Imperial College London

INDIVIDUAL PROJECT REPORT

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

Tackling Crohn's disease using deep learning

Author:
Feifan Fan

Supervisor:
Dr. Bernhard Kainz
Second Marker:
Prof. Ben Glocker

Submitted in partial fulfillment of the requirements for the MEng
Mathematics and Computer Science of Imperial College London

Abstract

This is the abstract about this thesis.

Acknowledgements

Thanks mum!

Contents

List of Figures

Chapter 1

Introduction

1.1 Crohn's Disease

Crohn's Disease [1, 2] is one of the primary types of Inflammatory Bowel Disease (IBD) [3], which is characterised by its chronic nature with inflammation in the gastrointestinal (GI) tract, as indicated in Figure 1.1 [4]. This condition can lead to long-term damage and complications, such as strictures, fistulas, and abscesses. Many people worldwide are struggling with IBD, and the management remains a challenge for medical professionals to address. A study from the University of Nottingham [5] reports that more than half a million individuals in the UK are affected by Crohn's Disease and Ulcerative Colitis, another significant IBD subtype. Unlike Ulcerative Colitis, which is limited to the colon and rectum, Crohn's disease can potentially develop lesions anywhere within the GI tract. Consequently, patients may experience diverse symptoms, including abdominal pain, diarrhoea, fatigue, and weight loss.

Although numerous research initiatives have been undertaken [6, 7], the precise aetiology of Crohn's disease remains elusive, rendering it incurable. However, fortunately, early diagnosis and appropriate treatment can alleviate patients' symptoms and substantially improve their quality of life. Various diagnostic methods are employed by clinicians, such as enteroclysis, endoscopy, colonoscopy, and radiographic techniques (including barium contrast X-rays, Computed Tomography (CT), and Magnetic Resonance Imaging (MRI)) to assist in the early diagnosis of the disease. MRI has become increasingly popular among radiographic techniques due to its non-invasive nature and enhanced imaging capabilities compared to CT. Nevertheless, manual MRI scan analysis remains a challenge since it is a time-consuming and labour-intensive process. Addition-
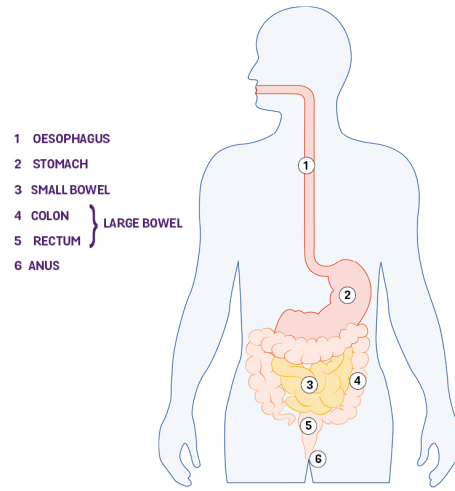
Figure 1.1: The gastrointestinal tract of a Human

ally, medical experts must examine each scan slice by slice painstakingly.

1.2 Motivation

The advancement of Machine Learning and Deep Learning technologies, notably Convolutional Neural Networks (CNNs), offers powerful means for automatic feature extraction from input imaging data, thereby supporting medical professionals in diagnostic tasks. One crucial aspect of Crohn's Disease diagnosis is the examination of the terminal ileum. Holland et al.'s study [8] in 2019 proposed a residual network specifically targeting the terminal ileum to facilitate automated detection of Crohn's Disease using MRI scans. The authors claimed that the efficacy of their framework was contingent upon the degree of localisation during the preprocessing stage. Consequently, they advocated for incorporating terminal ileal ground-truth segmentations to enhance the localisation of the terminal ileum and improve the performance of automated detection techniques.

In a subsequent study, Abidi et al. [9] advanced this line of research by developing an innovative deep-learning tool based on the nnU-Net architecture [10]. This approach enabled the automatic localisation of critical regions, particularly the terminal ileum, essential for radiologists during diagnostic assessments. The researchers addressed the previously identified limitations regarding the high dependence on localisation during the preprocessing phase [8]. Furthermore, their findings established a solid

foundation for a multi-class terminal ileum segmentation algorithm that combines transfer learning strategies with the nnU-Net architecture.

Inspired by the insights from [8, 9], this project aims to create a binary segmentation model capable of accentuating terminal ileum regions more accurately by incorporating advanced transfer and semi-supervised learning methodologies. The successful accomplishment of this objective will significantly aid clinicians in diagnosing Crohn's disease, ultimately contributing to enhanced patient outcomes.

1.3 Machine Learning Challenges

The efficacy of deep learning models is intrinsically linked to the training data's quality, quantity, and diversity. One of the principal challenges faced in this project is the limited availability of training data. The dataset at our disposal is relatively small, comprising only 233 patient cases, which pales compared to those utilised in other industry-leading deep learning systems. Furthermore, since the region of interest (ROI) occupies a minor portion of the MRI scan, additional preprocessing techniques, such as localisation and cropping, must be considered for enhancing the segmentation model's performance, as suggested by [9]. Another major challenge is the necessity for manual segmentation by clinical experts to develop gold-standard labels or point-wise centerlines for patient data, which is a complex, laborious, and inefficient endeavour. Consequently, acquiring high-quality and abundant patient data and gold-standard annotations poses significant challenges.

