AKGJUPZ. URL: https://wiki.cancerimagingarchive.net/x/xwxp (visited on 01/09/2023).

Brouwer, Charlotte L. et al. (Oct. 2015). "CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines". In: *Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology* 117.1, pp. 83–90. ISSN: 1879-0887. DOI: 10.1016/j.radonc.2015.07.041.

Cardoso, M. Jorge et al. (Nov. 4, 2022). *MONAI: An open-source framework for deep learning in healthcare*. DOI: 10.48550/arXiv.2211.02701. arXiv: 2211.02701[cs]. URL: http://arxiv.org/abs/2211.02701 (visited on 12/28/2022).

CHAOS Challenge - combined (CT-MR) healthy abdominal organ segmentation - ScienceDirect (n.d.). URL: https://www.sciencedirect.com/science/article/pii/S1361841520303145?via%3Dihub (visited on 12/18/2022).

Chen, Zhe et al. (Oct. 23, 2022). *Vision Transformer Adapter for Dense Predictions*. DOI: 10.48550/arXiv.2205.08534. arXiv: 2205.08534[cs]. URL: http://arxiv.org/abs/2205.08534 (visited on 12/30/2022).

Clark, Kenneth et al. (Dec. 2013). "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository". In: *Journal of Digital Imaging* 26.6, pp. 1045–1057. ISSN: 0897-1889. DOI: 10.1007/s10278-013-9622-7. URL:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3824915/ (visited on 01/09/2023).

CodaLab - Competition (n.d.). URL: https://competitions.codalab.org/competitions/17094 (visited on 12/27/2022).

Cordts, Marius et al. (2015). "The Cityscapes Dataset". In: *CVPR Workshop on The Future of Datasets in Vision*.

Cordts, Marius et al. (Apr. 7, 2016). *The Cityscapes Dataset for Semantic Urban Scene Understanding*. DOI: 10.48550/arXiv.1604.01685. arXiv: 1604.01685[cs]. URL: http://arxiv.org/abs/1604.01685 (visited on 12/30/2022).

Cortes, Corinna and Vladimir Vapnik (Sept. 1, 1995). "Support-vector networks". In: *Machine Learning* 20.3, pp. 273–297. ISSN: 1573-0565. DOI: 10.1007/BF00994018. URL: https://doi.org/10.1007/BF00994018 (visited on 12/16/2022).

Cybenko, G. (Dec. 1, 1989). "Approximation by superpositions of a sigmoidal function". In: *Mathematics of Control, Signals and Systems* 2.4, pp. 303–314. ISSN: 1435-568X. DOI: 10.1007/BF02551274. URL: https://doi.org/10.1007/BF02551274 (visited on 12/16/2022).

Database (n.d.). MedSeg. URL: https://www.medseg.ai/database (visited on 12/30/2022).

Diba, Ali et al. (July 2017). "Weakly Supervised Cascaded Convolutional Networks". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE, pp. 5131–5139. ISBN: 978-1-5386-0457-1. DOI: 10.1109/CVPR.2017.545. URL: https://ieeexplore.ieee.org/document/8100028/ (visited on 12/16/2022).

Donahue, Jeff et al. (2017). "Long-Term Recurrent Convolutional Networks for Visual Recognition and Description". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.4, pp. 677–691. DOI: 10.1109/TPAMI.2016.2599174.

Garyfallidis, Eleftherios et al. (Aug. 4, 2021). "FURY: advanced scientific visualization". In: *The Journal of Open Source Software*. DOI: 10.21105/joss.03384.

Gerven, M. van and S. Bohte (2017). "Editorial: Artificial neural networks as models of neural information processing". In: *Frontiers in Computational Neuroscience* 11, p. 114. URL: https://doi.org/10.1155/2022/7355233.

Gisev, Natasa, J. Simon Bell, and Timothy F. Chen (2013). "Interrater agreement and interrater reliability: key concepts, approaches, and applications". In: *Research in social & administrative pharmacy: RSAP* 9.3, pp. 330–338. ISSN: 1934-8150. DOI: 10.1016/j.sapharm.2012.04.004.

Gu, Zaiwang et al. (Oct. 2019). "CE-Net: Context Encoder Network for 2D Medical Image Segmentation". In: *IEEE Transactions on Medical Imaging* 38.10. Conference Name: IEEE Transactions on Medical Imaging, pp. 2281–2292. ISSN: 1558-254X. DOI: 10.1109/TMI.2019.2903562.

