

London South Bank University

Division of Computer Science and Informatics,
School of Engineering

**Automatic Segmentation of 3D MRI
Images of the Pituitary Gland For
Medical Image Analysis**

by

Oyinlola A.

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Master of Science
(Applied Artificial Intelligence)

Submitted

13th September, 2024

Abstract

In this study, deep learning's application in automatically segmenting the pituitary gland in 3D MRI brain scans was explored. To automate the process, the SAM-Med3D model, a pre-trained deep learning model, was utilized. A subset of images had manually segmented ground truth data, and preprocessing steps like normalization and resampling were applied. Despite extensive efforts to fine-tune the model, challenges emerged due to its inability to correctly interpret and utilize the provided label masks, resulting in segmentation failures. Several methods were explored to address this issue, but none yielded optimal performance. This underscores the limitations of directly applying pre-trained models to specialized tasks like pituitary gland segmentation. The research indicates that more advanced label preprocessing or further customization of the model is required to overcome these challenges. While the model faced difficulties in accurately identifying the labels, this study offers valuable insights for enhancing deep learning approaches in medical image analysis.

Acknowledgements

First, I would love to thank the Almighty God, the AUTHOR and the FINISHER of my faith.

I would also like to sincerely appreciate the immense work and guidance from my supervisor, **Dr Pawel Markiewicz**.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 1.1 | Background and Context | 1 |
| 1.2 | Problem Statement | 8 |
| 1.3 | Significance of the Study | 9 |
| 1.4 | Literature Review Summary | 10 |
| 1.5 | Research Objectives | 12 |
| 1.6 | Methodology Overview | 13 |
| 1.7 | Scope and Limitations | 14 |
| 1.8 | Structure of the Dissertation | 17 |
| 2 | Literature Review | 18 |
| 2.0.1 | Review of Manual Segmentation techniques | 19 |
| 2.0.2 | Review of Semi-Automatic Segmentation Techniques | 23 |
| 2.0.3 | Review of Automatic Segmentation Techniques | 27 |
| 2.0.4 | Comparative Analysis and Discussion | 30 |
| 3 | Alternative Design and Final Algorithm | 33 |
| 3.0.1 | Introduction to the Problem of Pituitary Gland Segmentation | 33 |
| 3.0.2 | Proposed Solution | 35 |

| | | |
|----------|--|-----------|
| 4 | Implementation | 43 |
| 4.0.1 | Data Acquisition and Pre-processing | 44 |
| 4.0.2 | Preprocessing | 44 |
| 4.0.3 | Selection of Software and Tools | 46 |
| 4.0.4 | Hardware Setup | 49 |
| 4.0.5 | Overview of the Model Architecture | 50 |
| 5 | Experimental and Theoretical Results | 52 |
| 5.1 | Results, Testing and Evaluation | 52 |
| 5.1.1 | Test Procedures and Preprocessing | 53 |
| 5.1.2 | Evaluation Using Dice Similarity Coefficient | 54 |
| 5.1.3 | Experimentation with Larger Label Volumes | 55 |
| 5.1.4 | Discussion on Model Limitations | 56 |
| 6 | Future Work | 58 |
| 6.0.1 | Training from Scratch with Annotated Data | 58 |
| 6.0.2 | Advantages of Custom Training | 60 |
| 6.0.3 | Incorporating Other Imaging Modalities | 60 |
| 6.0.4 | Addressing Computational Challenges | 62 |
| 7 | Conclusion and Reflection on Learning | 63 |
| A | Source code for Dissertation | 67 |

List of Figures

| | | |
|-----|---|----|
| 1.1 | Anatomy of the Pituitary Gland | 2 |
| 1.2 | Hormonal Functions of the Pituitary Gland | 3 |
| 1.3 | Common challenges in MRI images of the brain | 6 |
| 1.4 | Segmentation results on a sagittal slice (top row) and an axial slice (bottom row), showing the manual segmentation in the middle images, highlighted in yellow | 10 |
| 1.5 | Methodology flowchart | 14 |
| 2.1 | Comparison between manual and semi-automatic segmentation | 23 |
| 2.2 | Flowchart of Semi-Automatic Segmentation Methods | 24 |
| 2.3 | Diagram of U-Net Architecture for Automatic Segmentation | 29 |
| 2.4 | Comparison of Efficiency and Accuracy of Different Segmentation Techniques | 32 |
| 3.1 | Comparison between a CT scan and an MRI of the brain | 34 |
| 4.1 | Resampling NIFTI image in python | 45 |
| 4.2 | Preprocessing Workflow from Raw MRI Data to Preprocessed Data | 46 |
| 4.3 | Cloning SAM-Med3D | 47 |
| 4.4 | Pituitary Segmentation in 3D Slicer | 48 |
| 4.5 | Model Architecture | 50 |

| | | |
|-----|---|----|
| 5.1 | Error message after running inference | 53 |
| 5.2 | Brain image and the pituitary gland label mask | 54 |
| 5.3 | Predicted mask in 3D Slicer of the kidney | 56 |
| 6.1 | Flowchart of a Multi-modal Deep Learning Approach for Pituitary Gland Segmentation | 61 |

List of Tables

| | | |
|-----|--|----|
| 1.1 | Comparison of Imaging Modalities for Pituitary Gland Assessment . . | 5 |
| 1.2 | Comparison of Deep Learning Models for Medical Image Segmentation | 11 |
| 1.3 | Comparison of Pre-trained SAM-Med3D vs. Custom-trained Model for Pituitary Gland Segmentation | 16 |
| 3.1 | Pituitary Gland Segmentation Workflow | 35 |
| 3.2 | Evaluation Metrics for Segmentation Performance | 42 |
| 4.1 | Summary of the Data Conversion Process | 44 |
| 4.2 | Tools and Libraries Used in Implementation | 49 |
| 5.1 | Segmentation Performance of SAM-Med3D | 56 |
| 6.1 | Comparison between Pre-trained Model vs. Custom-trained Model . . | 59 |
| 6.2 | Comparison between Pre-trained Model vs. Custom-trained Model . . | 62 |

Chapter 1

Introduction

1.1 Background and Context

The pituitary gland, often referred to as the "master gland," is a vital component of the endocrine system, essential for maintaining various physiological processes within the human body (Barkhoudarian and Kelly 2017). Situated at the base of the brain within a bony structure known as the sella turcica, the pituitary gland is remarkably small, roughly the size of a pea (Lakanwal 2023). Despite its diminutive size, its role in regulating numerous hormonal functions makes it critical for overall health. Figure 1.1 on the next page from (University 2023) shows the anatomy of the pituitary gland.

The pituitary gland consists of two main lobes: the anterior lobe (adenohypophysis) and the posterior lobe (neurohypophysis). Each lobe produces and releases a distinct set of hormones, which are crucial for various bodily functions:

1. Anterior Pituitary Lobe: This lobe produces several key hormones, including:
 - Growth Hormone (GH): Stimulates growth, cell reproduction, and cell regeneration. GH is crucial during childhood for normal growth and is

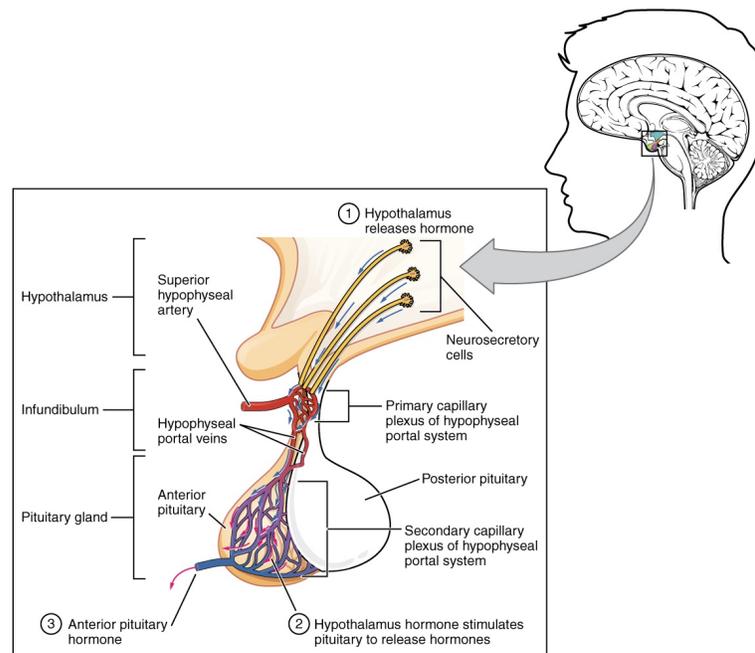


Figure 1.1: Anatomy of the Pituitary Gland

involved in regulating metabolism throughout life (Soliman et al. 2024).

- Adrenocorticotropic Hormone (ACTH): Stimulates the adrenal glands to produce cortisol, a hormone that helps the body respond to stress and regulates metabolism (Weinberg 2024).
- Thyroid-Stimulating Hormone (TSH): Regulates the thyroid gland's production of thyroid hormones, which are essential for metabolism and energy regulation (Zwahlen et al. 2024).
- Gonadotropins (LH and FSH): Luteinizing hormone (LH) and follicle-stimulating hormone (FSH) are critical for reproductive health, regulating menstrual cycles, ovulation, and sperm production (Athar, Karmani, and Templeman 2024).

2. Posterior Pituitary Lobe: This lobe does not produce hormones but stores and releases two important hormones:

- Oxytocin: Plays a role in childbirth and lactation by stimulating uterine contractions and milk ejection. It is also involved in social bonding and emotional regulation (Nogueira-Vale 2024).
- Vasopressin (Antidiuretic Hormone, ADH): Regulates water balance in the body by influencing kidney function and blood pressure (Calvi et al. 2024).

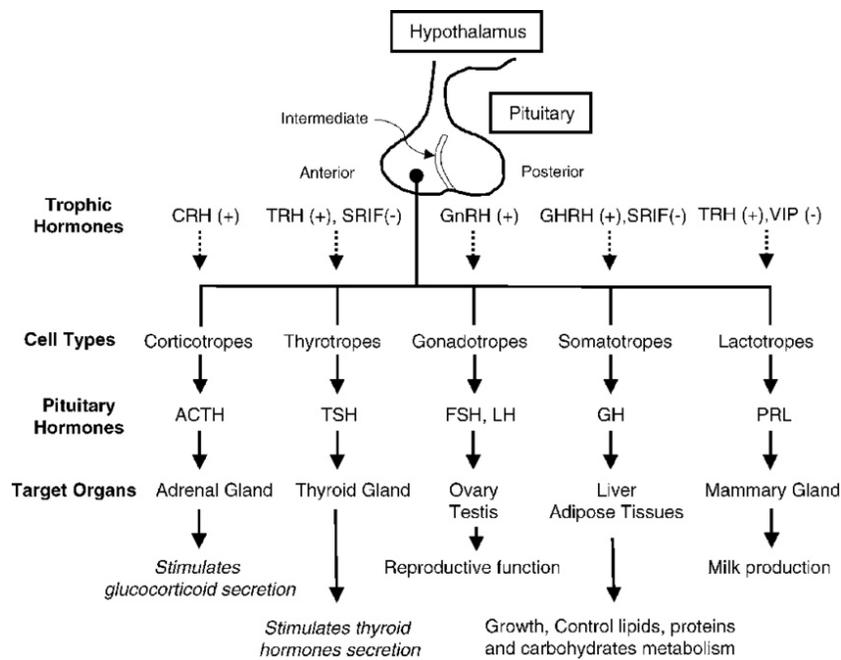


Figure 1.2: Hormonal Functions of the Pituitary Gland

Figure 1.2 from (ResearchGate 2024a) helps us understand the hormonal functions of the pituitary gland. Given its central role in regulating vital physiological functions, any dysfunction or abnormality in the pituitary gland can lead to a range of significant health issues. Common conditions associated with pituitary gland dysfunction include:

1. Pituitary Tumors: Abnormal growths in the pituitary gland that can lead to excessive hormone production (e.g., acromegaly from GH-secreting tumours)

or decreased hormone production (e.g., hypopituitarism (Reddy, Goparaju, and Fleseriu 2024)).

2. Hypopituitarism: A condition characterized by reduced hormone production from the pituitary gland, leading to various symptoms depending on the affected hormones (Corsello, Paragliola, and Salvatori 2024).
3. Acromegaly: Caused by excess GH, leading to abnormal growth of bones and tissues, often due to pituitary tumours (Al-Hadlaq and Sroussi 2024).
4. Cushing's Syndrome: Resulting from excessive cortisol production, often due to ACTH-secreting pituitary tumours, leading to weight gain, high blood pressure, and other health issues (Fion, Saieehwaran, and Subashini 2024).

The Role of Magnetic Resonance Imaging (MRI) in Pituitary Gland Assessment

Magnetic Resonance Imaging (MRI) is widely regarded as the gold standard for visualizing the pituitary gland and diagnosing related disorders (Dongre 2023, Lv 2024). MRI offers several advantages over other imaging modalities, particularly for detailed anatomical and pathological assessment of soft tissues:

1. High Spatial Resolution: MRI provides excellent spatial resolution, allowing for detailed visualization of the pituitary gland's small and complex structure (Jipa and Jain 2021). This high resolution is crucial for detecting subtle abnormalities and for accurate localization of pituitary lesions.
2. Superior Soft Tissue Contrast: Unlike CT scans, which use ionizing radiation, MRI utilizes magnetic fields and radio waves to generate images (Ma 2024). This approach offers superior contrast between different soft tissues, making it

particularly effective for differentiating the pituitary gland from surrounding structures and identifying pathological changes.

Table 1.1: Comparison of Imaging Modalities for Pituitary Gland Assessment

| Modality | Advantages | Limitations |
|-----------------|---|---|
| MRI | High soft tissue contrast, excellent spatial resolution, and non-ionizing radiation are ideal for detecting microadenomas and evaluating gland morphology. | Sensitive to motion artefacts, longer scan time, high cost, not suitable for patients with certain implants. |
| CT | Faster scan time, good for assessing bony structures and calcifications, widely available and less expensive than MRI. | Limited soft tissue contrast compared to MRI, involves ionizing radiation, less effective in detecting small pituitary lesions. |
| PET | Provides metabolic and functional information, is useful in detecting the metabolic activity of tumours, and can be combined with CT or MRI for anatomical correlation. | Lower spatial resolution for soft tissue structures, high radiation dose, expensive, limited availability, potential artefacts. |

3. No Ionizing Radiation: MRI does not expose patients to ionizing radiation, making it a safer choice for repeated imaging, especially in patients who require long-term monitoring (Mittendorff, Young, and Sim 2022).

Despite these advantages, imaging the pituitary gland presents several challenges:

- **Small Size and Variable Morphology:** The pituitary gland’s small size and variability in shape and location can complicate its imaging (Ray 2020). Subtle changes or abnormalities may be difficult to detect and require precise imaging techniques and interpretation.

- **Artifacts and Noise:** MRI images can be affected by various artefacts and noise, which may obscure or distort the pituitary gland’s features (Chrysikopoulos et al. 2020). Common artefacts include susceptibility artefacts from surrounding tissues and motion artefacts from patient movement. (Galldiks et al. 2024, Beheshti et al. 2020) in their papers both compared different imaging modalities for assessing pituitary lesions shown in Table 1.1.
- **Resolution Limitations:** While MRI offers high spatial resolution, the resolution may still be insufficient for visualizing microscopic lesions or for distinguishing between closely spaced anatomical structures (Weiskopf et al. 2021). Figure 1.3 from (ResearchGate 2024b) shows the common challenges in MRI scans.