To mitigate these concerns, we propose a proxy training task employing weak supervision to generate coarse-grained segmentation masks as a compromise for the scarcity of gold-standard segmentations. Upon completion of the proxy task, gold-standard segmentations will be integrated into the training process to produce the final segmentation model. However, prior research [9] indicates that training from scratch using the nnU-Net framework for proxy tasks is inefficient due to long convergence times and unstable performance. Implementing transfer learning for proxy training [11] or incorporating a related pre-training job may serve as potential solutions to address these limitations.

1.4 Objectives

The primary goal of this project is to leverage deep learning methodologies, building upon previous research, to devise an advanced segmentation model for the terminal ileum, employing sophisticated transfer and semi-supervised learning strategies. The objectives of this project include:

- Implementing a nnU-Net-based baseline segmentation model, incorporating relevant data preprocessing techniques to ensure optimal performance.

- Generating coarse-grained weak masks for data lacking ground-truth annotations, based on prior work, to establish the proxy training task.

- Training the proxy model using weak masks and utilising the proxy model and fully annotated data to develop the target segmentation model.

- Enhancing the weak mask generation step by integrating the **SegmentAnything** Model [12] from Meta AI and training a refined model accordingly.

- Evaluating the performance of the developed models quantitatively by comparing the Dice Similarity Coefficient (DSC), training efficiency, and generalisation gap.

Chapter 2

Related Work

2.1 Automatic Detection and Segmentation of Crohn's Disease Tissues from Abdominal MRI

The origins of applying deep learning techniques to medical image analysis can be traced back to approximately a decade ago. In 2013, Mahapara et al. [13] pioneered a machine learning-based method for segmenting bowel regions to detect Crohn's disease tissues in MRI scans.

The proposed pipeline begins with the over-segmentation of the input MR image test volume into supervoxels. Sequentially, Random Forest (RF) classifiers are employed to identify supervoxels containing diseased tissues, subsequently defining the Volume of Interest (VOI). Within the VOI, voxels are further examined to segment the affected region. An additional set of RF classifiers is applied to the test volume to generate a probability map, which delineates the likelihood of each voxel being classified as diseased tissue, normal tissue, or background.

Empirical results demonstrate that this approach achieved satisfactory segmentation performance, as evidenced by a Dice metric value of 0.90 ± 0.04 and a Hausdorff distance of 7.3 ± 0.8 mm. The significance of this research lies in the development of an automated pipeline for segmenting diseased bowel sections in abdominal MR images. This pipeline assists medical experts in identifying affected tissues, thereby facilitating the diagnosis and treatment of Crohn's Disease. The clinical validation of the results, showcasing high segmentation accuracy, further underscores its utility in supporting medical professionals in their work.

Nevertheless, the method is limited by computational inefficiency and complexity due to extended testing times for each instance, fine-tuning requirements for each pipeline stage, and ample opportunities for archi-

tectural improvements. Moreover, the research does not delve into finer details, such as the terminal ileum, which is particularly crucial for comprehensive analysis and early diagnosis.

2.2 Automatic Detection of Bowel Disease with Residual Networks

Building on several years of research in the field, Holland et al. [8] put forth a pioneering approach in 2019 to automate the detection of Crohn's disease from a limited dataset of MRI scans. The authors employed an end-to-end residual network [14], equipped by a soft attention layer [15]. This layer essentially magnified salient local features and added a layer of interpretability, providing a clearer understanding of the analytical process.

In a strategic departure from semantic segmentation strategies typically employed, their approach exclusively targets the terminal ileum. This focus served to underscore the potential feasibility of deep learning algorithms for the precise identification of terminal ileum Crohn's Disease within abdominal MRI scans.

The method's robustness is reflected in its experimental results. Under conditions of localized data within a semi-automatic setting, the model achieved a commendable weighted-f1 score of 0.83. This score is particularly noteworthy given its close correlation with the MaRIA [16] score, a clinical standard that enjoys widespread acceptance in the medical community. Beyond its performance metrics, the researchers accentuated the relative efficiency of their model, which necessitated only a fraction of the preparation and inference time compared to standard procedures. This aspect underlines the potential for significant time-saving benefits in a clinical context.

However, the research did reveal certain limitations. Notably, when applied in a fully automatic setting, the model's performance exhibited a marginal decrease in efficiency. Although this does not detract from the overall achievements of the study, it does highlight an area where further refinement and improvement could be pursued.

Reiterating their discoveries, Holland et al. proposed a strong correlation between model performance and the degree of localisation in the training data. They suggested the collection of gold-standard segmentation of the terminal ileum could prove beneficial as an antecedent task in efforts to enhance automatic detection performance. This proposition

opens up intriguing possibilities for research, including the work presented in this thesis, which explores these aspects in greater detail.

