Proposal for master's thesis

Foundation Models for Medical Segmentation

Author:	Rawan Mousa
Student ID:	2310854019
Study program:	Data Science
Advisor:	DI Dr. techn. Matthias Blaickner
Version:	3.0
Date:	1 December 2024

FH University of Applied Sciences TECHNIKUM WIEN

Problem area

Medical image segmentation is an important process in the field of medical diagnostics and treatment planning. Accurate segmentation of medical images such as MRI, CT scans and X-rays is required to identify anatomical structures, detect abnormalities and enable precise treatment actions that are difficult to detect in conventional treatments.

In recent years, numerous advances have been made in medical image segmentation using machine learning, especially deep learning. Olaf et al. [1] proposed a deep learning method using skip-connection convolutional neural networks (CNNs) to segment medical images. Based on the U-Net, Xiao et al. [2] and Guan et al. [3] developed the ResUnet and DenseUnet to improve the segmentation performance by modifying the backbone network. In addition, the developments in the field of Vision Transformer (ViT) [4] in deep learning led Chen et al. [5] to propose the Tran-Unet and Hu et al. [6] to propose the Swin-Unet. With the advancement of devices, more and more foundation models such as GPT and Segment Anything (SAM) [7, 8] are emerging, which are trained on very large datasets and therefore provide good results for various tasks.

In this master thesis, it is planned to apply three to four foundation models, namely 3D Segmentation [9], SAM [8] MedSAM and Encoder [10], on various medical image data such as MRI and CT scans. The aim is to investigate the performance of these models in segmenting medical images. The obtained results will then be compared with the Medical Volume Annotator (MVA) [11] annotation data to evaluate the accuracy and reliability of the models and to determine their suitability for clinical use.

Data description

The master's thesis employs medical imaging data to investigate deep learning techniques for segmentation. The data includes CT scans of anonymized patients provided by a hospital in Portugal and two additional datasets from the Medical Decathlon [12] repository: the Heart Dataset and the Spleen Dataset.

Dataset Details

Portugal Dataset:

- Target: upper respiratory tract
- Imaging Modality: CT
- Number of Study Folders: 110
- Total Number of Images: 818
- Patient Age Range: 1 to 78 years (average age 34 years)
- Gender Distribution: 55% male, 45% female
- Valid Images After Preprocessing: 570
- Image Resolution: 512x512 pixels
- Image Planes: Axial (286), Coronal (134), Sagittal (132)
- Slices per Image: Varies from a few to 600, with an average of 52
- Slice Thickness: 0.2 mm to 5 mm, averaging 2.47 mm
- Pixel Spacing: 0.1 mm to 0.9 mm, averaging 0.44 mm

Heart Dataset (Medical Decathlon):

- Target: Left Atrium
- Imaging Modality: Mono-modal MRI
- Number of Patients: 30 (20 training, 10 testing)
- Total Slices: 3,568 (2,271 training, 1,297 testing)
- Resolution: 256x256 pixels
- Slice Thickness: 1.5 mm
- Source: King's College London
- Challenge: Small training dataset with large variability
- Clinical Relevance: Focused on segmenting the left atrium for cardiac analysis
- Labels:
    - 0: Background
    - 1: Left Atrium

Spleen Dataset (Medical Decathlon):

- Target: Spleen
- Imaging Modality: CT
- Number of Patients: 61 (41 training, 20 testing)

- Total Slices: 5,277 (3,650 training, 1,627 testing)
- Resolution: 512x512 pixels
- Slice Thickness: 2.5 mm
- Source: Memorial Sloan Kettering Cancer Center
- Challenge: Large-ranging foreground size
- Clinical Relevance: Used for identifying spleen boundaries to aid in organ volume calculations
- Labels:
  - 0: Background
  - 1: Spleen

Data Preparation and Exclusion:

1. **Portugal Dataset:**
   - Exclusion Criteria:
     - Images with fewer than 20 slices
     - Missing labels
     - Duplicates
     - Calibration images
     - Insufficient airway coverage
     - Artifacts (e.g., metallic implants)
     - Patients under 16 years
     - Corrupted files
   - Labeling Methods: Manual outlining and semi-automatic level tracing functionality in 3D Slicer for segmenting regions of interest.

2. **Heart Dataset:**
   - All valid slices with complete annotations were retained.
   - Labeling Methods: Manual expert annotations focusing on the left atrium and myocardium

3. **Spleen Dataset:**
   - All valid slices with complete annotations were retained.
   - Labeling Methods: Semi-automated annotations validated by radiologists.

Data Augmentation:

- Methods: 90-degree rotation, horizontal and vertical flipping.
- Objective: To enhance dataset size and model quality.

Training and Testing:

- Training/Test Split:
  - Portugal Dataset: 154 for training (81%), 35 for testing (19%).

    - Heart and Spleen Datasets: 80/20 ratio for using and testing, respectively.
- Normalization: Applied min-max normalization to intensity values (pixel values) ranging from [0, 1].
- Evaluation Metrics: Dice Similarity Coefficient (DSC) and Hausdorff **distance**.