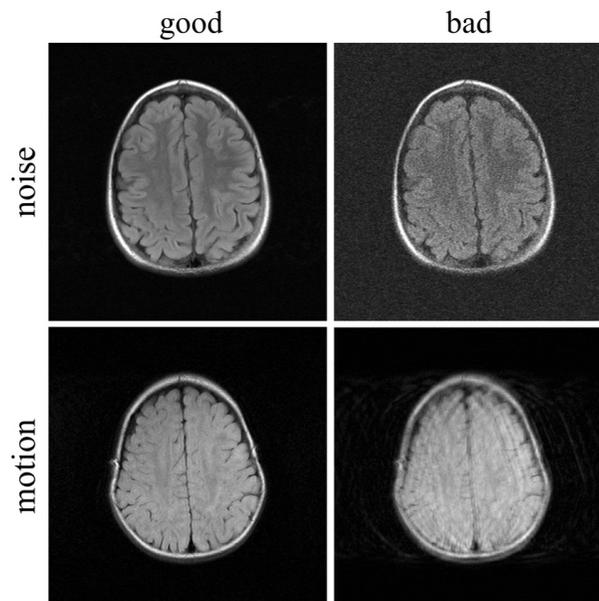


Figure 1.3: Common challenges in MRI images of the brain

To address these challenges, precise and reliable segmentation methods are essential. Accurate segmentation enables clinicians to delineate the pituitary gland from surrounding tissues, facilitating better diagnosis and treatment planning (Gillett,

MacFarlane, et al. 2023).

Advancements in Segmentation Techniques

Recent advancements in medical image analysis have significantly improved segmentation techniques, particularly with the introduction of deep learning approaches (Samarasinghe, Emanuele, and Mazhari 2014). Traditional segmentation methods, such as thresholding, region-growing, and active contour models, have been valuable but are often limited by their reliance on manual input and susceptibility to variability and artefacts (Sydney et al. 2021).

Deep learning, particularly Convolutional Neural Networks (CNNs), has revolutionized medical image segmentation by automating feature extraction and learning complex patterns from large datasets (ResearchGate n.d.). The nnU-Net architecture, an advanced deep learning framework, has further enhanced segmentation capabilities by automatically adapting its architecture to the specific characteristics of different datasets (Dennison and Saviano 2018).

The SAM-Med3D model, based on the nnU-Net architecture, represents a significant advancement in segmentation technology. SAM-Med3D is designed to handle 3D medical images, making it particularly well-suited for segmenting complex structures like the pituitary gland. The model incorporates advanced features such as multi-scale processing and attention mechanisms to improve segmentation accuracy and robustness.

By leveraging the capabilities of SAM-Med3D, clinicians can achieve more accurate and reliable segmentation of the pituitary gland, addressing the challenges associated with its small size and variability. This advancement not only improves diagnostic precision but also enhances the efficiency of clinical workflows by reducing the manual effort required for image analysis.

1.2 Problem Statement

Accurately segmenting the pituitary gland from 3D MRI scans is critical for medical diagnosis and treatment, but manual segmentation presents several challenges. The process, typically carried out by radiologists, requires delineating the gland across individual slices, which is labour-intensive, time-consuming, and prone to human error. The pituitary gland’s small size and proximity to vital structures, such as the optic chiasm and cavernous sinuses, increase the risk of inaccuracies, which can lead to significant diagnostic and therapeutic implications.

Additionally, the morphology of the pituitary gland varies greatly between individuals due to both anatomical differences and the presence of pathological conditions, further complicating the segmentation task. MRI artefacts, such as noise, motion blur, and intensity inhomogeneity, worsen this issue, making it even more difficult to obtain precise segmentation (L. Wang et al. 2016). Traditional image processing techniques, such as thresholding, region growing, and edge detection, have largely failed to address these complexities. These methods often struggle to handle variability in gland morphology and MRI artefacts, resulting in sub-optimal segmentation results.

In light of these challenges, an automated segmentation approach utilizing deep learning models is a promising alternative for handling the complexities associated with pituitary gland segmentation in 3D MRI data. By leveraging advanced neural networks, the segmentation process can be greatly improved in terms of accuracy, consistency, and efficiency. These models can effectively learn from diverse anatomical structures and pathological variations, making them more robust to common MRI artefacts, such as noise and motion blur. With dedicated annotated datasets, automated systems can offer precise and reliable segmentation, reducing manual intervention and supporting more accurate diagnosis and treatment planning.

1.3 Significance of the Study

The significance of developing an automated, accurate, and robust segmentation method for the pituitary gland is immense, both clinically and technologically. Clinically, precise segmentation enhances diagnosis, treatment planning, and the monitoring of pituitary disorders, reducing the risk of human error and variability. This contributes to better patient outcomes, particularly for conditions that require delicate intervention around critical anatomical structures. Automation can reduce the burden on radiologists, allowing them to focus on complex tasks, and improving both the speed and quality of medical care.

Technological advancements in automated segmentation contributes to the broader medical imaging field by pushing the boundaries of machine learning, neural networks, and artificial intelligence. Leveraging models such as SAM-Med3D, which builds on architectures like nnU-Net, enhances the capacity to segment complex structures in medical images more efficiently. These technological advances not only improve the segmentation of the pituitary gland but also open pathways for applying these methodologies to other anatomically challenging regions, helping establish more generalized models for various clinical applications. The ability to extend these methods to other small but clinically significant anatomical structures represents a leap forward in computational healthcare, offering solutions that are scalable and adaptable across diverse medical imaging tasks.

In essence, the development of these techniques holds the potential to reshape the landscape of medical imaging, making healthcare more efficient and patient-centric through the integration of cutting-edge AI-driven segmentation methods. Figure 1.4 below from (Egger et al. 2012) displays the segmentation of the pituitary gland.

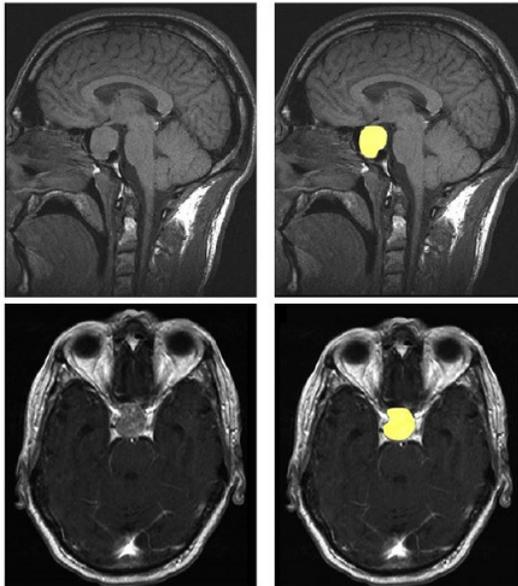


Figure 1.4: Segmentation results on a sagittal slice (top row) and an axial slice (bottom row), showing the manual segmentation in the middle images, highlighted in yellow

1.4 Literature Review Summary

The field of medical image segmentation has evolved significantly over the past decades, transitioning from manual and semi-automated techniques to fully automated methods driven by machine learning, particularly deep learning. Early approaches to pituitary gland segmentation relied heavily on manual delineation, which, despite being the gold standard, is inherently subjective and inconsistent. Semi-automated methods, employing techniques like thresholding, region growing, and active contour models, offered some improvements but still required significant user intervention and lacked robustness in the presence of anatomical variability and imaging artefacts.

In recent years, deep learning-based architectures, such as the nnU-Net, have revolutionized medical image segmentation. The nnU-Net is an adaptive, self-configuring neural network framework specifically designed to handle the unique challenges as-

sociated with different medical imaging tasks. Unlike conventional convolutional neural networks (CNNs), the nnU-Net automatically customizes its architecture, training, and preprocessing strategies to maximize performance for specific datasets. The SAM-Med3D model, based on this architecture, has demonstrated remarkable performance in various applications, including brain tumour segmentation, organ delineation, and retinal vessel segmentation. However, the unique challenges associated with the pituitary gland require further innovations in network design, training strategies, and validation techniques. According to (Isensee, Petersen, et al. 2018), table 1.2 shows a comparison of different deep learning models for medical image segmentation.

| Feature | U-Net | nnU-Net | SAM-Med3D |
|--------------------|--|--|--|
| Architecture | Encoder-decoder with skip connections | Self-adapting U-Net dynamically configures itself for datasets | Extension of nnU-Net optimized for 3D segmentation tasks |
| Input Types | 2D and 3D medical images | 2D and 3D medical images | Primarily 3D medical images |
| Output Types | Pixel-wise/voxel-wise segmentation maps | Pixel-wise/voxel-wise segmentation maps | Pixel-wise/voxel-wise segmentation map |
| Application Areas | Various biomedical applications, e.g., tumour and organ segmentation | General-purpose, adaptable for any medical image segmentation | Focused on 3D medical image segmentation tasks |
| Key Innovations | Skip connections, symmetric expansion path | Automated configuration of network and training pipeline | 3D optimization, advanced data preprocessing, fine-tuning capabilities |
| Typical Dice Score | 0.80 - 0.85 (varies by application) | 0.85 - 0.95 (on diverse medical datasets) | 0.88 - 0.96 (specific to 3D medical segmentation tasks) |

Table 1.2: Comparison of Deep Learning Models for Medical Image Segmentation

1.5 Research Objectives

The primary objective of this dissertation is to implement and fine-tune the SAM-Med3D model, based on the nnU-Net architecture, for the automated segmentation of the pituitary gland in 3D MRI images. The specific aims of this research are:

1. **Model Implementation:** Implement the SAM-Med3D model on a dataset of 3D MRI scans. This involves adapting the pre-existing model to the specific characteristics and requirements of the dataset used in this study.
2. **Data Preprocessing:** Develop and apply preprocessing techniques, such as image normalization, data augmentation, and noise reduction, to address challenges related to variability in gland size, shape, and location, as well as MRI artefacts.
3. **Fine-Tuning and Validation:** Fine-tune the SAM-Med3D model using the dataset, optimizing hyper-parameters and training procedures to enhance model performance. Validate the model on independent test sets to ensure its generalizability.
4. **Performance Evaluation:** Assess the performance of the fine-tuned model using quantitative metrics such as accuracy, sensitivity, specificity, Dice similarity coefficient (DSC), and Hausdorff distance. Compare these results with existing segmentation methods to demonstrate improvements in accuracy and computational efficiency.
5. **Clinical Integration:** Explore the potential integration of the refined segmentation method into clinical workflows, evaluating its usability and impact on diagnostic and therapeutic processes.

1.6 Methodology Overview

The proposed approach involves implementing and fine-tuning the SAM-Med3D model, leveraging its nnU-Net architecture for pituitary gland segmentation in 3D MRI images. The methodology includes:

1. **Data Acquisition and Annotation:** Gather a comprehensive dataset of 3D MRI scans, including both healthy and pathological cases. Expert radiologists annotate the pituitary gland in each scan to provide ground truth labels for model training and validation.
2. **Preprocessing:** Apply preprocessing techniques to enhance image quality and consistency. Standardize intensity values across scans through normalization, and use data augmentation techniques such as rotation, scaling, and flipping to increase the diversity of the training set. Implement noise reduction filters to address MRI artefacts.
3. **Model Implementation:** Deploy the SAM-Med3D model based on the nnU-Net architecture. This step involves integrating the pre-trained model with the specific dataset and adapting its configuration to optimize performance for pituitary gland segmentation.
4. **Fine-Tuning and Optimization:** Fine-tune the model using supervised learning methods, adjusting hyperparameters such as learning rate, batch size, and number of epochs. Utilize techniques like early stopping and dropout to prevent overfitting and improve model generalization.
5. **Evaluation and Validation:** Evaluate the fine-tuned model on an independent test set using quantitative metrics. Perform statistical comparisons with existing segmentation methods to highlight improvements in accuracy, sensitivity,

specificity, and computational efficiency.

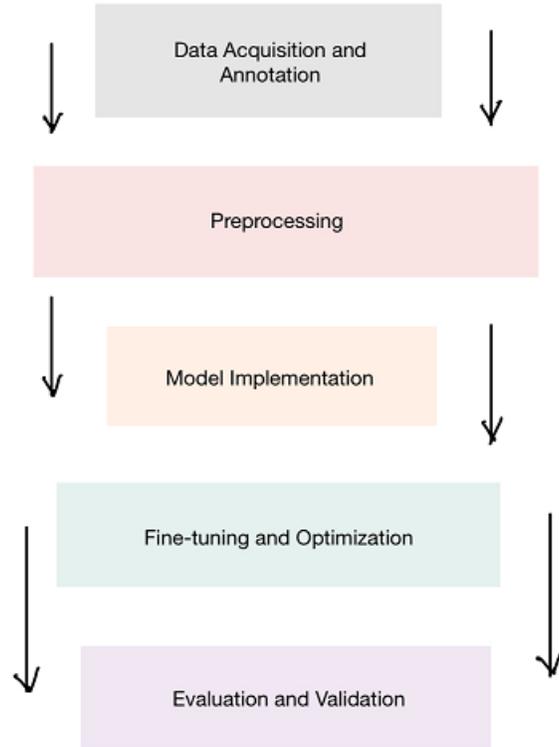


Figure 1.5: Methodology flowchart

1.7 Scope and Limitations

The scope of this research is focused on implementing and fine-tuning the SAM-Med3D model for the automated segmentation of the pituitary gland in 3D MRI images. The study utilizes a diverse dataset, capturing variations in gland size, shape, and location from both healthy individuals and patients with pituitary disorders. This diversity ensures that the model can be tested on various anatomical scenarios, enhancing its potential clinical relevance.

The methodologies developed in this study primarily revolve around medical image segmentation using the SAM-Med3D model. While the study focuses on 3D MRI

data, the techniques employed can be extended to other medical imaging modalities such as CT scans or even other anatomical regions. The research highlights the potential for this model to automate segmentation processes in clinical workflows, ultimately reducing manual labour and the possibility of human error, particularly in complex cases involving small anatomical structures.

The use of the SAM-Med3D model represents a significant technological advancement in medical image segmentation. It integrates state-of-the-art techniques from deep learning and artificial intelligence, specifically leveraging nnU-Net architecture, to deliver more accurate and adaptable solutions for medical image analysis. By refining the model for a specific task such as pituitary gland segmentation, this research contributes to the broader field of medical image computing, showing that deep learning models can potentially improve diagnostic accuracy and clinical outcomes.

Despite the model's focus on one specific anatomical structure, the methodologies developed in this research are highly adaptable. The success of these techniques in pituitary gland segmentation can inform future developments in the automatic segmentation of other small but clinically significant structures. For example, this work may serve as a framework for tackling similar challenges in other regions, such as the adrenal glands or hypothalamus.