Their insights illuminate the synergistic potential between manual analysis and automated methods in enhancing diagnostic capabilities. Importantly, they establish a pathway for integrating deep learning techniques to detect Crohn's disease from limited datasets, indicating a promising approach to tackle one of the significant challenges in machine learning: data scarcity. By leveraging soft attention mechanisms to intensify salient local features and augment interpretability, they provide a valuable tool for medical professionals to comprehend better the results generated by the algorithm.

These findings, particularly the proposed use of gold-standard segmentation of the terminal ileum, provide a solid foundation for the work pursued in this thesis.

2.3 Leveraging Machine Learning Methods for Accurate Prediction of Intestinal Damage in Crohn's Disease Patients

In 2020, Enchakalody et al. [17] embarked on an innovative study exploring the potential of machine learning methodologies to enhance the precision and reliability of diagnosing and monitoring Crohn's Disease. They applied these techniques to a small dataset of 207 CT-Enterography (CTE) scans, an approach that mirrors our own research focus. Their comprehensive analysis involved the intricate examination of cross-sectional views of small intestine segments and detailed detection of diseased tissues. Utilising two distinct classifier types - Random Forest (RF) with ensemble techniques and Convolutional Neural Network (CNN) algorithms, they quantitatively evaluated intestinal damage related to Crohn's Disease on each mini-segments.

The efficacy of both RF and CNN techniques was compellingly demonstrated in the experimental results, achieving accuracy rates of 96.3% and 90.7%, respectively, for classifying diseased and normal segments. Remarkably, these techniques mirrored the effectiveness of expert radiologists in distinguishing between diseased and normal small bowel tissue. This underscores the immense potential of machine learning, even when applied to small datasets, in elevating the precision of Crohn's Disease diagnoses.

The research conducted by Enchakalody et al. is particularly insight-

ful for our work. It not only demonstrates the successful application of deep learning techniques on small datasets but also opens the door to potentially revolutionising the diagnosis and treatment of Crohn's disease through machine learning. It highlights the possibility of a more precise and automated approach to detecting intestinal damage in such patients, a focus that aligns closely with our current research aims.

While it should be noted that this study primarily focused on data derived from CT-Enterography, differing slightly from our focus on MRI data, the methodology and findings offer valuable insights. As of the time of writing this report, despite the progress made, achieving a fully automatic approach for diagnosing Crohn's disease based on cross-sectional imaging that equates to the proficiency of expert radiologists continues to be an exciting area of ongoing research.

Chapter 3

Background

3.1 Delving into Image Segmentation

Image Segmentation is a fundamental concept in the realm of computer vision, which demonstrates how machines perceive and understand image data. But what exactly does the term "Image Segmentation" encapsulate? To explain this, let's take an illustrative example. Imagine a picture featuring two birds. The task of image segmentation, or more specifically, **"semantic segmentation"**, involves dissecting the entire image into distinct regions that are assigned different colour codes. These regions delineate the exact position of the birds within the image, effectively separating them from the background.

3.1.1 Expanding on Semantic Segmentation

Semantic segmentation transcends the act of partitioning an image into various regions. It assigns each segmented region with a label, i.e. a **semantic meaning**, indicating what the region denotes. An illustrative example is provided in Figure 3.1a, where the red region denotes the liver in a medical imaging scan. Returning to our bird image, we would assign these regions the label "bird" upon segmenting the regions corresponding to the birds. Similarly, the rest of the image, or the background region, would be labelled as "background".

One worth mentioning is that semantic segmentation distinguishes itself from other tasks that merely cluster images into coherent regions, as shown in Figure 3.1 [18]. The regions identified through semantic segmentation carry a specific value or meaning inherently linked to the task at hand. Simply put, not just 'where' but 'what' is just as crucial in semantic

segmentation. Having laid out an overview of **Semantic Segmentation**,

(a) Semantic Segementation   (b) Segmentation based on clustering
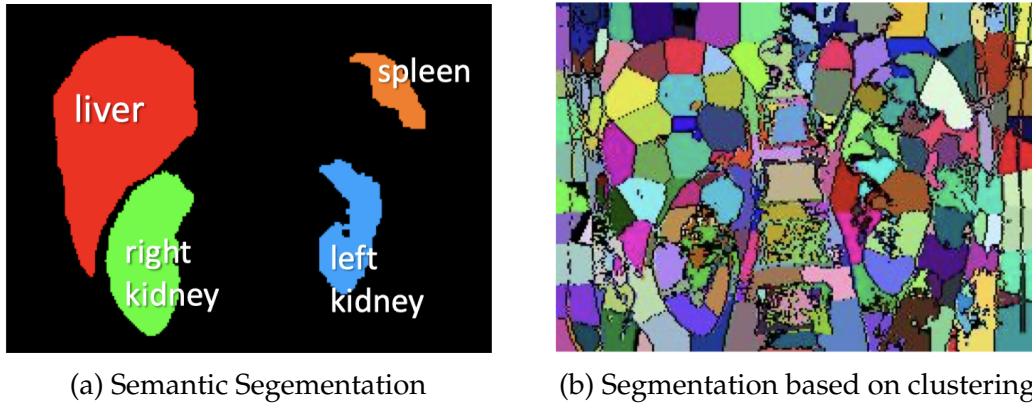
Figure 3.1: Different Image Segmentation Tasks

we will delve deeper into the distinctive methodologies utilised for segmentation tasks in the ensuing sections. This preliminary understanding provides a critical foundation for the exploration of more complex segmentation strategies and their applications in tasks such as diagnosing diseases or enhancing medical imaging methods.