Hanin, Boris (Oct. 18, 2019). "Universal Function Approximation by Deep Neural Nets with Bounded Width and ReLU Activations". In: *Mathematics* 7.10, p. 992. ISSN: 2227-7390. DOI: 10.3390/math7100992. arXiv: 1708.02691[cs,math,stat]. URL: http://arxiv.org/abs/1708.02691 (visited on 12/16/2022).

Hatamizadeh, Ali et al. (Jan. 4, 2022). *Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images*. DOI: 10.48550/arXiv.2201.01266. arXiv: 2201.01266[cs,eess]. URL: http://arxiv.org/abs/2201.01266 (visited on 12/19/2022).

Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh (July 2006). "A fast learning algorithm for deep belief nets". In: *Neural Computation* 18.7, pp. 1527–1554. ISSN: 0899-7667. DOI: 10.1162/neco.2006.18.7.1527.

Hochreiter, Sepp and Jürgen Schmidhuber (Dec. 1997). "Long Short-term Memory". In: *Neural computation* 9, pp. 1735–80. DOI: 10.1162/neco.1997.9.8.1735.

Hopfield, J J (Apr. 1982). "Neural networks and physical systems with emergent collective computational abilities." In: *Proceedings of the National Academy of Sciences* 79.8. Publisher: Proceedings of the National Academy of Sciences, pp. 2554–2558. DOI: 10.1073/pnas.79.8.2554. URL: https://www.pnas.org/doi/10.1073/pnas.79.8.2554 (visited on 12/16/2022).

Hornik, Kurt (Jan. 1, 1991). "Approximation capabilities of multilayer feedforward networks". In: *Neural Networks* 4.2, pp. 251–257. ISSN: 0893-6080. DOI: 10.1016/0893-6080(91)90009-T. URL: https://www.sciencedirect.com/science/article/pii/089360809190009T (visited on 12/16/2022).

Huang, Huimin et al. (Apr. 19, 2020). *UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation*. DOI: 10.48550/arXiv.2004.08790. arXiv: 2004.08790[cs,eess]. URL: http://arxiv.org/abs/2004.08790 (visited on 12/28/2022).

Hwang, Sangheum and Hyo-Eun Kim (2016). "Self-Transfer Learning for Fully Weakly Supervised Object Localization". In: *CoRR* abs/1602.01625. arXiv: 1602.01625. URL: http://arxiv.org/abs/1602.01625 (visited on 12/16/2022).

Isensee, Fabian et al. (Feb. 2021). "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In: *Nature Methods* 18.2. Number: 2 Publisher: Nature Publishing Group, pp. 203–211. ISSN: 1548-7105. DOI: 10.1038/s41592-020-01008-z. URL: https://www.nature.com/articles/s41592-020-01008-z (visited on 12/17/2022).

Ji, Yuanfeng et al. (Sept. 1, 2022). *AMOS: A Large-Scale Abdominal Multi-Organ Benchmark for Versatile Medical Image Segmentation*. DOI: 10.48550/arXiv.2206.08023. arXiv: 2206.08023[cs,eess]. URL: http://arxiv.org/abs/2206.08023 (visited on 12/18/2022).

Kavur, A. Emre et al. (Apr. 1, 2021). "CHAOS Challenge - combined (CT-MR) healthy abdominal organ segmentation". In: *Medical Image Analysis* 69, p. 101950. ISSN: 1361-8415. DOI: 10.1016/j.media.2020.101950. URL: https://www.sciencedirect.com/science/article/pii/S1361841520303145 (visited on 12/18/2022).

Lawrence, S. et al. (1997). "Face recognition: a convolutional neural-network approach". In: *IEEE Transactions on Neural Networks* 8.1, pp. 98–113. DOI: 10.1109/72.554195.

Lecun, Y. et al. (Nov. 1998). "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11. Conference Name: Proceedings of the IEEE, pp. 2278–2324. ISSN: 1558-2256. DOI: 10.1109/5.726791.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (May 1, 2015). "Deep learning". In: *Nature* 521.7553, pp. 436–444. ISSN: 1476-4687. DOI: 10.1038/nature14539. URL: https://doi.org/10.1038/nature14539.

Li, Feng et al. (Dec. 12, 2022). *Mask DINO: Towards A Unified Transformer-based Framework for Object Detection and Segmentation*. DOI: 10.48550/arXiv.2206.02777. arXiv: 2206.02777[cs]. URL: http://arxiv.org/abs/2206.02777 (visited on 12/30/2022).

Liu, Ze et al. (Aug. 17, 2021). *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows*. DOI: 10.48550/arXiv.2103.14030. arXiv: 2103.14030[cs]. URL: http://arxiv.org/abs/2103.14030 (visited on 12/30/2022).