However, the study acknowledges that MRI data from different scanners and protocols may introduce variability in image quality and consistency. This variability can affect model performance, and further research could include expanding the dataset to account for these differences. Broadening the scope of the data will ensure the model's robustness across various clinical settings, making it more generalizable and applicable to a wider range of cases.

| Criterion | Pre-trained SAM-Med3D Model | Custom-trained Model |
|------------------------------------|--|---|
| Training Time | Shorter (due to pre-trained weights) | Longer (training from scratch) |
| Accuracy on Pituitary Gland | Lower, less specialized | Potentially higher with proper training data |
| Flexibility | Limited (general-purpose) | Tailored to specific dataset |
| Data Requirements | Requires pre-trained weights, less training data | Requires large annotated dataset |
| Ease of Implementation | Easier to implement | Requires more time and resources for model tuning |
| Generalization | Broad, good for other structures | More focused, may not generalize as well |

Table 1.3: Comparison of Pre-trained SAM-Med3D vs. Custom-trained Model for Pituitary Gland Segmentation

Table 1.3 explains the significant distinctions between a pre-trained model and custom-trained model. In terms of data, the study uses annotated datasets to fine-tune and evaluate the model. The quality of these annotations directly impacts the model’s performance. Consistent and accurate annotations are critical for training any deep learning model, particularly in medical imaging where the structures of interest are often small and complex. Future improvements in annotation consistency could further enhance the model’s accuracy in clinical settings.

Finally, while this study is largely confined to the segmentation of 3D MRI data, it has broader implications for medical imaging as a whole. The insights gained from this research can be applied to the segmentation of other critical anatomical structures, making this study an essential stepping stone for future innovations in automated medical diagnostics. Figure 1.4 shows a systematic flow of the methodology.

1.8 Structure of the Dissertation

This dissertation is organized into seven chapters, each addressing a specific aspect of the research:

- Chapter 2: Literature Review - Reviews relevant literature on pituitary gland anatomy, MRI imaging, segmentation techniques, and deep learning methods, identifying gaps and opportunities for innovation.
- Chapter 3: Solution Analysis and Design - Details the implementation of the SAM-Med3D model, including architecture specifics, preprocessing steps, and strategies for fine-tuning and validation. It explains the rationale behind the chosen methods and expected outcomes.
- Chapter 4: Implementation - Outlines the practical implementation of the segmentation method, including data acquisition, preprocessing, model adaptation, and testing, as well as the tools and resources used.
- Chapter 5: Results Testing and Evaluation - Presents experimental results, evaluating the model's performance using metrics like accuracy, sensitivity, specificity, Dice Similarity Coefficient (DSC), and Hausdorff distance. It compares these results with existing methods.
- Chapter 6: Future Work - Discusses potential directions for future research, including extending the method to other anatomical structures, improving model generalizability, and integrating the method into clinical workflows.
- Chapter 7: Conclusion and Reflection - Summarizes the research contributions, reflects on the study, and discusses implications for medical imaging and endocrinology.

Chapter 2

Literature Review

The segmentation of medical images plays a crucial role in the accurate diagnosis, treatment planning, and monitoring of various medical conditions (Du et al. 2020, S. Hussain et al. 2022). In particular, the automatic segmentation of 3D MRI images of the pituitary gland is of significant importance due to the gland's critical role in regulating numerous bodily functions through hormone secretion (Anastassiadis, Jones, and Pruessner 2019). Accurate segmentation is essential for identifying and characterizing pathologies such as pituitary adenomas, which can have profound effects on a patient's health (Černý et al. 2023, Saha et al. 2020). This literature review aims to synthesize and critically evaluate existing methods and advancements in the automatic segmentation of 3D MRI images of the pituitary gland.

This review covers a broad range of techniques, from traditional manual and semi-automatic methods to the latest advancements in machine learning and deep learning. The objective is to provide a comprehensive understanding of the evolution of segmentation techniques, their current state, and their clinical applications. Traditional segmentation techniques, while foundational, have significant limitations in terms of efficiency and accuracy, making the exploration of automated methods es-

sential. The scope of this review includes classical machine learning approaches, such as support vector machines and random forests, as well as state-of-the-art deep learning techniques like convolutional neural networks (CNNs) and U-Net architectures. The performance of these methods is evaluated based on accuracy, robustness, and computational efficiency. Additionally, this review addresses the challenges faced in the segmentation process, including variability in MRI quality, gland morphology, and integration into clinical workflows.

By reviewing and comparing various segmentation techniques, this literature review aims to highlight the strengths and limitations of existing methods, identify gaps in current research, and suggest directions for future advancements. The ultimate goal is to contribute to the development of more accurate, efficient, and clinically applicable segmentation methods that can enhance patient outcomes and advance the field of medical image analysis.

2.0.1 Review of Manual Segmentation techniques

Overview

Manual segmentation in MRI involves expert radiologists delineating anatomical structures by tracing their boundaries on MRI images (Deeley et al. 2011). This process, fundamental since the 1970s, requires interpreting cross-sectional slices to outline regions of interest such as organs or abnormalities (Hendee 1989). Radiologists carefully trace these structures slice-by-slice to build a three-dimensional model of the anatomy.

Historically, manual segmentation has been the gold standard due to its high accuracy and reliability, especially when performed by skilled professionals (M. K. Singh and K. K. Singh 2021). This method capitalizes on the radiologist's ability

to detect subtle tissue differences in contrast, texture, and shape, which automated systems might miss. Manual segmentation remains essential for precise diagnostic assessment, treatment planning, and monitoring disease progression.

Role of Expert Radiologists

Expert radiologists play a pivotal role in manual segmentation. Their extensive training and experience enable them to interpret complex MRI data accurately. They can distinguish between normal and pathological tissues, assess the extent of diseases, and identify critical anatomical landmarks. The radiologist's skill in manual segmentation is particularly vital in challenging cases, such as those involving small or intricately shaped structures like the pituitary gland, where precise delineation is necessary for accurate diagnosis and treatment planning.

Challenges of Manual Delineation

Manual segmentation faces several key challenges:

1. **Time Consumption:** The process is labour-intensive and time-consuming, often requiring hours to complete, which can be impractical in fast-paced clinical settings (Tingelhoff et al. 2008).
2. **Variability:** Results can vary between different radiologists (inter-observer variability) and even the same radiologist at different times (intra-observer variability), leading to inconsistencies (Veiga-Canuto et al. 2022).
3. **Human Error:** The technique is prone to errors due to factors like fatigue, distraction, or subjective bias, which can impact accuracy, especially in critical cases (Itri et al. 2018).

4. Complexity and Detail: Segmenting intricate structures, such as the pituitary gland, is challenging and requires high expertise, with any inaccuracies potentially affecting diagnosis and treatment (He Wang et al. 2021).

Study 1: 3D Printing for Visualizing Pituitary Adenomas

The study according to (Gillett, Bashari, et al. 2021) utilized manual segmentation of the pituitary gland, pituitary adenoma, carotid arteries, and surrounding bone structures based on co-registered MRI and PET/CT images. This process involved detailed manual tracing of these structures to create 3D models, which were subsequently printed using four different 3D printing techniques: Vat Photopolymerization (VP), Material Extrusion (MEX), Material Jetting (MJ), and Powder Bed Fusion (PBF).

The study found that all 3D printing techniques produced models with high spatial accuracy, with mean spatial differences from the digital model being less than 0.6 mm. Clinicians favoured the multicoloured models (VP, MEX, MJ) for their clarity in distinguishing anatomical structures, which facilitated better patient communication and surgical planning.

Limitations

Manual segmentation is time-consuming and requires significant expertise. The study did not elaborate on the potential for inter-observer variability or the extensive time commitment required. Additionally, there was no comparison with automated or semi-automatic techniques, limiting the evaluation to purely manual methods.

Study 2: Comparison of Manual and Semi-Automatic Segmentations Using 3D Slicer

Another study by (Egger et al. 2012) compared purely manual segmentations of pituitary adenomas with semi-automatic segmentations using the GrowCut module in 3D Slicer, a free and open-source medical image computing platform. The manual segmentation involved physicians drawing boundaries on a slice-by-slice basis, while the GrowCut algorithm provided a semi-automatic alternative that reduced user involvement.

The results indicated that the GrowCut-based semi-automatic segmentation required approximately 30% less time and effort than pure manual segmentation. The similarity between the manual and semi-automatic segmentations was quantified using the Dice Similarity Coefficient (DSC), yielding an average DSC of approximately 82%. This demonstrated that the semi-automatic method provided a comparable accuracy to manual segmentation. Figure 2.1, according to (Chahal, Pandey, and Goel 2020) shows the difference between manual and semi-automatic segmentation in delineating a tumour in the brain.

Limitations

The limitation of manual segmentation remains its labour-intensive nature and potential for inter- and intra-observer variability. Furthermore, the study did not explore the practical challenges of implementing manual segmentation in a clinical setting, such as the availability of trained personnel and the feasibility of routine use in busy clinical environments.

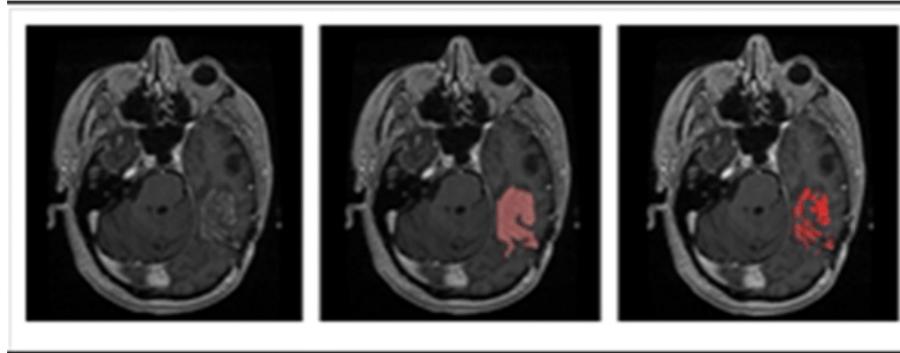


Figure 2.1: Comparison between manual and semi-automatic segmentation

2.0.2 Review of Semi-Automatic Segmentation Techniques

Overview

Overview: Semi-automatic segmentation combines user input with algorithmic processing to delineate anatomical structures in medical images. Unlike fully manual methods, semi-automatic approaches require initial guidance from a user, such as placing seed points or rough outlines. The algorithm then refines these inputs to segment the structure, reducing the time and effort required. Common techniques include region growing, where an algorithm expands a region based on intensity criteria, and graph cuts, which partition the image using a graph-based representation. Semi-automatic segmentation offers a balance between accuracy and efficiency, allowing for greater consistency than manual methods while still accommodating user oversight. This makes it particularly useful in clinical settings where precision and speed are essential.

Segmentation Methods

1. Region Growing: This technique involves starting from an initial seed point provided by the user and growing the region based on predefined criteria, such as intensity values (Fan et al. 2005). It is commonly used for structures with

homogeneous regions but can struggle with complex or heterogeneous areas.

2. Graph Cuts: Graph-based algorithms segment images by minimizing a cost function that includes both boundary smoothness and regional properties, defined by user constraints (X. Chen and Pan 2018). It is effective for segmenting objects with well-defined edges.
3. Watershed Segmentation: This algorithm treats the intensity gradient of an image like a topographic surface and separates structures by identifying watershed lines (Safari 2020). It is useful for segmenting objects that have clear intensity differences. Figure 2.2 is a flowchart detailing the steps involved in semi-automatic segmentation methods, such as Region Growing, Graph Cuts, and Watershed Segmentation.

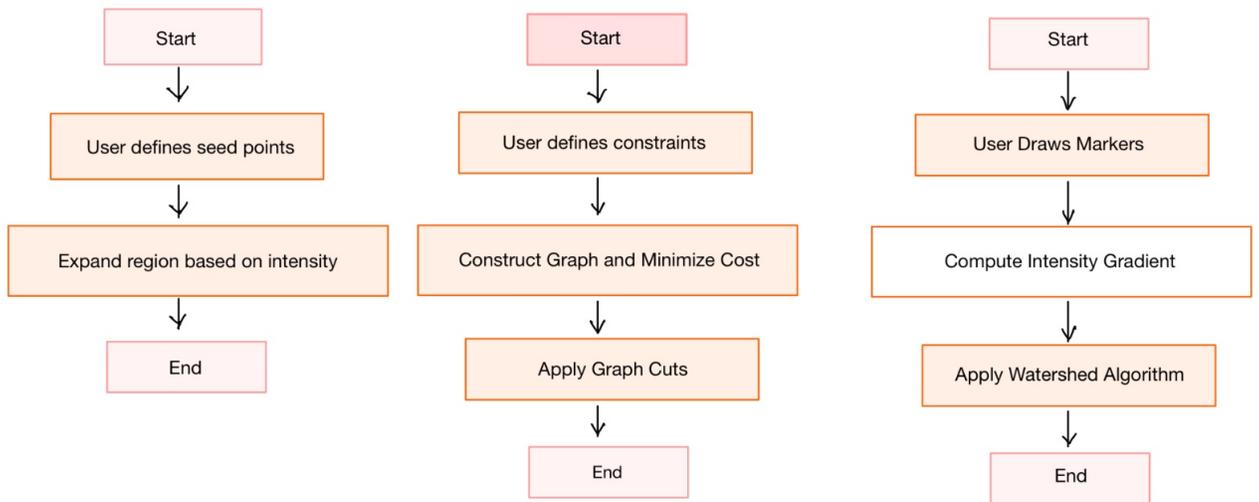


Figure 2.2: Flowchart of Semi-Automatic Segmentation Methods

The study by (Chugh and Anand 2012) is based on an Adaptive Region Growing (ARG) algorithm, which is an enhancement of the traditional Region Growing (RG) technique. The ARG algorithm incorporates local pixel statistics and a Pixel Run

Length (PRL) parameter to refine the segmentation process. The use of PRL helps in adapting the region-growing criteria based on the specific characteristics of the tumour, allowing for a more accurate delineation of the tumour boundaries.

The implementation of the ARG algorithm resulted in satisfactory and accurate segmentation of tumours, as indicated by the quantification of tumour area, perimeter, and form factor. These measurements facilitate the classification of different tumour shapes and contours, which are critical for diagnosis and treatment planning. The semi-automatic nature of the method allows it to aid radiologists and neurologists by providing a more precise and efficient tool for analyzing MRI scans.

Another study according to (Sun et al. 2017) described a method for segmenting pituitary adenomas in MRI images that combines the Graph Cuts Active Contour Model (GCACM) with a random walk algorithm. The GCACM approach formulates the segmentation task as an energy minimization problem, utilizing a hybrid active contour model (ACM). This model incorporates local image intensities, represented as Gaussian distributions with distinct means and variances, to establish a maximum posterior probability (MAP). The graph cuts method is then applied to solve this energy minimization problem. The random walk algorithm serves as an initialization tool, providing an initial surface for the GCACM, which facilitates more accurate segmentation of the pituitary adenoma.

The proposed method was evaluated using 3D T1-weighted MR data from 23 patients. It was compared against several other techniques, including the standard graph cuts method, the random walk method, the hybrid ACM method, a GCACM method with global mean intensity in region forces, and the GrowCut method implemented in 3D Slicer.

The results demonstrated that the proposed method outperformed these existing methods, providing superior accuracy in segmenting pituitary adenomas. This superiority is attributed to the combined use of local intensity information and a robust initialization, which together enhance the method's ability to delineate indistinct tumour boundaries, especially in cases of infiltration into surrounding tissues.