3.2 Manual Segmentation

Manual Segmentation, while basic, forms a cornerstone in the realm of image segmentation. In the context of our project, it involves utilising the specialised knowledge of medical experts to meticulously create 'gold standard' labels for MRI data. This process is not simply a cursory glance at the image; it requires careful slice-by-slice examination of the MRI data, highlighting the diseased from the healthy tissues. This level of detailed inspection ensures that such manual segmentation results in validated and trustworthy classifications.

However, manual segmentation does bring to light significant challenges. Notably, the time-intensive nature of this process becomes increasingly pronounced when dealing with large datasets. This situation is further compounded in scenarios requiring slice-by-slice analysis of MRI data. The resultant effect is that manual segmentation can be exceptionally time-consuming, limiting the volume of data that can be labelled within an acceptable timeframe.

This leads us to confront an ever-present challenge: the scarcity of gold standard labelled data. The considerable investment of time and skilled

resources required for manual segmentation naturally restricts the availability of such high-quality, expert-classified data sets. As we journey towards leveraging machine learning methodologies in the realm of medical imaging, this scarcity of gold-standard, manually segmented data surfaces as a substantial hurdle to overcome. Thus, innovative solutions are required to address this gap and enhance the effectiveness and efficiency of image segmentation processes.

Additionally, manual segmentation may introduce variability. This includes both inter-observer variability, where there could be disagreement between different human observers, and intra-observer variability, where inconsistencies could arise from the same observer at different occasions. This variability might potentially compromise the validity of the manually segmented data.

The process of manual segmentation may also unintentionally incorporate biases from human experts, which could potentially distort the results. The limitations inherent in manual approaches underscore the necessity for the integration of computational tools to bolster the segmentation process. By leveraging computer-assisted methodologies, we can introduce an intelligent strategy to carry out segmentation tasks. This approach not only enhances efficiency but also promotes consistency, thereby effectively minimising human error and variability. As a consequence, we can significantly enhance the overall reliability and accuracy of MRI data analysis. This confluence of human expertise and computational intelligence stands to advance the field of medical imaging analysis.

3.3   Threshold-based Methods

Thresholding is among the most rudimentary yet widely adopted methods for image segmentation in various industry applications. The premise of this technique rests on the hypothesis that the distribution of pixel intensity in an image contains multiple modes. In other words, the grayscale intensity of pixels can be differentiated into two (or more) distinct clusters. The strategy involves identifying thresholds within the 'gaps' separating these clusters to delineate the background from the foreground. For accomplishing this segmentation, pixels exhibiting an intensity lower than the threshold are designated as background, whilst those with equal to or higher than the threshold are identified as foreground. Furthermore, it is also feasible to define multiple thresholds, thereby enabling multi-class segmentation of pixels that fall within a specific intensity range only.

The inherent simplicity and speed of this algorithm is undeniably ad-

vantageous. However, its applicability is contingent on the homogeneity and distinctness of the regions of interest (ROIs), implying that pixels within the same region should exhibit similar intensity values. An additional challenge lies in identifying consistent threshold values across different images. This is due to the fact that pixel intensities are prone to variations between different images, thus introducing a degree of complexity in maintaining uniform thresholding norms across disparate datasets.

3.4 Region-based Methods

Region-based segmentation techniques, with the region-growing method standing as a prominent example, offer an alternative approach to image segmentation. This method hinges on the premise of homogeneity within the segmented regions. The algorithm initiates with a seed pixel and then expands the region by successively incorporating pixels similar to the initial seed. The process continues until the region growth reaches a predefined size or once the region achieves homogeneity - interpreted as the point when neighbouring pixels become significantly dissimilar [19].

Utilisation of the region-growing method often proves efficient, generating a connected region starting from the seed point. Unlike thresholding methods that rely on explicit image properties, region-growing methods facilitate segmentation based on pixel similarity. However, this method exhibits sensitivity to noise, as the algorithm may persist in growing the region even if the neighbouring pixels significantly deviate from the seed pixel's properties.

Another key challenge when employing the region-growing method is the critical significance of the initial seed point selection. An incorrect choice of the seed point could thwart proper region growth. This initial phase often becomes time-consuming and lacks accuracy when seeking the optimal seed point. Furthermore, it necessitates human intervention for evaluating the appropriateness of the chosen seed point.

To Address these concerns in the context of medical imaging, a study contucted by Poonguzhali et al. proposed an automated region-growing method tailored for ultrasound images [20]. Their approach introduces an automatic seed point selection mechanism predicated on textural features, such as co-occurrence and run-length features, thereby eliminating the need for manual intervention. The experimental results attested to the feasibility and efficacy of their proposed region-growing algorithm, showcasing its capacity to select seed points and segment the ROIs without requiring manual intervention.