Long, Jonathan, Evan Shelhamer, and Trevor Darrell (2015). "Fully convolutional networks for semantic segmentation". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440. DOI: 10.1109/CVPR.2015.7298965.

Lu, Zhou et al. (Nov. 1, 2017). *The Expressive Power of Neural Networks: A View from the Width*. DOI: 10.48550/arXiv.1709.02540. arXiv: 1709.02540[cs]. URL: http://arxiv.org/abs/1709.02540 (visited on 12/16/2022).

Luo, Xiangde et al. (Sept. 22, 2022). *WORD: A large scale dataset, benchmark and clinical applicable study for abdominal organ segmentation from CT image*. DOI: 10.48550/arXiv.2111.02403. arXiv: 2111.02403[cs,eess]. URL: http://arxiv.org/abs/2111.02403 (visited on 12/17/2022).

Ma, Jun et al. (July 20, 2021). *AbdomenCT-1K: Is Abdominal Organ Segmentation A Solved Problem?* DOI: 10.48550/arXiv.2010.14808. arXiv: 2010.14808[cs]. URL: http://arxiv.org/abs/2010.14808 (visited on 12/18/2022).

Marcus, Gary (Jan. 2, 2018). *Deep Learning: A Critical Appraisal*. DOI: 10.48550/arXiv.1801.00631. arXiv: 1801.00631[cs,stat]. URL: http://arxiv.org/abs/1801.00631 (visited on 12/16/2022).

McCulloch, Warren S. and Walter Pitts (Dec. 1, 1943). "A logical calculus of the ideas immanent in nervous activity". In: *The bulletin of mathematical biophysics* 5.4, pp. 115–133. ISSN: 1522-9602. DOI: 10.1007/BF02478259. URL: https://doi.org/10.1007/BF02478259.

MONAI Label — MONAI Label 0.6.0 Documentation (n.d.). URL: https://docs.monai.io/projects/label/en/latest/index.html (visited on 01/31/2023).

Multi-Atlas Labeling Beyond the Cranial Vault - Workshop and Challenge - syn3193805 - Wiki (n.d.). URL: https://doi.org/10.7303/syn3193805 (visited on 12/18/2022).

Multi-Modality Abdominal Multi-Organ Segmentation Challenge 2022 - Grand Challenge (n.d.). grand-challenge.org. URL: https://amos22.grand-challenge.org/ (visited on 12/27/2022).

Nature (Jan. 26, 2023). Nature. ISSN: 1476-4687. URL: https://www.nature.com/nature (visited on 01/30/2023).

Nievergelt, Jürg (July 1, 1969). "R69-13 Perceptrons: An Introduction to Computational Geometry". In: *Computers, IEEE Transactions on* C-18, pp. 572–572. DOI: 10.1109/T-C.1969.222718.

Nikolov, Stanislav et al. (July 12, 2021a). "Clinically Applicable Segmentation of Head and Neck Anatomy for Radiotherapy: Deep Learning Algorithm Development and Validation Study". In: *Journal of Medical Internet Research* 23.7, e26151. ISSN: 1438-8871. DOI: 10.2196/26151.

Nikolov, Stanislav et al. (Jan. 13, 2021b). *Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy*. DOI: 10.48550/arXiv.1809.04430. arXiv: 1809.04430[physics,stat]. URL: http://arxiv.org/abs/1809.04430 (visited on 12/30/2022).

Oktay, Ozan et al. (May 20, 2018). *Attention U-Net: Learning Where to Look for the Pancreas*. DOI: 10.48550/arXiv.1804.03999. arXiv: 1804.03999[cs]. URL: http://arxiv.org/abs/1804.03999 (visited on 12/28/2022).

Ouyang, Wanli et al. (June 1, 2015). *DeepID-Net: Deformable Deep Convolutional Neural Networks for Object Detection*. DOI: 10.48550/arXiv.1412.5661. arXiv: 1412.5661[cs]. URL: http://arxiv.org/abs/1412.5661 (visited on 12/16/2022).

Papernot, Nicolas et al. (Nov. 23, 2015). *The Limitations of Deep Learning in Adversarial Settings*. DOI: 10.48550/arXiv.1511.07528. arXiv: 1511.07528[cs,stat]. URL: http://arxiv.org/abs/1511.07528 (visited on 12/16/2022).

Papers with Code - The latest in Machine Learning (n.d.). URL: https://paperswithcode.com/ (visited on 01/30/2023).

PubMed (n.d.). PubMed. URL: https://pubmed.ncbi.nlm.nih.gov/ (visited on 01/30/2023).

Radiology (ACR), Radiological Society of North America (RSNA) {and} American College of (n.d.). *What are the benefits of CT scans?* Radiologyinfo.org. URL: https://www.radiologyinfo.org/en/info/safety-hiw_04 (visited on 01/30/2023).