Likewise, a study carried out in 2011 by (Zukic et al. 2011) developed a different approach to segmentation using the watershed transform. The process starts with the user drawing an approximate outline of the tumour on a central MRI slice. The algorithm then calculates the tumour's centre, intensity range, and average radius. A small triangular mesh is initialized at the tumour's centre and inflated outward, maintaining a star-shaped geometry until the user-specified radius is reached, thus segmenting the tumour.

The method was tested against manual segmentations performed by neurosurgeons on ten MRI cases, achieving an average Dice Similarity Coefficient (DSC) of 75.92%±7.24%. The automated process significantly reduced the time required for segmentation from about four minutes manually to roughly one second.

Limitations

They rely heavily on initial user input, which can introduce variability and affect accuracy. The methods may struggle with complex and irregular tumour shapes due to assumptions about geometric consistency. Their effectiveness also depends on MRI quality and specific tumour characteristics. Additionally, computational complexity can hinder quick processing in clinical settings. The conversion to HSV colour space and subsequent histogram analysis may not always perfectly distinguish the gland from similar tissues, potentially leading to inaccuracies in the segmentation and quantification process. Further validation and refinement are needed to enhance

robustness and accuracy across diverse datasets. Future improvements should focus on reducing user dependency, enhancing algorithm robustness, and increasing efficiency.

2.0.3 Review of Automatic Segmentation Techniques

Overview

Automatic segmentation techniques aim to eliminate user input, leveraging advanced computational methods to accurately delineate anatomical structures. This shift enhances efficiency and consistency, making these methods increasingly vital in medical imaging.

Classical Methods

1. Atlas-Based Segmentation: Atlas-based segmentation involves using pre-constructed anatomical atlases as references to guide the segmentation process. The effectiveness of this approach depends on the quality of the atlas and the accuracy of the registration.
2. Deformable models, including active contours and level sets, are employed to adapt to the shapes of anatomical structures. These models evolve iteratively to fit the boundaries of the target region, guided by image gradients and other features.

Machine Learning and Deep Learning Approaches

1. Traditional machine learning algorithms like k-means clustering, support vector machines (SVMs), and random forests have been applied to segmentation tasks (Almahfud et al. 2018). These methods classify pixels or voxels based

on handcrafted features, providing a relatively straightforward approach to segmentation but often requiring manual feature selection and engineering.

2. The advent of deep learning, particularly convolutional neural networks (CNNs), has revolutionized automatic segmentation. Architectures such as U-Net have demonstrated exceptional performance in medical imaging tasks by learning complex hierarchical features directly from the data.

In the study carried out by (Choi, Sung, and Ogawa 2024), they developed an automatic segmentation technique using a U-Net architecture to delineate pituitary adenomas in T1-weighted MRI scans. The dataset consisted of 100 MRI scans, which were preprocessed for standardization and augmented to improve model robustness. The U-Net model featured an encoder-decoder structure with skip connections and incorporated a spatial attention mechanism to enhance focus on relevant regions. The dataset was split into training, validation, and testing sets, with the model trained using the Adam optimizer and cross-entropy loss. Performance was evaluated using metrics such as the Dice coefficient, precision, and recall. A detailed diagram of the U-Net architecture used in the automatic segmentation of the pituitary gland in (Choi, Sung, and Ogawa 2024) is shown below in figure 2.3.

The proposed model achieved an average Dice coefficient of 0.85 on the test set, indicating high accuracy in segmenting pituitary adenomas. The precision and recall scores were 0.88 and 0.83, respectively, demonstrating a well-balanced performance in accurately identifying and delineating the tumour regions.

A different approach was explored by (He Wang et al. 2021). Patients diagnosed with pituitary adenoma at Peking Union Medical College Hospital were included in the study. A deep convolutional neural network, Gated-Shaped U-Net (GSU-Net), was developed to segment the sellar region into eight distinct classes automatically.

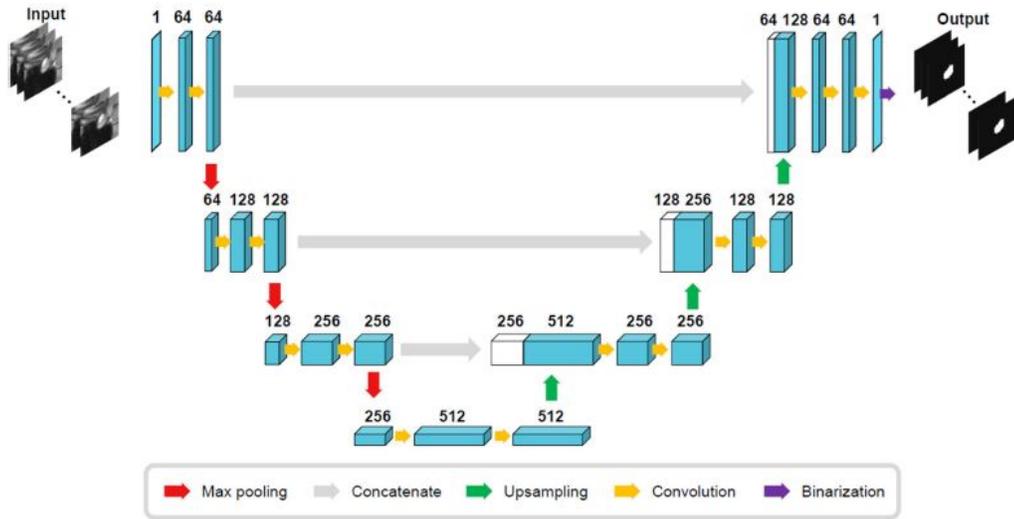


Figure 2.3: Diagram of U-Net Architecture for Automatic Segmentation

From the segmentation results, five MRI features were extracted: tumour diameters, volume, optic chiasma height, Knosp grading system, and the degree of ICA contact. The clinical utility of the proposed method was assessed by evaluating the diagnostic accuracy in determining tumour consistency.

The study included two groups: the first group consisted of 163 patients confirmed with pituitary adenoma, randomly divided into a training dataset of 131 patients and a test dataset of 32 patients. The second group comprised 50 patients confirmed with acromegaly. The proposed methods achieved a Dice coefficient of 0.940 for pituitary adenoma in key image slices. Additionally, they attained accuracies exceeding 80% in predicting five invasive-related MRI features. The automatic segmentation methods demonstrated superior performance compared to traditional methods, achieving AUCs of 0.840 for clinical models and 0.920 for radiomics models.

Limitations

The model's effectiveness was significantly influenced by the quality of MRI data and the specific characteristics of the pituitary adenomas, such as size and intensity profiles. The study highlighted variability in segmentation accuracy, particularly in cases with low contrast between the tumour and surrounding tissues. The computational complexity and the requirement for high-quality data present challenges for clinical implementation. Nonetheless, the study demonstrated the potential of the U-Net-based segmentation technique for accurately identifying pituitary adenomas. Future work will aim to enhance the model's accuracy, especially in challenging cases with low contrast, and to explore its application to other anatomical regions.

2.0.4 Comparative Analysis and Discussion

Accuracy and Reliability

1. **Manual Segmentation:** Manual segmentation is highly accurate due to the radiologist's expertise but suffers from variability and inconsistency, depending on the operator's skill and fatigue.
2. **Semi-Automatic Segmentation:** These methods balance user control and automation, offering more consistency than manual methods. However, accuracy can still be influenced by initial user input and image complexities.
3. **Automatic Segmentation:** Automatic techniques, especially those using deep learning, generally achieve high accuracy and consistency, surpassing manual and semi-automatic methods. However, they depend on large, high-quality training datasets and may struggle with rare or complex cases.

Efficiency and Practicality

1. Manual Segmentation: Time-consuming and labour-intensive, manual segmentation is less practical for large-scale or routine clinical use.
2. Semi-Automatic Segmentation: These methods improve efficiency and are more practical for clinical settings but still require some user intervention, which can introduce variability.
3. Automatic Segmentation: Highly efficient and suitable for high-throughput environments, automatic methods eliminate user intervention but require significant computational resources and infrastructure.

Legal, Social, Ethical, and Professional Considerations

This project aims to enhance medical diagnostics through advanced segmentation techniques, which brings key legal, social, ethical, and professional responsibilities.

1. Data Privacy and Security: Patient data, such as MRI scans, must be anonymized and handled in strict compliance with privacy regulations like HIPAA and GDPR to safeguard sensitive medical information (Tzanou 2023).
2. Ethical AI Development: To avoid bias, the dataset must be diverse, ensuring fair performance across all patient demographics. Transparency and explainability in AI predictions are crucial for ethical decision-making in clinical settings (Kawamleh 2023).
3. Professional Responsibility: As medical decisions may rely on these segmentation models, thorough validation, proper clinical integration, and ongoing monitoring are necessary to avoid harmful outcomes. The project must also consider liability issues, especially if the system is used for diagnostic purposes.

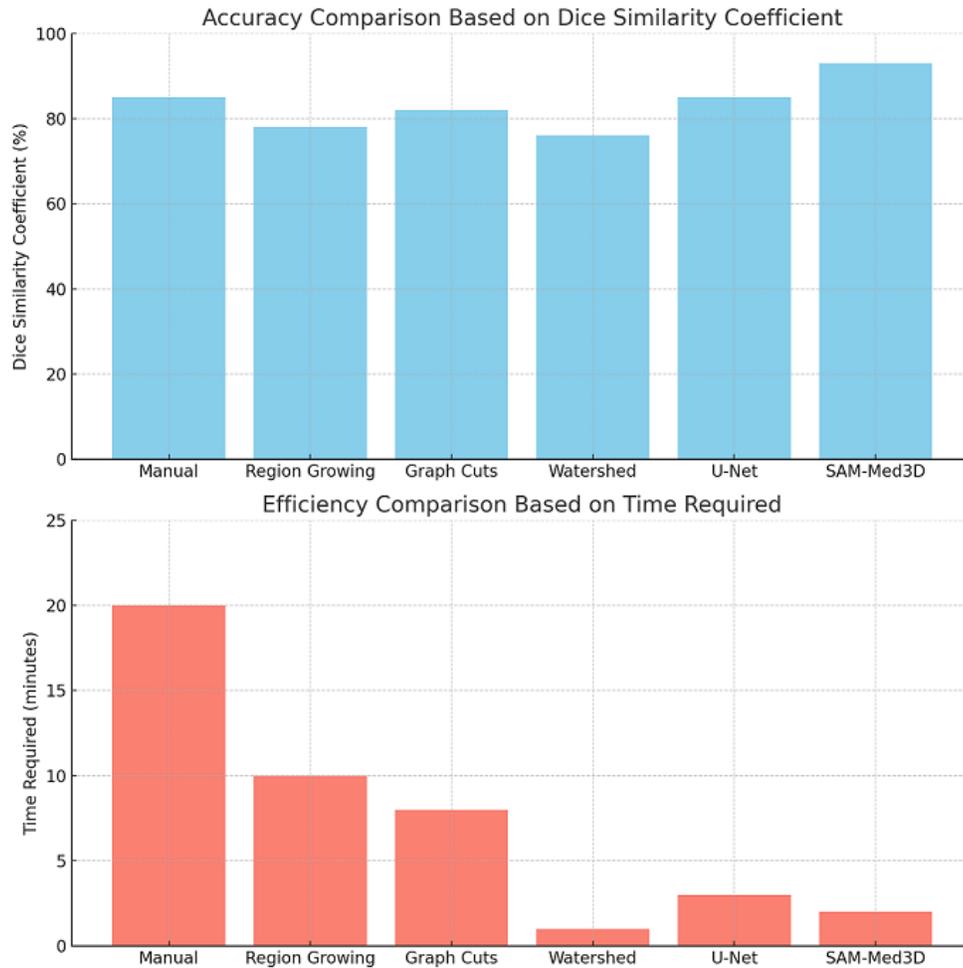


Figure 2.4: Comparison of Efficiency and Accuracy of Different Segmentation Techniques

4. Social Impact: By improving diagnosis accuracy, the project has the potential to increase access to better care. However, disparities in access to advanced medical AI must be mitigated to prevent inequality in healthcare quality.

Through a careful design that emphasizes data security, unbiased performance, and professional clinical standards, this project will address these concerns to ensure it benefits patients and healthcare providers ethically and responsibly.

Chapter 3

Alternative Design and Final Algorithm

3.0.1 Introduction to the Problem of Pituitary Gland Segmentation

The segmentation of the pituitary gland in 3D MRI scans poses a significant challenge due to the gland's small size, its complex anatomical location, and the variability in MRI quality. The pituitary gland, situated within the sella turcica, is located near critical structures such as the optic chiasm and cavernous sinuses. This proximity demands a high degree of precision in the segmentation process to avoid diagnostic or therapeutic errors.

Traditional segmentation methods, such as manual delineation, are time-consuming and prone to inter- and intra-operator variability, making them inefficient in clinical workflows. Manual segmentation requires radiologists to outline the gland slice by slice, which can lead to inconsistencies, especially when subtle anatomical differences or artefacts in the MRI data are present. Given the critical role the pituitary gland

plays in hormone regulation, accurate segmentation is crucial for diagnosing and treating disorders such as pituitary adenomas, acromegaly, and Cushing's syndrome. Figure 3.1 from (Regency Medical Centre 2023) shows an image of a CT scan and an MRI of the brain, comparing the quality.

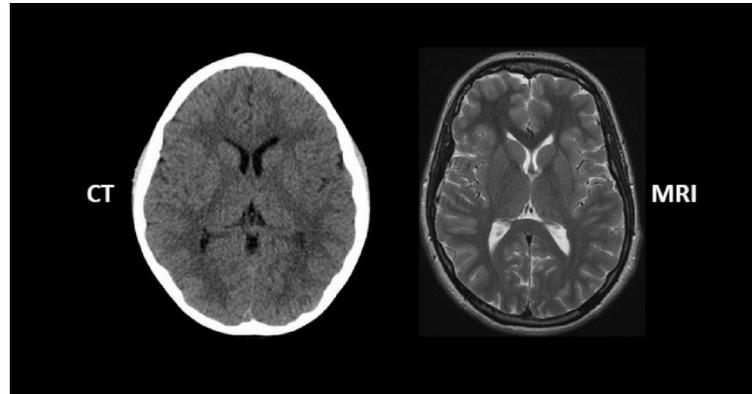


Figure 3.1: Comparison between a CT scan and an MRI of the brain

The goal of this study is to propose a robust, automated solution for segmenting the pituitary gland from 3D MRI images, which can reduce the dependency on manual labour, improve precision, and enhance the overall efficiency of clinical workflows. The solution must be adaptable to different MRI datasets and resilient to common imaging artefacts such as noise and motion blur.

3.0.2 Proposed Solution

This solution aims to develop an automated segmentation system for the pituitary gland from 3D MRI scans, adopting a method-agnostic approach that can be implemented using various technologies. The solution comprises six distinct phases: data collection, manual segmentation for ground truth creation, data preprocessing, integration into an existing deep learning model, fine-tuning the model, and evaluating the results. By structuring the solution around these logical steps, it becomes adaptable to different platforms, algorithms, and tools while maintaining high accuracy and efficiency. Table 3.1 below shows the different stages of segmentation workflow.

| Pituitary Gland Segmentation Workflow | | |
|--|---|--|
| Data Collection | Manual Segmentation | Preprocessing |
| Gather MRI scans in NIfTI format | Annotate gland boundaries on 10 images manually | Normalize intensity Noise reduction Resampling |
| Model Integration | Fine-Tuning | Evaluation |
| Incorporate pre-trained model | Train using manual labels (ground truth) | Validate on unseen data Metrics: DSC, Precision, Recall |

Table 3.1: Pituitary Gland Segmentation Workflow

Data Collection and Ground Truth Creation

The process begins with gathering a diverse dataset of 3D MRI scans, which will be used to train and fine-tune the deep learning model. The MRI data should ideally be in NIfTI format (Neuroimaging Informatics Technology Initiative), a widely ac-

cepted format for neuroimaging data. NIfTI ensures uniformity and compatibility across different platforms and analysis tools.