15

3.5 Deep Learning Methods

3.5.1 U-Net: An Automated Deep Learning Method for Image Segmentation

U-Net [21] represents a transformative approach in the realm of deep learning methodologies for image segmentation, particularly aligned with our project's objectives. As an evolved variant of Fully Convolutional Networks (FCN) [22], U-Net integrates a high-level contextual extraction path with a symmetric localisation pathway. This unique arrangement both captures a broad context of the image and enables precise localisation. This is ideal for creating an automated pipeline for TI segmentations, and an essential requirement in our research.

The end-to-end learning facilitated by U-Net directly generates pixel-wise segmentation masks from raw pixels, creating a valuable asset for our project. The architecture of this robust network is illustrated in Figure 3.2.
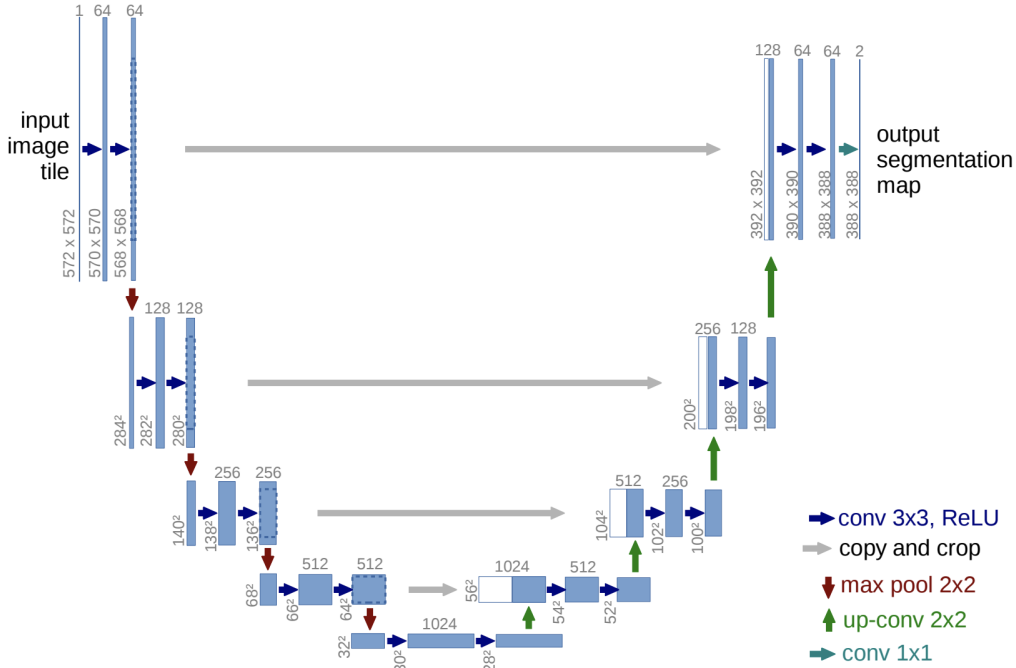

Figure 3.2: An example of the U-net architecture.

Significantly, U-net has also been developed to handle 3D data, which aligns perfectly with our scenario, where the MRI scans are inherently 3-dimensional images. Furthermore, each blue box in Figure 3.2 encapsu-

lates a multi-channel feature map. The number of channels is denoted at the top of each box, while the *x-y* dimensions are annotated at the lower left corner of the box. Meanwhile, the white boxes signify sets of copied feature maps, with arrows demonstrating varied operations.

On the contracting (left) side of the network, the design entails a series of recurrent steps. Each step initiates with two 3x3 convolutions, proceeding to double the number of feature maps. The result is then subjected to a Rectified Linear Unit (ReLU) activation function, followed by a downsampling operation using 2x2 max pooling with a stride of 2.

The contracting path of the network (left) contains recurrent steps, each initiating with two 3x3 convolutions that effectively double the feature maps count. Subsequently, a Rectified Linear Unit (ReLU) activation function is applied, followed by a 2x2 max pooling downsampling operation with a stride of 2.

On the expansive side (right) of the network, the process commences with an upsampling step. A 2x2 convolution follows, halving the feature channels, which are then concatenated with their corresponding features from the contracting pathway. To conclude, a pair of 3x3 convolutions are applied to the image, succeeded by ReLU activations. The final layer employs a 1x1 convolution that maps each 64-component feature vector to the desired classes.

Overall, U-Net offers a comprehensive and highly effective tool for achieving granular image segmentation—a capability integral to the success of our research exploration.

3.5.2 nnU-Net: A Self-configuring Framework Aligned with Our Research Goals

The nnU-Net [10] offers a progressive step towards personalised segmentation techniques. As a framework built upon U-Nets, it embodies a self-configuring segmentation mechanism, which autonomously orchestrates the configuration of preprocessing, network architecture, training, and post-processing steps in a segmentation pipeline. Crucially, the configuration selected is not static but instead adapts to the specificities of the medical data used for training.