The combination of datasets from Portugal and the Medical Decathlon ensures a diverse and comprehensive foundation for evaluating Foundation models, allowing for robust analysis across different modalities and anatomical targets.

Research questions / hypotheses

1. How can the performance of foundation models be optimized for the segmentation of medical images?

2. What is the impact of different foundation models on the accuracy, efficiency, and reliability of medical image segmentation?

3. How do the results of these models compare to the annotations provided by the Medical Volume Annotator (MVA)?

Hypotheses:

- Foundation models will significantly improve the accuracy and efficiency of medical image segmentation.

- Different foundation models will exhibit varying levels of performance, with some models excelling in specific aspects of segmentation.

- The comparison with MVA annotations will highlight the strengths and limitations of each model.

Methods

- Data Collection and Preprocessing: Collect and preprocess the medical image data, including normalization and augmentation.

- Model: execute various foundation models (3D Segmentation, SAM, MedSAM, Encoder) on the datasets.

- Evaluation: Evaluate the models using metrics such as Dice Similarity Coefficient and Hausdorff Distance.

- Comparison: Compare the segmentation results with MVA annotations to assess model performance.

Expected results

- A comprehensive evaluation of foundation models for medical image segmentation.

- Insights into the performance of different models and their applicability to various medical imaging tasks.

- High-resolution segmented images that can be used for further research and clinical applications.

Overview Research questions – methods – expected results

Research questions/ hypotheses	*Method(s) per research question*	*Expected kind of result per method*
Performance optimization of foundation models	Model Evaluation	Improved segmentation accuracy and efficiency
Impact of different models on segmentation accuracy	Comparative Analysis	Varying performance metrics
Comparison with MVA annotations	Benchmarking and Validation	Strengths and limitations of each model

Timetable

Milestone plan

Nr	Name	Plan	Adapted by	Actual date
M1	Master thesis started	02.05.2024		
M2	Initial literature research and review	15.05.2024	01.12.2024	
M3	Proposal v1.0	01.06.2024	15.06.2024	
M4	Final Proposal approved	09.09.2024	01.12.2024	
M5	Intermediate appointment held with advisor	16.09.2024		
M6	Data Exploration	17.09.2024	05.11.2024	
M7	Data prepared for Segmentation	01.11.2024	01.12.2024	
M8	Start with the writing phase	01.02.2025		
M9	Models Evaluation	31.03.2025		
M10	Comparative Analysis done	14.04.2025		
M11	Comparison with MVA annotations	21.04.2025		
M12	Complete first version handed in	27.04.2025		
M13	Edit the first full version	31.05.2025		
M14	Final version handed in	01.06.2025		