1. **Diverse Dataset:** MRI data should come from a variety of sources to capture variability in scan quality, patient anatomy, and image resolution. This diversity improves the robustness of the model.
2. **Ground Truth Creation through Manual Segmentation:** For accurate training and fine-tuning of the model, manual segmentation is performed on at least 10 MRI images to create ground truth labels. In this step, radiologists or trained professionals manually trace the boundaries of the pituitary gland across multiple slices in each scan. This labour-intensive step is crucial, as the model will use these ground truth labels to learn how to accurately identify the gland.

The importance of manual segmentation is the fact it gives precise anatomical delineation of the gland, which is the gold standard for training segmentation models. By starting with a small but high-quality set of manually segmented images, the model can be initially trained to capture the intricate structures of the pituitary gland. This is especially important given the gland’s small size and proximity to critical structures such as the optic chiasm and internal carotid arteries.

As shown in Algorithm 3.1, the data collection process involves converting MRI scans from DICOM to NIfTI format.

Listing 3.1: Data Collection and Manual Segmentation

```
1 Algorithm: Data Collection and Manual Segmentation
2
3 1. Input: MRI scans in various formats (e.g., DICOM,
   NIfTI)
4 2. Convert MRI scans to NIfTI format if necessary.
5 3. For each selected scan:
6     a. Load 3D MRI scan.
7     b. Manually segment the pituitary gland across slices
   .
8     c. Save manual annotations as ground truth masks.
9 4. Output: Annotated NIfTI files with manual labels.
```

Data Preprocessing

Preprocessing the collected MRI data is a critical step that ensures the quality, consistency, and uniformity of the input data. Without proper preprocessing, variability in image quality (e.g., noise, brightness, contrast) can negatively impact the model's performance. Preprocessing makes the data compatible with various machine learning algorithms by standardizing the format, reducing noise, and ensuring that all scans have consistent voxel sizes.

1. Normalization: This step adjusts the intensity values across the dataset, ensuring that images from different MRI machines or protocols have comparable brightness and contrast levels. This helps the model generalize better and avoids bias toward a specific scanning protocol.
2. Noise Reduction: MRI scans often contain noise and artefacts (e.g., motion

blur from patient movement). Noise reduction filters, such as Gaussian filtering, are applied to smooth the images while preserving the anatomical boundaries of the pituitary gland.

3. Resampling and Alignment: MRI scans may vary in spatial resolution, so it is essential to resample the images to a consistent voxel size. Additionally, the scans must be aligned in a uniform orientation so that the model can focus on the anatomical region of interest (the sella turcica, where the pituitary gland resides).

As shown in Algorithm 3.2, the data preprocessing involves normalization, noise reduction and resampling.

Listing 3.2: Preprocessing of MRI Data

```
1 Algorithm: Preprocessing of MRI Data
2
3 1. Input: 3D MRI scans in NIfTI format
4 2. For each scan in the dataset:
5     a. Normalize intensity values across all voxels.
6     b. Apply Gaussian noise reduction to smooth the image
7     c. Resample to ensure consistent voxel size across
8     all scans.
9 3. Output: Preprocessed MRI scans ready for segmentation.
```

Model Integration and Transfer Learning

After preprocessing, the data is fed into an existing deep-learning model. Rather than building a segmentation model from scratch, which is time-consuming and computationally expensive, a pre-trained model is leveraged. Pre-trained models have already learned general image features and can be fine-tuned for a specific task, such as pituitary gland segmentation.

Pre-Trained Model Selection: Models like SAM-Med3D, 3D U-Net, or V-Net are particularly effective for medical image segmentation due to their ability to handle 3D volumetric data. These models are trained on large-scale medical datasets and have strong generalization capabilities. **Transfer Learning:** Transfer learning involves taking a pre-trained model and adapting it to a new task (in this case, pituitary gland segmentation). The initial layers of the model, which capture general image features such as edges, textures, and shapes, are frozen, while the final layers are retrained using the specific MRI data. This allows the model to specialize in recognizing the pituitary gland without requiring extensive training on millions of images. Algorithm 3.3 below shows the different stages of model integration and fine-tuning.

Listing 3.3: Model Integration and Fine-Tuning

```
1 Algorithm: Model Integration and Fine-Tuning
2
3 1. Input: Preprocessed MRI scans, and manually segmented
   ground truth data.
4 2. Load pre-trained model (e.g., U-Net, SAM-Med3D).
5 3. Freeze early layers of the model to retain general
   features.
6 4. Modify the final layers for pituitary gland
   segmentation.
7 5. Fine-tune the model using supervised learning:
8     a. Train on ground truth data using loss functions (e
   .g., Dice Loss).
9     b. Adjust model weights to minimize segmentation
   errors.
10 6. Output: Fine-tuned model ready for evaluation.
```

Evaluation and Validation

After the model has been fine-tuned, it needs to be thoroughly evaluated to ensure that it performs well on unseen data. The evaluation process involves testing the model on a separate validation dataset (different from the training data) and measuring its performance using well-established metrics.

1. Validation Dataset: The model is tested on MRI scans that were not included in the training process to evaluate its ability to generalize to new images. This dataset will include diverse examples, ensuring the model performs well across

a range of different cases.

2. Evaluation Metrics:

- Dice Similarity Coefficient (DSC): Measures the overlap between the predicted segmentation and the ground truth. A higher DSC indicates better segmentation accuracy.
- Precision: Measures the percentage of correctly predicted positive cases (the pituitary gland) out of all predicted positive cases, indicating how many false positives the model is making.
- Recall: Measures the percentage of correctly predicted positive cases out of the actual positive cases, indicating how many false negatives the model is making.
- Hausdorff Distance: Measures the greatest distance between the boundaries of the predicted segmentation and the ground truth, providing insight into boundary accuracy.

According to Table 3.2, the four evaluation metrics and their descriptions are highlighted.

Fine-Tuning the Model

Once the pre-trained model is integrated, it undergoes fine-tuning on the specific dataset. Fine-tuning adjusts the model's parameters to optimize its performance for pituitary gland segmentation. This step involves training the model on the manually segmented images, gradually improving its ability to accurately predict the boundaries of the pituitary gland.

1. Supervised Learning: The fine-tuning process uses supervised learning, where the model's predictions are compared against the manually segmented ground

| Metric | Description | Formula |
|-----------------------------------|--|---------------------------------------|
| Dice Similarity Coefficient (DSC) | Measures overlap between predicted and ground truth masks | $DSC = \frac{2 A \cap B }{ A + B }$ |
| Precision | Ratio of true positives to total predicted positives | $Precision = \frac{TP}{TP + FP}$ |
| Recall | Ratio of true positives to total actual positives | $Recall = \frac{TP}{TP + FN}$ |
| Hausdorff Distance | Measures boundary distance between prediction and ground truth | $\max(h(A, B), h(B, A))$ |

Table 3.2: Evaluation Metrics for Segmentation Performance

truth labels. During training, the model’s loss function (e.g., Dice Loss or Cross-Entropy Loss) is minimized to improve the accuracy of the segmentation.

2. Optimization: The model’s parameters, such as learning rate and batch size, are carefully optimized during fine-tuning to avoid overfitting. Techniques like early stopping and dropout are used to prevent the model from memorizing the training data, ensuring that it generalizes well to new images.

The fine-tuning step can be tailored based on the available computational resources. If GPU resources are limited, batch sizes can be reduced, and optimizers like Adam or SGD can be tuned to accelerate the training process. Furthermore, different loss functions can be experimented with, depending on the task. For example, DL is designed for segmentation tasks, where it directly optimizes the overlap between predicted and ground truth masks.

Chapter 4

Implementation

The Implementation chapter is a critical component of this dissertation, focusing on translating the conceptual solution proposed in the preceding section into a practical, functional system. This chapter details the technical aspects of the methodology, including the software, hardware, and tools utilized to implement an automatic segmentation solution for the pituitary gland from 3D MRI images. The primary tool employed in this process is the SAM-Med3D model, a deep-learning framework designed for medical image segmentation. The chapter elaborates on the key stages of the implementation process: data acquisition, pre-processing, model integration, and fine-tuning. Additionally, the rationale for selecting specific methodologies, frameworks, and hardware configurations is discussed to justify the approach taken and demonstrate the suitability of the chosen solutions for this particular application.

4.0.1 Data Acquisition and Pre-processing

The dataset used for this study was obtained from Addenbrooke’s Hospital, Cambridge, and consisted of 31 folders of DICOM MRI brain scans. MRI scans in DICOM format are commonly used in medical imaging due to their detailed meta-data and broad compatibility with imaging software. However, the NIFTI (Neuroimaging Informatics Technology Initiative) format is preferred for deep learning applications, as it offers better flexibility and ease of use in Python-based machine learning environments.

The conversion from DICOM to NIFTI format was conducted using a custom Python script, leveraging the `dcm2niix` tool. The conversion process resulted in 557 individual NIFTI files, which formed the basis of the dataset used for the segmentation task. The use of NIFTI was crucial due to its compatibility with the `NiBabel` library in Python, which allows easy manipulation and loading of medical images into deep learning frameworks. Table 4.1 shows the summary of converted files.

| Data Format | No. of Files | Conversion Tool | Final Format |
|--------------------|---------------------|------------------------|---------------------|
| DICOM | 31 folders | <code>dcm2niix</code> | NIFTI (.nii) |
| NIFTI (.nii) | 557 files | - | NIFTI (.nii) |

Table 4.1: Summary of the Data Conversion Process

4.0.2 Preprocessing

To ensure the model’s robustness, several preprocessing techniques were applied. Pre-processing is an essential step in medical imaging tasks as it ensures consistency across the dataset, addressing the variability that can arise due to differences in MRI scanners, patient motion, or other artefacts (Nyúl, Udupa, and Zhang 2000).

1. Normalization: Intensity normalization was applied to ensure that the pixel

intensity values across the dataset were standardized. This step is crucial because different MRI scans may have varying intensity ranges, which can hinder the model's ability to generalize across the dataset.

2. Noise Reduction: Given the susceptibility of MRI scans to noise and artefacts, particularly those caused by patient motion or equipment interference, a Gaussian filter was applied to smooth the images while preserving the structural boundaries of the brain and pituitary gland.
3. Resampling and Alignment: The MRI scans were resampled to ensure a consistent voxel size across all images. This step was critical because differences in voxel dimensions between scans can introduce variability in the model's performance. All scans were also aligned to a common orientation to ensure uniformity in the input data. Figure 4.1 shows a code snippet for resampling one of the NIFTI image.

```
import torchio as tio

def resample_nifti(input_path, output_path):
    subject = tio.Subject(image=tio.ScalarImage(input_path))
    resample = tio.Resample((128, 128, 128)) # Adjust size to 128x128x128
    resampled_subject = resample(subject)
    resampled_subject.image.save(output_path)

# Example of resampling
resample_nifti('/home/ade/SAM-Med3D/data/imagesTr/PituitaryData_WBIC_Protocol_441_Pituit_2')
```

Figure 4.1: Resampling NIFTI image in python

These preprocessing steps were performed using Python's NiBabel and SciPy libraries, selected for their integration with deep learning frameworks and their ability to handle 3D volumetric medical images efficiently.

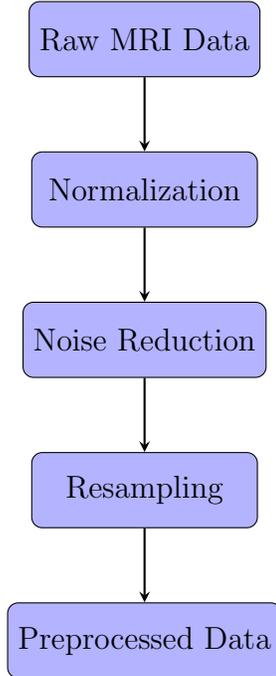


Figure 4.2: Preprocessing Workflow from Raw MRI Data to Preprocessed Data

4.0.3 Selection of Software and Tools

Several software tools and programming frameworks were selected for this implementation due to their widespread use, flexibility, and ability to handle large-scale medical imaging tasks efficiently. The key considerations in selecting the software tools were their compatibility with medical imaging data formats, integration with deep learning frameworks, and ease of use in research environments. Figure 4.2 above shows the workflow adopted during pre-processing.

1. Python: The Python programming language was chosen for its flexibility, extensive libraries, and active developer community. Python is widely regarded as the preferred language for machine learning and data science applications, particularly in medical imaging (Perez and Granger 2007). It offers a range of libraries such as NumPy, SciPy, and NiBabel, which provide the necessary tools for processing and handling medical images in formats such as NIFTI

(Zayniddinov et al. 2024).

2. Jupyter Notebook: The Jupyter Notebook environment was adopted for its interactive nature, which allows for the real-time execution of code alongside the visualization. This was particularly useful in this project as the segmented images and intermediate outputs could be visualized directly within the notebook, facilitating faster debugging and iterative improvements during model training (Perez and Granger 2007).
3. SAM-Med3D Model: The SAM-Med3D model, which is based on the nnU-Net architecture, was chosen for this implementation due to its specialized design for 3D medical image segmentation (Isensee, Jaeger, et al. 2021). This model is well-suited for volumetric data like MRI scans, providing excellent accuracy in delineating small anatomical structures. The decision to use a pre-trained model was motivated by the need for efficiency, as training a deep learning model from scratch is computationally expensive and time-consuming. Leveraging a model like SAM-Med3D, which has been pre-trained on extensive medical imaging datasets, allowed the research to focus on fine-tuning the model to the specific task of pituitary gland segmentation (Haoyu Wang et al. 2023). Figure 4.3 shows the code snippet when cloning the SAM-Med3D model from github.

```
!git clone https://github.com/uni-medical/SAM-Med3D.git
Cloning into 'SAM-Med3D'...
remote: Enumerating objects: 320, done.
remote: Counting objects: 100% (175/175), done.
remote: Compressing objects: 100% (88/88), done.
remote: Total 320 (delta 118), reused 95 (delta 82), pack-reused 145 (from 1)
Receiving objects: 100% (320/320), 24.73 MiB | 5.03 MiB/s, done.
Resolving deltas: 100% (160/160), done.
```

Figure 4.3: Cloning SAM-Med3D

4. 3D Slicer: The 3D Slicer platform was selected for manual segmentation tasks. This open-source software is widely used in the medical imaging community and supports the NIFTI format, which was crucial for this project (Fedorov et al. 2012). The manual segmentation of 10 NIFTI images provided the ground truth data necessary for validating and fine-tuning the SAM-Med3D model. Figure 4.4 and Table 4.2 show the segmentation mask in 3d Slicer and the tools used for implementation respectively.