Ranjbarzadeh, Ramin et al. (May 25, 2021). "Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images". In: *Scientific Reports* 11.1. Number: 1 Publisher: Nature Publishing Group, p. 10930. ISSN: 2045-2322. DOI: 10.1038/s41598-021-90428-8. URL: https://www.nature.com/articles/s41598-021-90428-8 (visited on 12/16/2022).

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by Nassir Navab et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 234–241. ISBN: 978-3-319-24574-4. DOI: 10.1007/978-3-319-24574-4_28.

Rosenblatt, F. (1958). "The perceptron: A probabilistic model for information storage and organization in the brain." In: *Psychological Review* 65.6, pp. 386–408. ISSN: 1939-1471, 0033-295X. DOI: 10.1037/h0042519. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/h0042519 (visited on 12/16/2022).

Rumelhart, David E., James L. McClelland, and CORPORATE PDP Research Group, eds. (1986). *Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations*. Cambridge, MA, USA: MIT Press. 547 pp. ISBN: 978-0-262-68053-0.

Schmidhuber, Juergen (Jan. 2015). "Deep Learning in Neural Networks: An Overview". In: *Neural Networks* 61, pp. 85–117. ISSN: 08936080. DOI: 10.1016/j.neunet.2014.09.003. arXiv: 1404.7828[cs]. URL: http://arxiv.org/abs/1404.7828 (visited on 12/16/2022).

Schoppe, Oliver et al. (Nov. 6, 2020). "Deep learning-enabled multi-organ segmentation in whole-body mouse scans". In: *Nature Communications* 11.1. Number: 1 Publisher: Nature Publishing Group, p. 5626. ISSN: 2041-1723. DOI: 10.1038/s41467-020-19449-7. URL: https://www.nature.com/articles/s41467-020-19449-7 (visited on 12/19/2022).

ScienceDirect.com | Science, health and medical journals, full text articles and books. (N.d.). URL: https://www.sciencedirect.com/ (visited on 01/30/2023).

Self-supervised learning (n.d.). *Self-supervised learning: The dark matter of intelligence*. URL: https://ai.facebook.com/blog/self-supervised-learning-the-dark-matter-of-intelligence/ (visited on 12/29/2022).

Sengupta, Saptarshi et al. (2020). "A review of deep learning with special emphasis on architectures, applications and recent trends". In: *Knowledge-Based Systems* 194, p. 105596. ISSN: 0950-7051. DOI: https://doi.org/10.1016/j.knosys.2020.105596. URL: https://www.sciencedirect.com/science/article/pii/S095070512030071X.

Shelhamer, Evan, Jonathan Long, and Trevor Darrell (Apr. 2017). "Fully Convolutional Networks for Semantic Segmentation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.4. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 640–651. ISSN: 1939-3539. DOI: 10.1109/TPAMI.2016.2572683.

Tang, Yucheng et al. (Mar. 28, 2022). *Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis*. DOI: 10.48550/arXiv.2111.14791. arXiv: 2111.14791[cs]. URL: http://arxiv.org/abs/2111.14791 (visited on 12/19/2022).

TCIA Test & Validation Radiotherapy CT Planning Scan Dataset (Nov. 13, 2022). originaldate: 2018-08-09T08:12:05Z. URL: https://github.com/deepmind/tcia-ct-scan-dataset (visited on 12/22/2022).

The Medical Segmentation Decathlon | Nature Communications (n.d.). URL: https://www.nature.com/articles/s41467-022-30695-9 (visited on 12/23/2022).

Vaswani, Ashish et al. (Dec. 5, 2017). *Attention Is All You Need*. DOI: 10.48550/arXiv.1706.03762. arXiv: 1706.03762[cs]. URL: http://arxiv.org/abs/1706.03762 (visited on 12/18/2022).

Wang, Xiao-Feng, De-Shuang Huang, and Huan Xu (Mar. 1, 2010). "An efficient local Chan–Vese model for image segmentation". In: *Pattern Recognition* 43.3, pp. 603–618. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2009.08.002. URL: https://www.sciencedirect.com/science/article/pii/S0031320309002982 (visited on 12/19/2022).

Wang, Yu et al. (Sept. 2019). "Lednet: A Lightweight Encoder-Decoder Network for Real-Time Semantic Segmentation". In: *2019 IEEE International Conference on Image Processing (ICIP)*. 2019 IEEE International Conference on Image Processing (ICIP). ISSN: 2381-8549, pp. 1860–1864. DOI: 10.1109/ICIP.2019.8803154.