Project Milestones Gantt Chart

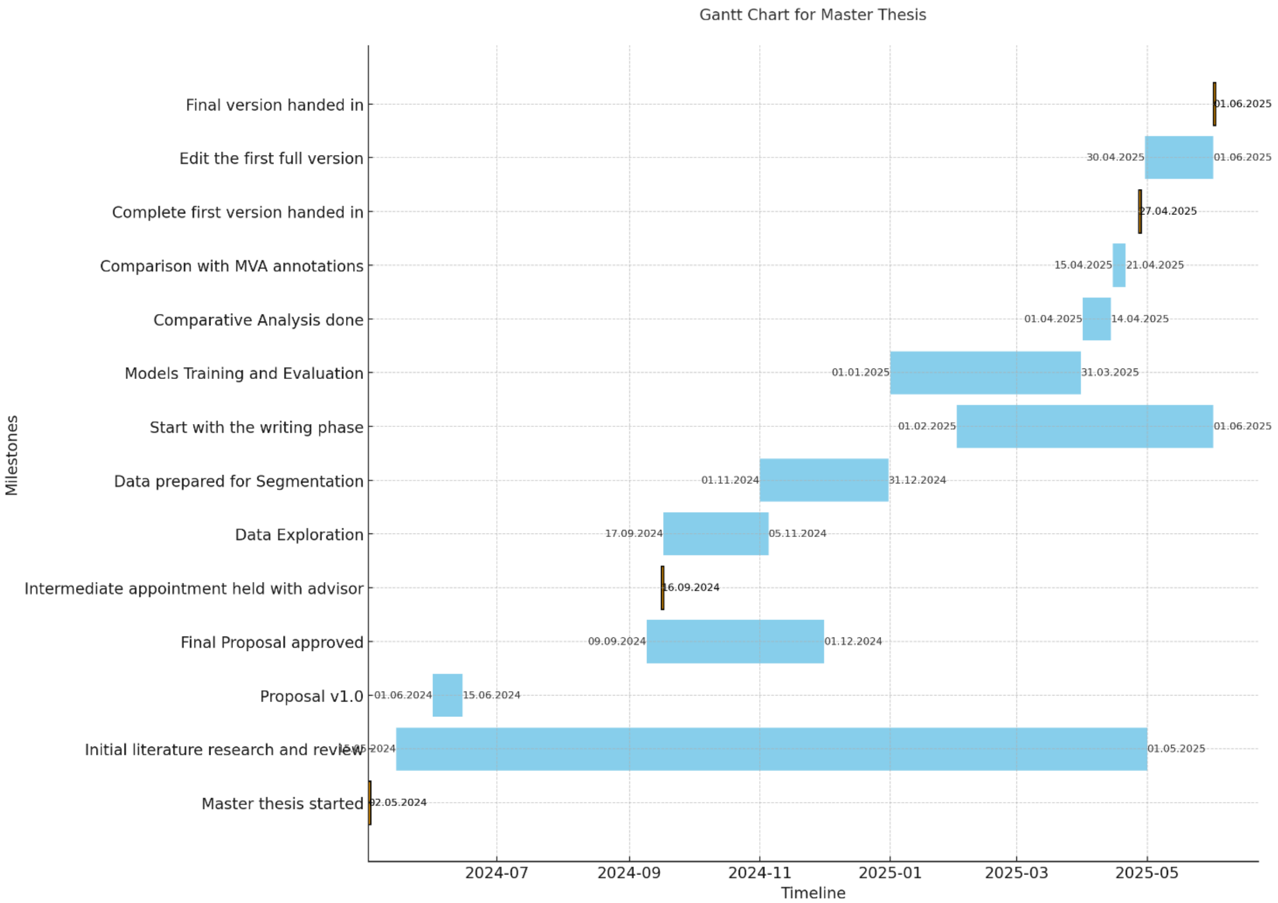

Table of contents of the planned work

- Title Page
- Declaration of authenticity
- Abstract
  - Motivation
  - problem description
  - solution/approach
  - results
  - consequences
- Acknowledgements
- Table of contents
- Introduction
  - Topic
  - Research question
  - Aim of thesis
  - Structure of the work
- First chapter: Literature review
  - Foundation Models
  - Types of Medical images
  - Medical images Segmentation
- Second chapter: Methods
- Third chapter: Main part
  - Preparing the data
  - Cleaning the data
  - Models Evaluation
  - Comparison of the Foundation Models
  - Comparison with MVA
- Discussion
- Conclusion – Summary – Outlook
- Appendix and indexes

Literature

[1] Ronneberger, O., Fischer, P. and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lecture Notes in Computer Science*, 9351, pp.234–241. doi: https://doi.org/10.1007/978-3-319-24574-4_28

[2] X. Xiao, S. Lian, Z. Luo and S. Li, "Weighted Res-UNet for High-Quality Retina Vessel Segmentation," 2018 9th International Conference on Information Technology in Medicine and Education (ITME), Hangzhou, China, 2018, pp. 327-331, doi: 10.1109/ITME.2018.00080.

[3] S. Guan, A. A. Khan, S. Sikdar and P. V. Chitnis, "Fully Dense UNet for 2-D Sparse Photoacoustic Tomography Artifact Removal," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 2, pp. 568-576, Feb. 2020, doi: 10.1109/JBHI.2019.2912935.

[4] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. and Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv:2010.11929 [cs]. [online] Available at: https://arxiv.org/abs/2010.11929.

[5] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L. and Zhou, Y. (2021). TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. arXiv:2102.04306 [cs]. [online] Available at: https://arxiv.org/abs/2102.04306.

[6] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q. and Wang, M. (2021). Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. arXiv:2105.05537 [cs, eess]. [online] Available at: https://arxiv.org/abs/2105.05537.

[7] Floridi, L., Chiriatti, M. GPT-3: Its Nature, Scope, Limits, and Consequences. Minds & Machines 30, 681–694 (2020). https://doi-1org-1000342a419c3.han.technikum-wien.at/10.1007/s11023-020-09548-1

[8] Dong, G., Wang, Z., Chen, Y. et al. An efficient segment anything model for the segmentation of medical images. Sci Rep 14, 19425 (2024). https://doi-1org-1000342a419a9.han.technikum-wien.at/10.1038/s41598-024-70288-8

[9] Wasserthal, J. (2023). TotalSegmentator. [online] GitHub. Available at: https://github.com/wasserth/TotalSegmentator.

[10] Pai S, Bontempi D, Prudente V, Hadzic I, Sokač M, Chaunzwa TL, Bernatz S, Hosny A, Mak RH, Birkbak NJ, Aerts HJ. Foundation Models for Quantitative Biomarker Discovery in Cancer Imaging. medRxiv [Preprint]. 2023 Sep 5:2023.09.04.23294952. doi: 10.1101/2023.09.04.23294952. PMID: 37732237; PMCID: PMC10508804.

[11] www.mva.ai. (2023). MVA.ai – Medical Volume Annotator. [online] Available at: https://www.mva.ai/ [Accessed 15 Sep. 2024].

[12] Medical Decathlon, "Medical Decathlon," [Online]. Available: http://medicaldecathlon.com/. [Accessed: Dec. 1, 2024].