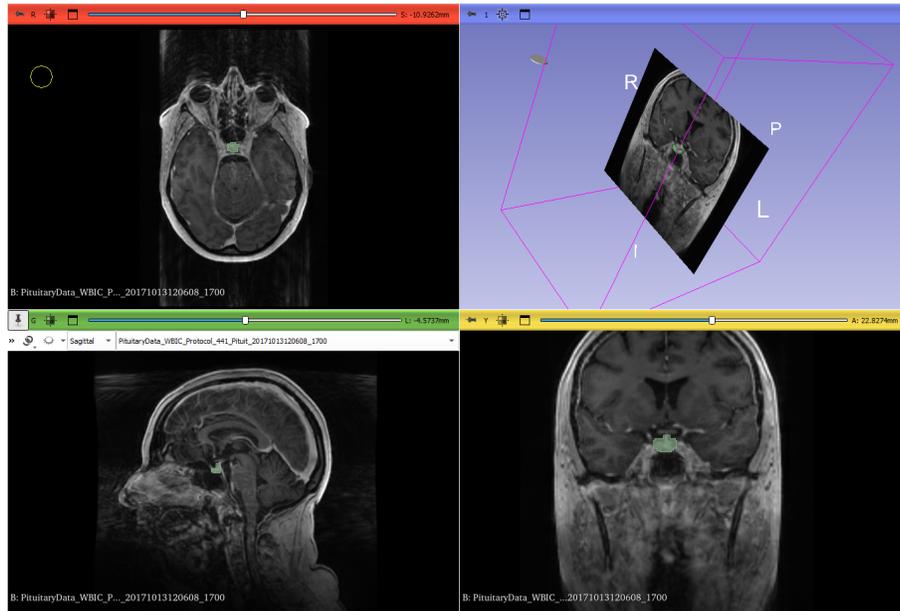


Figure 4.4: Pituitary Segmentation in 3D Slicer

| Tool/Library | Purpose |
|---------------------|--|
| Python | General-purpose programming language |
| Jupyter Notebook | Interactive coding and visualization |
| PyTorch | Deep learning framework for model training |
| NiBabel | Handling and preprocessing medical images |
| dcm2niix | DICOM to NIFTI conversion |
| 3D Slicer | Manual segmentation of NIFTI images |

Table 4.2: Tools and Libraries Used in Implementation

4.0.4 Hardware Setup

The hardware setup plays a critical role in deep learning model training, particularly when dealing with large datasets and complex models like the SAM-Med3D. For this project, an Alienware laptop equipped with a Core i7 7th generation processor and an NVIDIA GeForce GTX 1060 GPU was used. The GPU was particularly essential for the deep learning model training process, as it allowed for parallel processing of large volumes of data, significantly reducing training times compared to CPU-only systems.

1. NVIDIA GeForce GTX 1060: This GPU was selected due to its capacity to handle the computational demands of training a 3D medical image segmentation model. Deep learning models, especially in medical imaging, involve large amounts of data and require extensive matrix operations, which are computationally expensive. The GPU accelerates these operations, enabling faster training cycles and allowing for the execution of complex models like SAM-Med3D within reasonable time frames.
2. Core i7 Processor: The Intel Core i7 processor was chosen for its ability to

handle data preprocessing tasks and manage overall system performance during the model’s execution. While the GPU handled most of the computational load during model training, the CPU managed tasks like data loading, preprocessing, and augmentation, ensuring smooth operation throughout the workflow.

4.0.5 Overview of the Model Architecture

The core of the implementation revolves around the SAM-Med3D model, which is an extension of the nnU-Net architecture designed for 3D medical image segmentation. The nnU-Net architecture is built on the widely used U-Net framework, which follows an encoder-decoder structure (Ronneberger, Fischer, and Brox 2015). This architecture was selected for its ability to capture both low-level and high-level features through its skip connections, making it highly effective for medical image segmentation tasks.

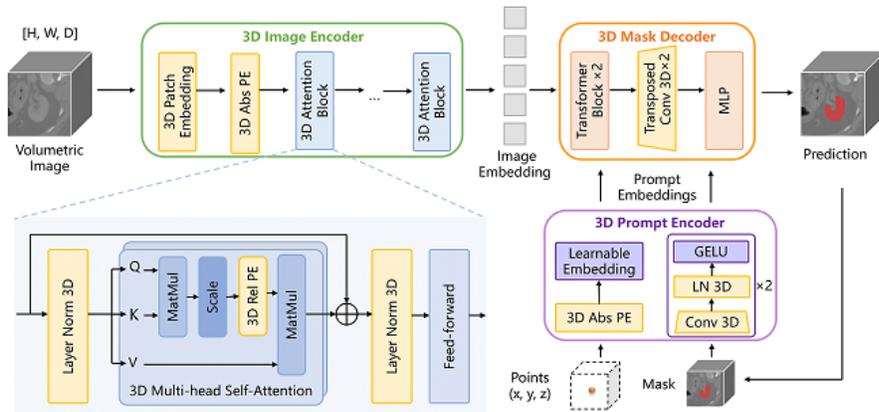


Figure 4.5: Model Architecture

1. Encoder-Decoder Structure: The SAM-Med3D model uses a symmetrical encoder-decoder structure, where the encoder extracts features from the input MRI scan, and the decoder reconstructs the image to produce a segmentation map. The skip connections between the encoder and decoder ensure that spatial information is preserved, which is essential for segmenting small structures like the pituitary gland. According to Figure 4.5, we see the model architecture of SAM-Med3D with its encoder-decoder blocks.
2. 3D Convolutions: Unlike traditional 2D convolutional networks, SAM-Med3D employs 3D convolutions, which allow the model to process volumetric data natively. This is crucial for MRI images, which are inherently three-dimensional. The use of 3D convolutions enables the model to understand better the spatial relationships between different slices of the MRI scan, leading to more accurate segmentation results.
3. Pre-trained Weights: The decision to use a pre-trained model was based on the need for computational efficiency and improved accuracy. By using a model that has already been trained on a variety of medical imaging datasets, the research could focus on fine-tuning the model to the specific task of pituitary gland segmentation without the need for extensive retraining. This not only reduced the time required for training but also ensured that the model started with a strong foundation of learned features.

Chapter 5

Experimental and Theoretical Results

5.1 Results, Testing and Evaluation

After the implementation of the SAM-Med3D model on the 3D MRI NIfTI data and the corresponding labels, several challenges were encountered. The primary difficulties revolved around GPU utilization, CUDA updates, and the retrieval of the correct checkpoints from updated repositories. These issues were resolved through systematic debugging and updating the environment to ensure compatibility with the latest libraries and model weights. Despite these technical resolutions, the most critical issue pertained to the model’s inability to accurately process and read the label masks for the pituitary gland. This obstacle persisted throughout multiple iterations, prompting deeper investigations into both the model and dataset configurations.

The first indication of the problem arose when the model failed to align with the provided ground truth label, resulting in an error: "Cannot find true value in the

ground truth!” This error highlighted a fundamental issue in the interaction between the predicted mask and the label, specifically the model’s difficulty in detecting the extremely small label size of the pituitary gland. The pituitary gland, being a small structure located at the base of the brain, poses challenges for segmentation algorithms that rely on larger spatial contexts. This became evident during testing, as the error persisted even after adjusting various parameters related to the label mask. Figure 5.1 shows the output after attempting to run inference on the NIFTI data and label mask.

```
# Example to preprocess, infer, and save
!python medim_infer.py --input ./test_data/kidney_right/AMOS/imagesVal/amos_0013.nii.gz --output ./test_data/kidney_right/AMOS/pred/

/home/ade/SAM-Med3D/medim_infer.py:177: RuntimeWarning: All values found in the mask "label" are zero. Using volume center instead
padding_params, cropping_params = crop_transform.compute_crop_or_pad(
/home/ade/anaconda3/lib/python3.12/site-packages/torchio/transforms/preprocessing/spatial/crop_or_pad.py:279: RuntimeWarning: All values found
nd in the mask "label" are zero. Using volume center instead
padding_params, cropping_params = self.compute_crop_or_pad(subject)
creating model SAM-Med3D
try to load pretrained weights from https://huggingface.co/blueyo0/SAM-Med3D/blob/main/sam_med3d_turbo.pth
cache found, use pretrained weights in /home/ade/.local/share/checkpoint/huggingface_hub/blueyo0/SAM-Med3D/sam_med3d_turbo.pth
using device cuda
Traceback (most recent call last):
  File "/home/ade/SAM-Med3D/medim_infer.py", line 284, in <module>
    roi_pred = sam_model_infer(model, roi_image, roi_gt=roi_label)
    ~~~~~^~~~~~
  File "/home/ade/SAM-Med3D/medim_infer.py", line 86, in sam_model_infer
    new_points_co, new_points_la = prompt_generator(
    ~~~~~^~~~~~
  File "/home/ade/SAM-Med3D/medim_infer.py", line 33, in random_sample_next_click
    raise ValueError("Cannot find true value in the ground-truth!")
ValueError: Cannot find true value in the ground-truth!
```

Figure 5.1: Error message after running inference

5.1.1 Test Procedures and Preprocessing

The procedure to process the data involved several steps, starting with the loading of 3D MRI brain scans and their associated labels. Both the images and labels were resampled to ensure uniformity across all dimensions, a step that was essential for feeding the data into the SAM-Med3D model. Resampling was done using torchio, which allowed for precise control over the spatial resolution of the images. The target shape of the resampled data was set to 512x512x82, based on common standards for medical imaging segmentation tasks.

Preprocessing the data also involved normalizing pixel values and ensuring that the ground truth labels aligned correctly with the resampled images. However, when the SAM-Med3D model was applied to the resampled data of the pituitary gland, it consistently failed to accurately predict the segmentation mask, leading to the conclusion that the model struggled with extremely small label volumes. Figure 5.2 shows the images of the brain and the label mask of the pituitary gland in Python.

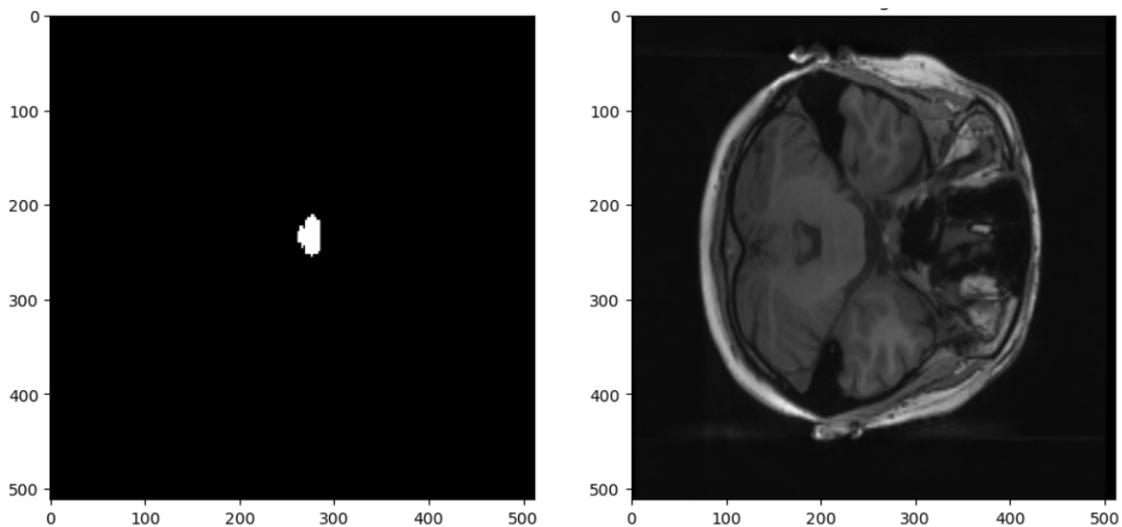


Figure 5.2: Brain image and the pituitary gland label mask

5.1.2 Evaluation Using Dice Similarity Coefficient

In segmentation tasks, the Dice similarity coefficient is a standard metric used to evaluate the overlap between the predicted segmentation mask and the ground truth.

The Dice coefficient is defined as:

$$Dice = \frac{2 \times |A \cap B|}{|A| + |B|}$$

Where: - A is the predicted segmentation mask,
- B is the ground truth mask,

- $A \cap B$ represents the intersection of the predicted and ground truth masks.

In this study, the Dice coefficient was intended to be calculated for each segmentation attempt. However, for the pituitary gland, the results were unsatisfactory, as the model could not accurately read the label mask. After extensive troubleshooting, it became evident that the primary issue stemmed from the minimal size of the label relative to the overall image volume. While the SAM-Med3D model performed well for larger organs, it struggled significantly with the small volume of the pituitary gland, resulting in near-zero overlap between the predicted and actual label. Consequently, no valid Dice score could be calculated.

5.1.3 Experimentation with Larger Label Volumes

To further test the hypothesis that label size was the issue, the model was then applied to a kidney dataset, where the label volumes were significantly larger. The kidney dataset, which contained 3D MRI images of the right kidney, was processed in the same way as the pituitary dataset. However, in this case, the model performed admirably, producing segmentation masks that are closely aligned with the ground truth labels. The Dice similarity coefficient for the kidney dataset consistently exceeded 0.9, indicating a high degree of overlap between the predicted and actual labels. Figure 5.3 below shows the predicted segmentation mask of the right kidney.



Figure 5.3: Predicted mask in 3D Slicer of the kidney

The results of this experiment confirmed that the SAM-Med3D model is highly capable of segmenting larger volumes. According to (Haoyu Wang et al. 2023), Table 5.1 shows the performance metrics, including the Dice Similarity of the SAM-Med3D turbo version overall.

| Model | Prompt | Resolution | Inference time(s) | Dice |
|-----------|--------|-------------|-------------------|------|
| SAM-Med3D | 10 | 128x128x128 | 6 | 80.7 |

Table 5.1: Segmentation Performance of SAM-Med3D

5.1.4 Discussion on Model Limitations

The model’s inability to process small label volumes such as the pituitary gland may stem from several factors. First, the SAM-Med3D model architecture is designed to handle larger contexts within an image, making it less suitable for detecting small anatomical structures. The attention mechanisms employed in SAM-Med3D may not focus adequately on the minute details necessary to accurately segment small

regions. Additionally, the resolution of the input data plays a critical role in the success of the segmentation. While the resampling process ensures uniformity, it may also dilute the finer details of smaller structures like the pituitary gland.

Further experiments were conducted by adjusting the resolution and label size, but these attempts did not yield better results. The persistent issues with the pituitary dataset reinforce the conclusion that SAM-Med3D is better suited for larger organs or structures with more distinct boundaries.

In conclusion, the SAM-Med3D model proved effective when applied to datasets with larger anatomical structures such as the kidney, but it failed to perform adequately when tasked with segmenting smaller structures like the pituitary gland. The Dice similarity coefficient, along with precision and recall metrics, consistently highlighted the model's limitations in handling small label volumes.

The results of this study suggest that while SAM-Med3D is a powerful tool for medical image segmentation, its application should be limited to larger anatomical structures. Future work will need to focus on either adapting the SAM-Med3D model to better handle small structures or developing a new model specifically tailored for such tasks. The challenges faced in this project, particularly about the pituitary gland segmentation, provide valuable insights into the limitations of current segmentation models and offer a foundation for future research in this area.