A standout feature of nnU-Net is its unique approach towards determining hyperparameters. The framework utilises "data-fingerprints" allied with heuristic rules to pinpoint the optimal hyperparameter configuration for a given dataset prior to processing the training data. This data fingerprint concept is further leveraged to generate pipeline configura-

tions, encapsulating both inferred parameters (such as image resampling, normalisation, batch, and patch size) and blueprint parameters (such as loss function, optimiser, and network architecture).

With these pre-selected hyperparameters and generated pipeline configurations, nnU-Net proceeds to facilitate network training for 2D, 3D full-resolution, and 3D-Cascade U-Nets [23]. The platform then select an ensemble of configurations from these three networks to achieve optimal performance (for instance, maximising the average dice coefficient). Once identified, this optimal configuration is subsequently deployed and evaluated on the test dataset.
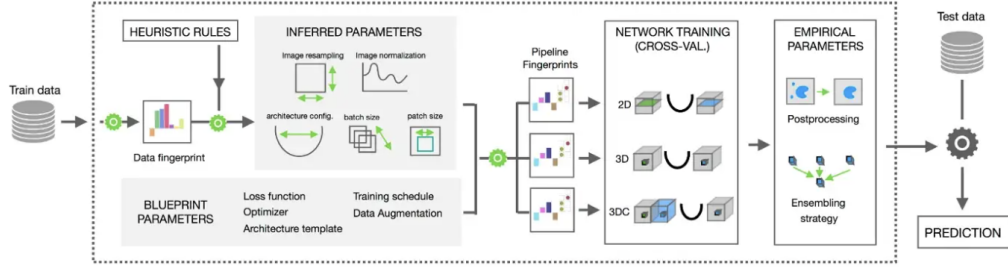

Figure 3.3: The pipeline representation of nnU-Net.

Notably, nnU-Net has been shown to deliver state-of-the-art performance in various medical imaging tasks, including segmenting brain tumours, prostate tissues, and liver structures [24].

However, it is essential to understand that the 'adaptive' nature of nnU-Net does not imply universal applicability. Once the framework is trained, it performs optimally when applied to new data similar to the training data. For instance, if nnU-Net is trained using abdominal MRI images for liver segmentation, it will yield the best performance in similar tasks involving abdominal MRI images.

Therefore, while nnU-Net enables a highly adaptive and automated segmentation workflow, the utility of the trained model is maximised when deployed on tasks that closely resemble the context and content of the training data. In our specific case, we will train the nnU-net model with our specific type of 3D MRI scans for TI segmentation tasks to ensure optimal results. This approach ensures leveraging the full potential of nnU-Net's adaptive capabilities tailored to our research goals.

As our research strives to establish an automated pipeline for TI segmentations on 3D MRI scans, the integration of the nnU-Net framework is poignantly relevant. With its inherent capacity to adapt to different

datasets and consistently yield precise image segmentation, nnU-Net is excellently positioned to enhance the rigour and precision of our research endeavour.

3.6 Generating Weak Labels: A Strategic Response to Annotation Scarcity

Given the scaricity of gold standard annotations for segmentation tasks, as discussed in section 3.2, it becomes imperative to identify effective strategies that can enhance our training outcomes. One potential approach is the use of weak labels or weakly-segmented masks derived from the data. By adopting such methods, we aim to extract critical image features that can contribute to the learning efficacy of our model, despite limited access to manually-annotated, gold standard data.

3.6.1 Harnessing Pretrained Models: The Segment Anything Model (SAM)

A notable unsupervised method we contemplate employing is the Simple Linear Iterative Clustering (SLIC) [25]. This algorithm stands at the forefront of superpixel segmentation techniques, providing a powerful tool to cluster pixels within an image into compact and uniformly labelled regions, referred to as 'superpixels'.

The key strengths of SLIC are its simplicity, ease of implementation, and adaptability across diverse scenarios. These virtues enable SLIC to effectively handle boundary adherence issues while simultaneously reducing the computational burden associated with image segmentation tasks. Essentially, SLIC operates as a spatially constrained iterative k-means clustering method with a pre-determined superpixel size m. Consequently, in the generated weak mask, each superpixel generated incorporates a cluster center and m pixels. The SLIC algorithm adapted from [25] is illustrated in Algorithm 1:

Algorithm 1 SLIC superpixel segmentation

1: Initialise cluster centers $C_k = [l_k, a_k, b_k, x_k, y_k]^\top$ by sampling pixels at regular grid with grid intervel S.
2: Perturb cluster centers in a $n \times n$ neighbourhood, to the loweset gradient position.
3: **while** $E >$ threshold **do**
4:     **for** each cluster center C_k **do**
5:         Assign the best matching pixels from a $2S \times 2S$ square neighbourhood around the cluster according to the distance measure [25].
6:     **end for**
7:     Compute new cluster centers and residual error E.
8: **end while**
9: Enforce Connectivity.

Note the term "Perturb" in line 2 essentially means to re-assign the cluster center in a $n \times n$ neighbourhood, where the re-assigned center is the point has lowest gradient position compare to the original cluster. Additionally, the residual error E is defined as the L_1 distance between previous centers and recomputed centers.

