SegEVOLution: Enhanced Medical Image Segmentation with Multimodality Learning

Matey Krastev, Miklos Hamar, Serghei Mihailov, Zsombor Fulop, Danilo Toapanta

Deep Learning 2 (2023/2024)
Supervisor: Stefanos Achlatis
UvA

Background

SegVol [1] is a cutting-edge **3D segmentation model**, excelling in medical image benchmarks. It enables **universal** and **interactive segmentation** by integrating a ViT backbone with text, bounding box, and point prompts. Its success is partly due to extensive pre-training on **96,000 unlabelled CT** volumes and fine-tuning on **6,000 labelled CT** volumes.

Mixture-of-Adapters

Mixture-of-Adapters (MoA) [2] utilizes multiple lightweight adapter modules within a model to handle diverse tasks and modalities. We apply a *top-1* gated mixture combining an ***identity adapter (CT)*** with ***LoRA adapter (MRI)*** to guarantee:

- **Preserved base performance**
- **Almost no additional runtime costs**

Context-Aware Priors

Taking inspiration from HERMES [3], we apply **prior fusion** to infuse the hidden representations with task- and modality-specific information. We incorporate:

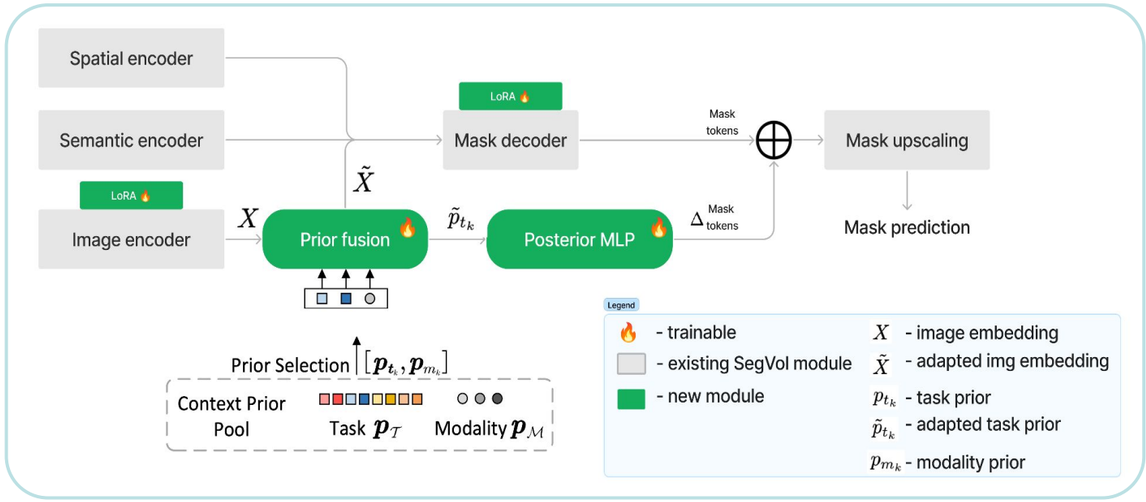
- **Context prior pool.** Priors adapt the image representation to the specified task and modality, essentially conditioning the segmentation.
- **Posterior prototype.** We use the posterior tokens obtained from prior fusion to adapt the mask decoder output.

Proposed Architecture

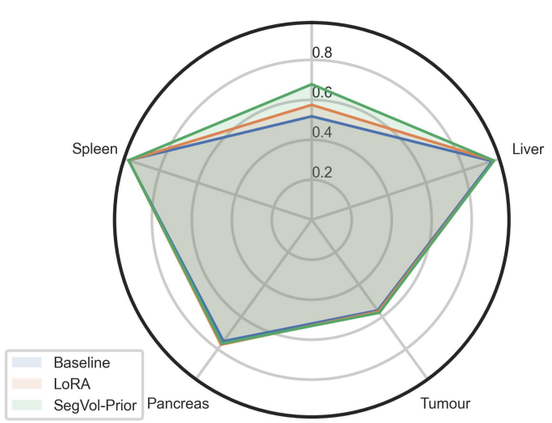
- We extend the base SegVol architecture by introducing a **prior fusion layer** – a self- and cross-attention block that introduces **biases** that provide a more robust mask decoding strategy.
- We combine them with **learnable positional embeddings** in order to encode prior source (task, modality, image, spatial prompts).
- Finally, we apply an **MLP that computes "posterior"** for the generated tokens.