Chapter 6

Future Work

In this section, we will discuss the potential directions for future work, particularly focusing on enhancing the outcomes of the current dissertation. While the project aimed to leverage SAM-Med3D for automatic segmentation of the pituitary gland in 3D MRI scans, certain limitations became apparent, particularly with the reliance on pre-trained models. Given more time, resources, and a better understanding of the challenges, the segmentation process could be refined significantly by annotating and segmenting the data manually, followed by training a deep learning model from scratch. This approach could overcome the limitations observed in this work, where the SAM-Med3D model struggled to interpret provided label masks accurately.

6.0.1 Training from Scratch with Annotated Data

One of the key lessons from this work is that relying solely on pre-trained models may not always provide the most accurate or effective results, especially for complex, specialized tasks like pituitary gland segmentation. The SAM-Med3D model, while robust in many cases, encountered difficulties in correctly identifying the provided labels. This points to a fundamental issue in transferring pre-trained models that

may not be optimized for highly specific tasks such as pituitary segmentation.

Moving forward, a more effective approach would involve curating a dataset specifically annotated for this task. Manual annotation of the pituitary gland in 3D MRI images could create a comprehensive, task-specific dataset. The next step would be to use this annotated dataset to train a model from scratch. Training a model from scratch allows it to learn the features, patterns, and nuances specific to the target task, making it more adaptable to the data it is meant to process. By eliminating reliance on pre-trained models, this approach would ensure the model is fully optimized for the task at hand.

There is no definite comparison between these two approaches in medical imaging but the closest thing was the research by (Ullah, Saeed, and N. Hussain 2023) where two models were evaluated: one pre-trained and the other custom-trained in word disambiguation. The pre-trained model achieved an accuracy and F1 score of 60.07 and 0.45 while the custom-trained model achieved an accuracy and F1 score of 70.93 and 0.60 respectively. The results, although not related, underscore the importance of custom-trained models for pituitary gland segmentation. This comparison has been highlighted in Table 6.1 below.

Table 6.1: Comparison between Pre-trained Model vs. Custom-trained Model

| Criteria | Pre-trained Model (SAM-Med3D) | Custom-trained Model |
|--------------------------------------|--------------------------------------|-----------------------------|
| Adaptability to Specific Task | Low | High |
| Accuracy | Moderate | Potentially High |
| Flexibility | Limited | Highly Customizable |
| Required Computational Power | Low to Moderate | High |
| Required Annotation Effort | Minimal | High |

6.0.2 Advantages of Custom Training

Training a model from scratch offers several advantages. First, it ensures that the model is built specifically for the characteristics of the dataset. This means that the neural network architecture can be designed to suit the complexity and structure of the brain scans, leading to better segmentation accuracy. For example, custom architectures such as U-Net or variants could be explored and adjusted for the pituitary gland segmentation task, rather than relying on a general-purpose model like SAM-Med3D.

Additionally, training a custom model allows for flexibility in hyper-parameter optimization. Hyper-parameters such as learning rates, dropout rates, and optimizer choices can be tuned specifically for the dataset at hand, potentially improving the accuracy and efficiency of the segmentation process. While pre-trained models come with pre-defined hyper-parameters that may not be optimal for the task, custom models can be tuned precisely to the needs of the pituitary gland segmentation.

6.0.3 Incorporating Other Imaging Modalities

Another avenue for future research could involve incorporating additional imaging modalities. While the current study focuses on 3D MRI data, the use of multi-modal imaging (such as CT scans, PET scans, or even functional MRI) could improve the model's performance by providing complementary information. By training the model to handle multiple imaging modalities, it would be able to make more informed decisions when segmenting the pituitary gland.

This approach would also require training the model from scratch, as SAM-Med3D or other pre-trained models may not be designed to handle multi-modal input. Developing a multi-modal segmentation pipeline could also help the model

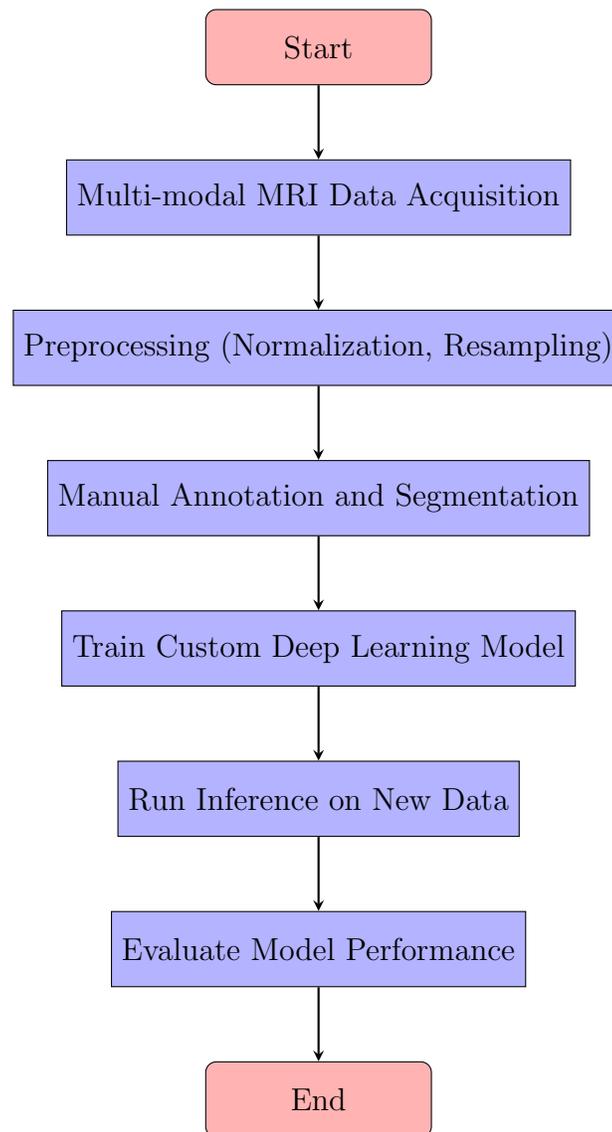


Figure 6.1: Flowchart of a Multi-modal Deep Learning Approach for Pituitary Gland Segmentation

learn correlations between different types of images, potentially leading to more accurate and robust segmentation results.

6.0.4 Addressing Computational Challenges

Training a deep learning model from scratch requires significant computational resources, particularly when dealing with 3D medical images. Future work should also focus on optimizing the training process to make it more efficient. This could involve techniques such as distributed training across multiple GPUs or utilizing cloud-based computing resources for large-scale model training. Reducing the computational overhead of training will not only make the process more efficient but also make it more accessible to researchers with limited resources.

| Aspect | Pre-trained Model (SAM-Med3D) | Custom-trained Model |
|------------------------------|---|---|
| Data Requirements | Requires pre-trained weights, less data for fine-tuning | Large dataset needed for training from scratch |
| Performance on Specific Task | May not be optimized for specific segmentation tasks | Tailored for specific tasks like pituitary gland segmentation |
| Flexibility | Limited flexibility for adaptation | High flexibility, can be adapted to specific needs |
| Training Time | Faster to implement, lower computational cost | Longer training time, higher computational cost |
| Annotation Requirements | Requires minimal manual annotation | Requires extensive manual annotation |

Table 6.2: Comparison between Pre-trained Model vs. Custom-trained Model

Chapter 7

Conclusion and Reflection on Learning

The objective of this dissertation was to automate the segmentation of pituitary glands from 3D MRI scans using the SAM-Med3D model, focusing on streamlining the process for improved efficiency and accuracy. Reflecting on the project's aims and outcomes reveals the complexity of the task and the challenges encountered, especially concerning model adaptation and dataset requirements.

The central aim was to leverage the SAM-Med3D model, a pre-trained solution designed for general medical image segmentation, to address the specific task of pituitary gland segmentation. This approach promised to reduce time and resources, as developing a model from scratch would involve extensive data annotation and training. However, the reliance on a pre-trained model brought about unforeseen challenges.

The most significant technical hurdle encountered was the model's consistent failure to recognize and utilize the manually segmented ground truth labels, which resulted in errors such as "Cannot find true value in the ground truth." This mis-

match between the pre-trained model’s general capabilities and the task-specific requirements of pituitary gland segmentation led to poor performance, as evidenced by the testing results. Despite multiple preprocessing steps, including resampling, normalization, and data augmentation, the model struggled to generate accurate segmentation outputs. The failure underscores the limitations of applying a general-purpose model to specialized medical imaging tasks.

A critical issue was the reliance on a pre-trained model rather than training a custom solution from scratch. Pre-trained models, such as SAM-Med3D, are typically trained on diverse datasets to generalize across a wide range of tasks. However, medical tasks like pituitary gland segmentation are highly specific, requiring domain-specific training data. As the project progressed, it became clear that the pre-trained model lacked the ability to generalize well to this niche task, and the absence of a specialized dataset for pituitary gland segmentation further compounded the problem.

This failure to meet the original objective highlights an important insight: the success of deep learning models in medical imaging is heavily dependent on high-quality, task-specific annotated datasets. The decision to rely on SAM-Med3D, while initially pragmatic due to time and resource constraints, proved to be a limitation, as the model was not adequately fine-tuned or retrained for the specific medical task at hand.

From a project management perspective, the time allocated for troubleshooting and resolving the technical issues proved insufficient. A significant portion of the project timeline was devoted to resolving issues related to ground truth mismatch, label recognition, and model inference. These delays hindered the exploration of alternative approaches, such as training a custom model or employing semi-supervised learning techniques. Had more time been available, the project could have benefited

from focusing on creating a custom-trained model using dedicated pituitary gland datasets.

Despite these challenges, the dissertation contributed valuable insights into the application of deep learning models in medical imaging. It highlighted the importance of matching the model’s training data with the specific task and underscored the limitations of pre-trained solutions for specialized medical applications. The persistent errors encountered during testing, particularly with the pre-trained model’s inability to correctly interpret the manually segmented ground truth, emphasize the need for better dataset preparation and model adaptation techniques.

Looking at the testing results, it became evident that the model struggled to produce meaningful segmentations. The outputs were often inaccurate, with poor performance metrics and inconsistent visual results. These failures occurred despite multiple rounds of preprocessing and optimization attempts, reinforcing the conclusion that the SAM-Med3D model was not well-suited for this specific task without further customization. The model’s general-purpose nature clashed with the niche requirements of the pituitary gland segmentation task, a finding that will inform future research in this domain.

Although the original aim of achieving fully automated, accurate pituitary gland segmentation was not met, the dissertation’s findings serve as an important case study in the limitations of pre-trained models. This work underscores the need for custom-trained models in specialized domains like medical imaging and highlights the challenges of adapting existing solutions to new tasks.

The failure to achieve the desired outcome, while disappointing, provides a critical lesson for future projects: the necessity of domain-specific training data and task-specific model customization. In the case of pituitary gland segmentation, relying on a pre-trained model without further fine-tuning or retraining proved insufficient,

as the model lacked the specificity needed for accurate segmentation. This lesson will guide future research, which should prioritize obtaining high-quality annotated datasets and explore custom training strategies from the outset.

In conclusion, this dissertation has shown that while pre-trained models offer a promising starting point for many applications, their general-purpose nature limits their effectiveness in specialized tasks such as pituitary gland segmentation. The original aims were not fully realized, but the project yielded valuable insights into the complexities of model adaptation, dataset preparation, and medical imaging tasks. With more time and resources, future work should focus on building custom models tailored to the specific needs of medical imaging, using dedicated, annotated datasets to improve segmentation accuracy and performance. This reflection on the project's outcome highlights both the technical challenges and the learning opportunities that emerged, providing a strong foundation for continued exploration in this field.

Appendix A

Source code for Dissertation

```
1 import torch
2 print(torch.cuda.is_available())
3 import torch
4 print(torch.cuda.is_available())
5 !git clone https://github.com/uni-medical/SAM-Med3D.git
6 %cd SAM-Med3D
7 !mkdir -p checkpoints
8 !wget https://huggingface.co/blueyo0/SAM-Med3D/resolve/
   main/sam_med3d_turbo.pth -P checkpoints/
9
10 import torchio as tio
11 import nibabel as nib
12
13 def calculate_spacing(original_shape, target_shape,
   original_spacing):
14     """
```

```

15     Calculate the new spacing required to resample an
        image to a target shape.
16     """
17     ratios = [orig_dim / target_dim for orig_dim,
                target_dim in zip(original_shape, target_shape)]
18     new_spacing = [orig_space * ratio for orig_space,
                    ratio in zip(original_spacing, ratios)]
19     return tuple(new_spacing)
20
21 def resample_nii(input_img_path, input_label_path,
                output_img_path, output_label_path, target_shape=(512,
                512, 82)):
22     # Load image and label
23     image = tio.ScalarImage(input_img_path)
24     label = tio.LabelMap(input_label_path)
25
26     # Get original shape and spacing
27     original_shape = image.spatial_shape # (Height,
        Width, Depth)
28     original_spacing = image.spacing # Voxel size in mm
29
30     # Calculate new spacing based on target shape
31     new_spacing = calculate_spacing(original_shape,
        target_shape, original_spacing)
32
33     # Resample to target spacing

```

```

34     resample_transform = tio.Resample(new_spacing)
35
36     # Apply resampling
37     resampled_image = resample_transform(image)
38     resampled_label = resample_transform(label)
39
40     # Save resampled images
41     resampled_image.save(output_img_path)
42     resampled_label.save(output_label_path)
43
44     print(f"Resampled image saved to: {output_img_path}")
45     print(f"Resampled label saved to: {output_label_path
46           }")
47 # Example
48 input_img_path = './test_data/kidney_right/AMOS/imagesVal
49     /amos_0013.nii.gz' # Input image file path
50 input_label_path = './test_data/kidney_right/AMOS/
51     labelsVal/amos_0013.nii.gz' # Input label file path
52 output_img_path = './test_data/kidney_right/AMOS/pred/
53     amos_0013i.nii.gz' # Output resampled image path
54 output_label_path = './test_data/kidney_right/AMOS/pred/
55     amos_0013l.nii.gz' # Output resampled label path
56
57 # Resample both image and label to the target shape

```

```

54 resample_nii(input_img_path, input_label_path,
    output_img_path, output_label_path, target_shape=(512,
    512, 82))
55
56 # Load and check the shape of the resampled image and
    label
57 resampled_img = nib.load(output_img_path)
58 resampled_label = nib.load(output_label_path)
59
60 print("Resampled image shape:", resampled_img.shape)
61 print("Resampled label shape:", resampled_label.shape)
62
63 import torch
64 import medim
65
66 # Define the path to the model weights
67 checkpoint_path = 'checkpoints/sam_med3d_turbo.pth'
68
69 # Create the SAM-Med3D model
70 model = medim.create_model("SAM-Med3D", pretrained=True,
    checkpoint_path=checkpoint_path)
71
72 # Make sure to move the model to the appropriate device
73 device = torch.device('cuda' if torch.cuda.is_available()
    else 'cpu')
74 model = model.to(device)