Wasserthal, Jakob (Dec. 27, 2022). *TotalSegmentator*. original-date: 2022-01-19T12:24:33Z. URL: https://github.com/wasserth/TotalSegmentator (visited on 12/27/2022).

Wasserthal, Jakob et al. (2022). *TotalSegmentator: robust segmentation of 104 anatomical structures in CT images*. DOI: 10.48550/ARXIV.2208.05868. URL: https://arxiv.org/abs/2208.05868.

Werbos, Paul John (1994). *The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting*. USA: Wiley-Interscience. ISBN: 0471598976.

Widrow, B. and M.A. Lehr (Sept. 1990). "30 years of adaptive neural networks: perceptron, Madaline, and backpropagation". In: *Proceedings of the IEEE* 78.9. Conference Name: Proceedings of the IEEE, pp. 1415–1442. ISSN: 1558-2256. DOI: 10.1109/5.58323.

Winter and Widrow (July 1988). "MADALINE RULE II: a training algorithm for neural networks". In: *IEEE 1988 International Conference on Neural Networks*. IEEE 1988 International Conference on Neural Networks, 401–408 vol.1. DOI: 10.1109/ICNN.1988.23872.

Wolf, Ivo et al. (Dec. 2005). "The medical imaging interaction toolkit". In: *Medical Image Analysis* 9.6, pp. 594–604. ISSN: 1361-8415. DOI: 10.1016/j.media.2005.04.005.

Wu, Xiang, Ran He, and Zhenan Sun (2015). "A Lightened CNN for Deep Face Representation". In: *CoRR* abs/1511.02683. arXiv: 1511.02683. URL: http://arxiv.org/abs/1511.02683.

Xu, Xuanang et al. (Jan. 2019). "Efficient Multiple Organ Localization in CT Image Using 3D Region Proposal Network". In: *IEEE Transactions on Medical Imaging* 38, pp. 1885–1898. DOI: 10.1109/TMI.2019.2894854.

Yin, Xiao-Xia et al. (Apr. 15, 2022). "U-Net-Based Medical Image Segmentation". In: *Journal of Healthcare Engineering* 2022, p. 4189781. ISSN: 2040-2295. DOI: 10.1155/2022/4189781. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9033381/ (visited on 12/28/2022).

Yushkevich, Paul A. et al. (July 1, 2006). "User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability". In: *NeuroImage* 31.3, pp. 1116–1128. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2006.01.015.

Zhao, Zhong-Qiu, De-Shuang Huang, and Bing-Yu Sun (Sept. 1, 2004). "Human face recognition based on multi-features using neural networks committee". In: *Pattern Recognition Letters* 25.12, pp. 1351–1358. ISSN: 0167-8655. DOI: 10.1016/j.patrec.2004.05.008. URL: https://www.sciencedirect.com/science/article/pii/S0167865504001047 (visited on 12/16/2022).

Zhou, Bolei et al. (July 2017). "Scene Parsing through ADE20K Dataset". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE, pp. 5122–5130. ISBN: 978-1-5386-0457-1. DOI: 10.1109/CVPR.2017.544. URL: http://ieeexplore.ieee.org/document/8100027/ (visited on 12/30/2022).

Zhou, Bolei et al. (Oct. 16, 2018a). *Semantic Understanding of Scenes through the ADE20K Dataset*. arXiv: 1608.05442[cs]. URL: http://arxiv.org/abs/1608.05442 (visited on 12/30/2022).

Zhou, Zongwei et al. (2018b). "UNet++: A Nested U-Net Architecture for Medical Image Segmentation". In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Ed. by Danail Stoyanov et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 3–11. ISBN: 978-3-030-00889-5. DOI: 10.1007/978-3-030-00889-5_1.

Zhou, Zongwei et al. (June 2020). "UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation". In: *IEEE Transactions on Medical Imaging* 39.6. Conference Name: IEEE Transactions on Medical Imaging, pp. 1856–1867. ISSN: 1558-254X. DOI: 10.1109/TMI.2019.2959609.

Zuley, Margarita L. et al. (2016). *The Cancer Genome Atlas Head-Neck Squamous Cell Carcinoma Collection (TCGA-HNSC)*. In collab. with TCIA Team. Version Number: 5 Type: dataset. DOI: 10.7937/K9/TCIA.2016.LXKQ47MS. URL: https://wiki.cancerimagingarchive.net/x/VYG0 (visited on 01/09/2023).

Çiçek, Özgün et al. (2016). "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Ed. by Sebastien Ourselin et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 424–432. ISBN: 978-3-319-46723-8. DOI: 10.1007/978-3-319-46723-8_49.