By integrating SLIC as part of our weak label generation strategy, we aim to utilise these inherent benefits to enhance the robustness of our machine learning models, thereby bringing rich insights from our medical imaging data in the absence of extensive manually-curated annotations.

3.6.2 Harnessing Pretrained Models: The Segment Anything Model (SAM)

Emerging from the innovative approaches of Meta AI, the Segment Anything (SA) project presents a new paradigm for image segmentation. The cornerstone of this project is a pioneering model, known as SAM, which exhibits an impressive degree of adaptability and transferability. Specifically designed and trained to be promptable, SAM showcases remarkable capabilities in zero-shot transfer to new image distributions and tasks.

This inherent agility of SAM instigates impressive zero-shot performance outcomes. Indeed, comparative analysis with prior fully supervised results reveals that SAM often delivers competitive, if not superior, performance metrics [12].

Extending this versatile approach to medical imaging, MedSAM of-

fers compelling prospects. Derived from the parent SAM model, Med-SAM has shown considerable promise in segmenting medical images. One study unearthed that SAM's functionality can be enhanced with manual prompts, such as points and boxes, indicating intended objects in medical images [26]. This enhancement strategy aligns perfectly with our proposed approach to initially employ SLIC for coarse segmentation, and subsequently refine the segmentation process by obtaining finer-grained weak masks from SAM.

However, it's worth noting that whilst SAM exhibits remarkable performance with certain objects and modalities, it may fall short or even fail in other scenarios. This observation underscores the importance of particularising the model's utility in context-specific applications, a perspective central to our research objective on TI segmentations with 3D MRI scans. With its inspiring capabilities, SAM holds immense potential to enhance our quest for automated and precision-driven segmentation techniques, thereby bolstering our research outcomes.

Chapter 4

Ethical Discussion

Given the sensitive nature of our project that involves the handling of medical data, we are steadfastly committed to ensuring a robust ethical framework guides all phases of our work. This initiative encompasses further processing and augmentation of previously collected medical data, as well as merging existing datasets.

To safeguard personal identity and ensure strict adherence to privacy standards, all MRI scans are carefully processed under the oversight of clinical radiologists at St Mark Hospital. Comprehensive measures are employed to remove any personally identifiable data, such as names, genders, ages, and ID numbers, from the medical records. As a result, it is impossible to trace back any individual's identity from the processed medical data.

Our commitment to ethical considerations extends into the model training procedure. We employ the nnU-Net framework, which utilises convolutional layers to learn from data via feature extraction. This learning procedure does not store original training data; instead, it creates feature maps that represent distilled, valuable insights from the data. This process ensures that the original training data cannot be recovered from the model, thereby preserving individual privacy.

In essence, our project stands on a foundational ethical commitment that respects personal anonymity and data confidentiality. Recognising the crucial importance of trust in scientific inquiry, particularly when dealing with sensitive medical data, we are dedicated to exemplifying conscientious practices that uphold the highest standards of research ethics.

Chapter 5

Evaluation Plan

Ensuring the effectiveness of our approach in developing an advanced segmentation model for the terminal ileum necessitates formulating a rigorous evaluation plan. This plan outlines systematic stages and metrics to assess our system's performance and alignment with our project objectives:

1. **Baseline Model Evaluation**: The initial step involves establishing and evaluating the performance of the nnU-Net-based segmentation model. We conduct baseline tests using annotated data to measure its initial performance, specifically focusing on Dice Similarity Coefficient (DSC) as an indicative metric. These results will serve as our benchmark against which to evaluate further improvements.

2. **Generation and Inspection of Coarse-grained Weak Masks**: Leveraging the Simple Linear Iterative Clustering (SLIC) method, we generate coarse-grained weak masks for data lacking ground-truth annotations. We will evaluate the produced masks, both qualitatively and quantitatively, to gauge the quality and adequacy of these initial segmentation outputs.

3. **Proxy Model Training and Evaluation**: Following the production of coarse-grained weak masks, we employ them to train the proxy model. We evaluate its performance based on a series of parameters: training stability, learning rate, convergence speed, and DSC score, among others. The results obtained provide insight into the efficacy of using SLIC-created masks in training models.

4. **Integration and Evaluation of SegmentAnything Model**: To refine the generated weak masks, we integrate the SegmentAnything Model

from Meta AI into our pipeline, with additional information of the ROI, including centerline coordinates and bounding boxes provided. We will assess the quality of the finer-grained weak masks produced by SAM through visual inspection and statistical measures, comparing them against the coarse-grained masks created by SLIC.

5. **Target Model Training and Evaluation**: Utilizing the refined weak masks and fully annotated data, we develop the target segmentation model. The performance of this model is evaluated based on important parameters such as DSC, training efficiency, and generalization gap.

6. **Overall Performance Evaluation**: In the final stage, we perform a comprehensive analysis to compare our derived models with our original objectives and benchmark model. Ultimately, we aim to identify the most significant performance enhancements, evaluate the value of our model-refinement process, and highlight possible directions for future research.