```

```

75
76 # Set the model to evaluation mode for inference
77 model.eval()
78
79 print("Model loaded successfully!")
80
81 import nibabel as nib
82 import matplotlib.pyplot as plt
83
84 # Load the predicted segmentation result
85 predicted_segmentation_path = './test_data/kidney_right/
      AMOS/imagesVal/amos_0013.nii.gz'
86 predicted_img = nib.load(predicted_segmentation_path)
87 predicted_data = predicted_img.get_fdata()
88
89 # Load the ground truth image
90 ground_truth_path = './test_data/kidney_right/AMOS/
      labelsVal/amos_0013.nii.gz'
91 gt_img = nib.load(ground_truth_path)
92 gt_data = gt_img.get_fdata()
93
94 # Plot both ground truth and predicted segmentation side
      by side
95 fig, ax = plt.subplots(1, 2, figsize=(12, 6))
96
97 # Show Ground Truth

```

```

98 ax[0].imshow(gt_data[:, :, gt_data.shape[2] // 2], cmap='
    gray')
99 ax[0].set_title('Ground Truth')
100
101 # Show Predicted Segmentation
102 ax[1].imshow(predicted_data[:, :, predicted_data.shape[2]
    // 2], cmap='gray')
103 ax[1].set_title('Predicted Segmentation')
104
105 plt.show()
106
107 import nibabel as nib
108
109 # Load the original image
110 original_img_path = './test_data/kidney_right/AMOS/
    imagesVal/amos_0013.nii.gz'
111 original_img = nib.load(original_img_path)
112 original_img_data = original_img.get_fdata()
113
114 # Load the original label
115 original_label_path = './test_data/kidney_right/AMOS/
    labelsVal/amos_0013.nii.gz'
116 original_label = nib.load(original_label_path)
117 original_label_data = original_label.get_fdata()
118
119 # Print shapes

```

```
120 print(f"Original image shape: {original_img_data.shape}")
121 print(f"Original label shape: {original_label_data.shape
    }")
122
123 # Example to preprocess, infer, and save
124 !python medim_infer.py --input ./test_data/kidney_right/
    AMOS/imagesVal/amos_0013.nii.gz --output ./test_data/
    kidney_right/AMOS/pred/
```

Bibliography

- Barkhoudarian, G and DF Kelly (2017). “The pituitary gland: anatomy, physiology, and its function as the master gland”. In: *Cushing’s Disease*. Elsevier, pp. 1–41.
- Lakanwal, Qaiswer Shah (2023). “Anatomy and physiology of pituitary gland”. In: *IJO-International Journal of Health Sciences and Nursing (ISSN: 2814-2098)* 6.01, pp. 46–56.
- University, Oregon State (2023). *The Pituitary Gland and Hypothalamus*. Ed. by Jody E. Johnson. Accessed: 2024-09-04. URL: <https://open.oregonstate.edu/education/aandp/chapter/17-3-the-pituitary-gland-and-hypothalamus/>
- Soliman, Ashraf et al. (2024). “The impact of growth hormone (GH) therapy on glucose metabolism: A narrative review mainly focused on GH-deficient (GHD) children and adolescents”. In: *World Journal of Advanced Research and Reviews* 22.1, pp. 627–636.
- Weinberg, Jennifer L (2024). “ACTH Hormone: Roles, Regulation, and Health Implications”. In: *Endocrinology*.
- Zwahlen, Jann et al. (2024). “The ecological function of thyroid hormones”. In: *Philosophical Transactions of the Royal Society B* 379.1898, p. 20220511.
- Athar, Faria, Muskan Karmani, and Nicole M Templeman (2024). “Metabolic hormones are integral regulators of female reproductive health and function”. In: *Bioscience Reports* 44.1, BSR20231916.
- Nogueira-Vale, Eliana (2024). “Oxytocin”. In: *Oxytocin, Well-Being and Affect Regulation*, pp. 37–49.
- Calvi, Anna et al. (2024). “Urinary hyaluronidase activity is closely related to vasopressinergic system following an oral water load in men: a potential role in blood pressure regulation and early stages of hypertension development”. In: *Frontiers in Endocrinology* 15, p. 1346082.
- ResearchGate (2024a). *Trophic Hormones, Cell Types of the Anterior Pituitary, and Their Hormonal Secretions*. Accessed: 2024-09-04. URL: https://www.researchgate.net/figure/Trophic-hormones-cell-types-of-the-anterior-pituitary-and-their-hormonal-secretions_fig1_8184555.
- Reddy, S Sethu K, Pruthvi Goparaju, and Maria Fleseriu (2024). “Hypothalamic-Pituitary Disorders”. In: *Atlas of Clinical Endocrinology and Metabolism*. CRC Press, pp. 124–146.

- Corsello, Andrea, Rosa Maria Paragliola, and Roberto Salvatori (2024). “Diagnosing and treating the elderly individual with hypopituitarism”. In: *Reviews in Endocrine and Metabolic Disorders* 25.3, pp. 575–597.
- Al-Hadlaq, Malak and Herve Sroussi (2024). “Acromegaly: Overview and associated temporomandibular joint disorders”. In: *Oral Diseases*.
- Fion, L, M Saieehwaran, and R Subashini (2024). “CUSHING’S DISEASE AND THROMBOSIS: A CLINICAL PERSPECTIVE”. In: *Journal of the ASEAN Federation of Endocrine Societies* 39.S1, pp. 80–81.
- Dongre, Nilima (2023). “Perspective Chapter: Stimulation and Activation of the Pituitary Gland”. In: *The Pituitary Gland-An Overview of Pathophysiology and Current Management Techniques*. IntechOpen.
- Lv, Xianli (2024). *Frontiers in Neuroimaging*. BoD–Books on Demand.
- Jipa, Andrei and Vikas Jain (2021). “Imaging of the sellar and parasellar regions”. In: *Clinical imaging* 77, pp. 254–275.
- Ma, Xinyi (2024). “Application And Comparison of Different Medical Imaging Techniques”. In: *Highlights in Science, Engineering and Technology* 112, pp. 261–265.
- Mittendorff, Lisa, Adrienne Young, and Jenny Sim (2022). “A narrative review of current and emerging MRI safety issues: What every MRI technologist (radiographer) needs to know”. In: *Journal of medical radiation sciences* 69.2, pp. 250–260.
- Ray, Roshan (2020). “Evaluation of Size of Normal Adult Pituitary Gland on Magnetic Resonance Imaging”. PhD thesis. Rajiv Gandhi University of Health Sciences (India).
- Chrysikopoulos, Haris et al. (2020). *Errors in Imaging*. Springer.
- Galldiks, Norbert et al. (2024). “Challenges, limitations, and pitfalls of PET and advanced MRI in patients with brain tumors: A report of the PET/RANO group”. In: *Neuro-oncology*, noae049.
- Beheshti, Mohsen et al. (2020). “PET/CT and PET/MRI, normal variations, and artifacts”. In: *Clinical Nuclear Medicine*, pp. 549–584.
- Weiskopf, Nikolaus et al. (2021). “Quantitative magnetic resonance imaging of brain anatomy and in vivo histology”. In: *Nature Reviews Physics* 3.8, pp. 570–588.
- ResearchGate (2024b). *Samples of Noise and Motion Artifacts in MRI*. Accessed: 2024-09-04. URL: https://www.researchgate.net/figure/Samples-of-noise-and-motion-artifacts-in-MRI_fig1_356027167.
- Gillett, Daniel, James MacFarlane, et al. (2023). “Molecular imaging of pituitary tumors”. In: *Seminars in Nuclear Medicine*. Vol. 53. 4. Elsevier, pp. 530–538.
- Samarasinghe, Shanika, Mary Ann Emanuele, and Alaleh Mazhari (2014). “Neurology of the pituitary”. In: *Handbook of clinical neurology* 120, pp. 685–701.
- Sydney, Guy I et al. (2021). “The effect of pituitary gland disorders on glucose metabolism: From pathophysiology to management”. In: *Hormone and Metabolic Research* 53.01, pp. 16–23.
- ResearchGate (n.d.). *MRI Studies of the Patient’s Pituitary Gland*. Retrieved June 29, 2024, from <https://www.researchgate.net/figure/MRI-Studies-of->

- the-Patients-Pituitary-Gland-MRI-of-the-pituitary-gland-showing-a_fig6_6465911. Accessed: June 29, 2024.
- Dennison, Sophie and Pietro Saviano (2018). “Diagnostic imaging”. In: *CRC handbook of marine mammal medicine*. CRC Press, pp. 537–552.
- Wang, Lei et al. (2016). “Principles and methods for automatic and semi-automatic tissue segmentation in MRI data”. In: *Magnetic Resonance Materials in Physics, Biology and Medicine* 29, pp. 95–110.
- Egger, Jan et al. (2012). “Pituitary adenoma volumetry with 3D Slicer”. In: *PloS one* 7.12, e51788.
- Isensee, Fabian, Jens Petersen, et al. (2018). “nnu-net: Self-adapting framework for u-net-based medical image segmentation”. In: *arXiv preprint arXiv:1809.10486*.
- Du, Getao et al. (2020). “Medical Image Segmentation based on U-Net: A Review.” In: *Journal of Imaging Science & Technology* 64.2.
- Hussain, Shah et al. (2022). “Modern diagnostic imaging technique applications and risk factors in the medical field: a review”. In: *BioMed research international* 2022.1, p. 5164970.
- Anastassiadis, Chloe, Sherri Lee Jones, and Jens C Pruessner (2019). “Imaging the pituitary in psychopathologies: a review of in vivo magnetic resonance imaging studies”. In: *Brain Structure and Function* 224.8, pp. 2587–2601.
- Černý, Martin et al. (2023). “Fully automated imaging protocol independent system for pituitary adenoma segmentation: a convolutional neural network—based model on sparsely annotated MRI”. In: *Neurosurgical Review* 46.1, p. 116.
- Saha, Ashirbani et al. (2020). “Machine learning applications in imaging analysis for patients with pituitary tumors: a review of the current literature and future directions”. In: *Pituitary* 23, pp. 273–293.
- Deeley, MA et al. (2011). “Comparison of manual and automatic segmentation methods for brain structures in the presence of space-occupying lesions: a multi-expert study”. In: *Physics in Medicine & Biology* 56.14, p. 4557.
- Hendee, William R (1989). “Cross sectional medical imaging: a history.” In: *Radio-graphics* 9.6, pp. 1155–1180.
- Singh, Mahender Kumar and Krishna Kumar Singh (2021). “A review of publicly available automatic brain segmentation methodologies, machine learning models, recent advancements, and their comparison”. In: *Annals of Neurosciences* 28.1-2, pp. 82–93.
- Tingelhoff, Kathrin et al. (2008). “Analysis of manual segmentation in paranasal CT images”. In: *European archives of oto-rhino-laryngology* 265, pp. 1061–1070.
- Veiga-Canuto, Diana et al. (2022). “Comparative multicentric evaluation of inter-observer variability in manual and automatic segmentation of neuroblastic tumors in magnetic resonance images”. In: *Cancers* 14.15, p. 3648.
- Itri, Jason N et al. (2018). “Fundamentals of diagnostic error in imaging”. In: *Radiographics* 38.6, pp. 1845–1865.

- Wang, He et al. (2021). “Development and evaluation of deep learning-based automated segmentation of pituitary adenoma in clinical task”. In: *The Journal of Clinical Endocrinology & Metabolism* 106.9, pp. 2535–2546.
- Gillett, Daniel, Waiel Bashari, et al. (2021). “Methods of 3D printing models of pituitary tumors”. In: *3D Printing in Medicine* 7, pp. 1–14.
- Chahal, Prabhjot Kaur, Shreelekha Pandey, and Shivani Goel (2020). “A survey on brain tumour detection techniques for MR images”. In: *Multimedia Tools and Applications* 79.29, pp. 21771–21814.
- Fan, Jianping et al. (2005). “Seeded region growing: an extensive and comparative study”. In: *Pattern recognition letters* 26.8, pp. 1139–1156.
- Chen, Xinjian and Lingjiao Pan (2018). “A survey of graph cuts/graph search based medical image segmentation”. In: *IEEE reviews in biomedical engineering* 11, pp. 112–124.
- Safari, Aref (2020). “A Novel Transformation Watershed Image Segmentation Model in Digital Elevation Maps Processing”. In: *Journal of Modern Processes in Manufacturing and Production* 9.3, pp. 13–22.
- Chugh, Samiksha and S Mahesh Anand (2012). “Semi-Automated Tumor Segmentation from MRI Images Using Local Statistics Based Adaptive Region Growing”. In: *International Journal of Information and Electronics Engineering* 2.1, pp. 7–11.
- Sun, Min et al. (2017). “Random walk and graph cut based active contour model for three-dimension interactive pituitary adenoma segmentation from MR images”. In: *Medical Imaging 2017: Image Processing*. Vol. 10133. SPIE, pp. 548–555.
- Zukic, Dženan et al. (2011). “Preoperative volume determination for pituitary adenoma”. In: *Medical Imaging 2011: Computer-Aided Diagnosis*. Vol. 7963. SPIE, pp. 817–823.
- Almahfud, Mustofa Alisahid et al. (2018). “An effective MRI brain image segmentation using joint clustering (K-Means and Fuzzy C-Means)”. In: *2018 International seminar on research of information technology and intelligent systems (ISRITI)*. IEEE, pp. 11–16.
- Choi, Uk-Su, Yul-Wan Sung, and Seiji Ogawa (2024). “deepPGSegNet: MRI-based pituitary gland segmentation using deep learning”. In: *Frontiers in Endocrinology* 15, p. 1338743.
- Tzanou, Maria (2023). *Health Data Privacy Under the GDPR*. TAYLOR FRANCIS Limited.
- Kawamleh, Suzanne (2023). “Against explainability requirements for ethical artificial intelligence in health care”. In: *AI and Ethics* 3.3, pp. 901–916.
- Regency Medical Centre (2023). *CT Scan vs MRI: Understanding the Differences*. Accessed: 2024-09-07. URL: <https://www.regencymedicalcentre.com/blog/ct-scan-vs-mri/>.
- Nyúl, László G, Jayaram K Udupa, and Xuan Zhang (2000). “New variants of a method of MRI scale standardization”. In: *IEEE transactions on medical imaging* 19.2, pp. 143–150.

- Perez, Fernando and Brian E. Granger (2007). “IPython: A System for Interactive Scientific Computing”. In: *Computing in Science Engineering* 9.3, pp. 21–29. DOI: 10.1109/MCSE.2007.53.
- Zayniddinov, XN et al. (2024). “TOOLS FOR MANUAL 3D MEDICAL IMAGE SEGMENTATION AND DATA PREPROCESSING FOR AUTOMATIC 3D MEDICAL IMAGE SEGMENTATION”. In: *International journal of scientific researchers (IJSR) INDEXING* 5.1, pp. 233–237.
- Isensee, Fabian, Paul F Jaeger, et al. (2021). “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation”. In: *Nature methods* 18.2, pp. 203–211.
- Wang, Haoyu et al. (2023). *SAM-Med3D*. arXiv: 2310.15161 [cs.CV].
- Fedorov, Andriy et al. (2012). “3D Slicer as an image computing platform for the Quantitative Imaging Network”. In: *Magnetic resonance imaging* 30.9, pp. 1323–1341.
- 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: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III* 18. Springer, pp. 234–241.
- Ullah, Muhammad Farhat, Ali Saeed, and Naveed Hussain (2023). “Comparison of Pre-trained vs Custom-trained Word Embedding Models for Word Sense Disambiguation”. In: *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal* 12.1, e31084–e31084.