Discussion

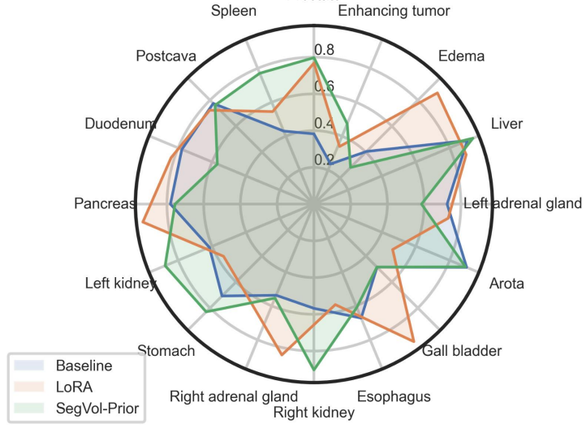
- **The Fusion embedding and finetuning shows clear improvement over the SegVol baseline**
- **Introduction of Context Priors enhances performance even further**
- **Good performance on MRI data**
- **The high-dimensional space of the prior fusion self-attention block shows high-correlation between similar features.**

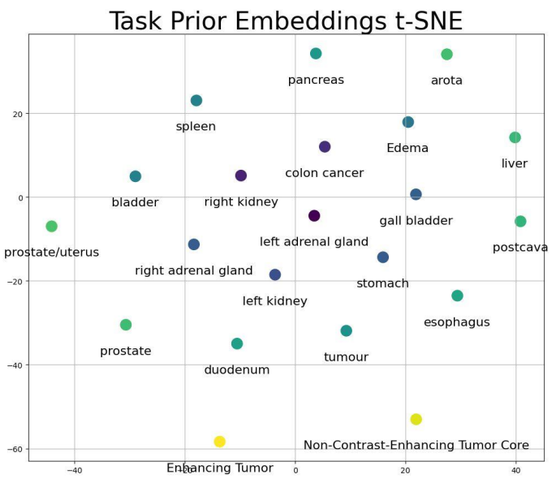


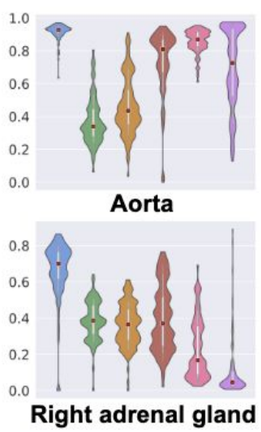
Model performance on the 400 dataset. (Modality: CT)

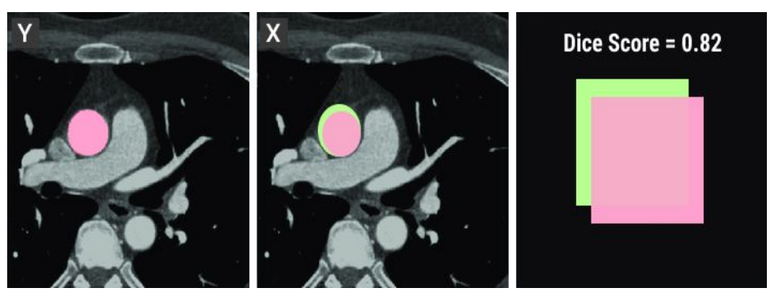


Model performance on the 400 dataset. (Modality: MRI)









Limitations & Future Work

- **Limited amount of MRI data, privacy issues**
- **Class unbalance leads hamper model accuracy**
- **Pseudo masks using Felzenswalb-Huttenlocher are noisy, a better**
- **Contrastive language-medical image training for the medical domain**

SOURCE CODE

References

[1] Yuxin Du et al. SegVol: Universal and Interactive Volumetric Medical Image Segmentation. arXiv 2311.13385v (2024)
[2] Haisong Liu et al. Training Like a Medical Resident: Universal Medical Image Segmentation via Context Prior Learning. arXiv 2306.02416v (2023)