By implementing this robust evaluation plan, we strive to validate the success of our project through meticulous testing, thereby assuring the effectiveness and accuracy of our terminal ileum segmentation model.

Chapter 6

Dataset Analysis

Chapter 7

Methodology

Chapter 8

Conclusion

Chapter 9

Future Work

Bibliography

[1] Daniel C Baumgart and William J Sandborn. Crohn's disease. *The Lancet*, 380(9853):1590–1605, 2012.

[2] NHS. Crohn's disease - nhs. URL https://www.nhs.uk/conditions/crohns-disease/.

[3] Centers for Disease Control and Prevention. What is inflammatory bowel disease? (ibd) | ibd. URL https://www.cdc.gov/ibd/what-is-IBD.htm.

[4] The human digestive system. https://crohnsandcolitis.org.uk/media/ftsl0iea/digestion-graphic.jpg. (Accessed on 06/13/2023).

[5] University of Nottingham. Rates of crohn's and colitis have been vastly underestimated for decades, says new study. URL https://www.nottingham.ac.uk/news/rates-of-crohns-and-colitis-have-been-vastly-underestimated-for-decades-says-new-stu

[6] Gautier Hoarau, PK Mukherjee, C Gower-Rousseau, C Hager, J Chandra, MA Retuerto, Christel Neut, Séverine Vermeire, J Clemente, Jean-Frederic Colombel, et al. Bacteriome and mycobiome interactions underscore microbial dysbiosis in familial crohn's disease. *MBio*, 7(5):e01250–16, 2016.

[7] Joseph D Feuerstein, Edith Y Ho, Eugenia Shmidt, Harminder Singh, Yngve Falck-Ytter, Shanaz Sultan, Jonathan P Terdiman, Shahnaz Sultan, Benjamin L Cohen, Karen Chachu, et al. Aga clinical practice guidelines on the medical management of moderate to severe luminal and perianal fistulizing crohn's disease. *Gastroenterology*, 160(7): 2496–2508, 2021.

[8] Robert Holland, Uday Patel, Phillip Lung, Elisa Chotzoglou, and Bernhard Kainz. Automatic detection of bowel disease with resid-

ual networks. In *International Workshop on PRedictive Intelligence In MEdicine*, pages 151–159. Springer, 2019.

[9] Ali Abidi. Tackling crohn's disease using deep learning. Master's thesis, Imperial College London, 2022.

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

[11] Jong-Hwan Jang, Tae Young Kim, and Dukyong Yoon. Effectiveness of transfer learning for deep learning-based electrocardiogram analysis. *Healthcare informatics research*, 27(1):19–28, 2021.

[12] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. *arXiv:2304.02643*, 2023.

[13] Dwarikanath Mahapatra, Peter J Schüffler, Jeroen AW Tielbeek, Jesica C Makanyanga, Jaap Stoker, Stuart A Taylor, Franciscus M Vos, and Joachim M Buhmann. Automatic detection and segmentation of crohn's disease tissues from abdominal mri. *IEEE transactions on medical imaging*, 32(12):2332–2347, 2013.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

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

[16] Jordi Rimola, Sonia Rodríguez, Orlando García-Bosch, Ingrid Ordás, Edgar Ayala, Montserrat Aceituno, Maria Pellisé, Carmen Ayuso, Elena Ricart, Lluís Donoso, et al. Magnetic resonance for assessment of disease activity and severity in ileocolonic crohn's disease. *Gut*, 58 (8):1113–1120, 2009.

[17] Binu E Enchakalody, Brianna Henderson, Stewart C Wang, Grace L Su, Ashish P Wasnik, Mahmoud M Al-Hawary, and Ryan W Stidham. Machine learning methods to predict presence of intestine damage

in patients with crohn's disease. In *Medical Imaging 2020: Computer-Aided Diagnosis*, volume 11314, pages 742–753. SPIE, 2020.

[18] Ben Glocker. Image segmentation - machine learning for imaging course, 2023.

[19] Rolf Adams and Leanne Bischof. Seeded region growing. *IEEE Transactions on pattern analysis and machine intelligence*, 16(6):641–647, 1994.

[20] S Poonguzhali and G Ravindran. A complete automatic region growing method for segmentation of masses on ultrasound images. In *2006 International Conference on Biomedical and Pharmaceutical Engineering*, pages 88–92. IEEE, 2006.

[21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[22] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[23] Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*, 2018.

[24] Fabian Isensee, Jens Petersen, Simon AA Kohl, Paul F Jäger, and Klaus H Maier-Hein. nnu-net: Breaking the spell on successful medical image segmentation. *arXiv preprint arXiv:1904.08128*, 1(1-8):2, 2019.

[25] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels. Technical report, 2010.

[26] Yuhao Huang, Xin Yang, Lian Liu, Han Zhou, Ao Chang, Xinrui Zhou, Rusi Chen, Junxuan Yu, Jiongquan Chen, Chaoyu Chen, Haozhe Chi, Xindi Hu, Deng-Ping Fan, Fajin Dong, and Dong Ni. Segment anything model for medical images?, 2023.