SegMed: Implementation of MedSAM segmentation and enhancement with FeatUp and GAFL

Hasaan Maqsood, Iana Kulichenko, Daniil Volkov, Sergey Egorov

Code Submission

Our code, including the baseline models and additional parts such as data preprocessing, data augmentation, and accuracy metrics, is available at Github URL.

Introduction

Brain tumors critically affect patient health, survival, and quality of life. Manual segmentation of tumors from MRI images, essential for treatment planning, is both slow and prone to errors. Our project aims to automate this process with advanced deep learning techniques, focusing on the development of the MedSAM model, a specialized adaptation of the Segment Anything Model (SAM) for medical imaging. By enhancing the accuracy and efficiency of tumor segmentation, we seek to support more informed clinical decisions, optimize treatment strategies, and ultimately improve patient outcomes.

Dataset Details

We have changed dataset(see Implementation Challenges below)

LGG Segmentation Dataset

The dataset consists of MRI brain images paired with FLAIR abnormality segmentation masks derived from The Cancer Imaging Archive (TCIA). It includes 110 cases from The Cancer Genome Atlas (TCGA), featuring lowergrade glioma patients. Each case includes pre-contrast, FLAIR, and post-contrast sequences, with segmentation masks provided as binary, single-channel images. The dataset is partitioned into training, testing, and validation sets, with 2828, 393, and 708 samples, respectively.

Approach

Preprocessing

Segmentation

- Data Acquisition: MRI scans with corresponding segmentation masks are downloaded from Kaggle.
- One-hot Encoding and Reverse Encoding: Segmentation labels are one-hot encoded for compatibility with neural network models.
- Custom Dataset Class: The LGGDataset class handles image loading, one-hot encoding, and augmentations.

- Image Size Divisibility by 32: Ensuring image dimensions are divisible by 32 for proper alignment of encoder and decoder features.
- Data Augmentation: Techniques include RandomCrop, HorizontalFlip, VerticalFlip, RandomRotate90, CLAHE, RandomBrightnessContrast, RandomGamma, and Normalize.

Training

Segmentation

- Encoder: EfficientNetB5
- Segmentation Models: SAM, U-Net, U-Net++, Pyramid Attention Network (PAN), and DeepLabV3+.
- Loss Function: For SAM Focal loss, for baseline models: Combined Dice loss and Binary Cross-Entropy (BCE) loss.
- Training: Conducted about 5-7 epochs on a CUDA-enabled device, used T4 GPU.

Methodology Adjustments on Med SAM

Model Architecture (SAM)

Vision Transformer (ViT)

The SamModel uses the Vision Transformer (ViT) to process images as sequences of patches. Each image is divided into patches, which are embedded and fed into a transformer encoder composed of self-attention and feed-forward networks.

Encoders

- Vision Encoder: Extracts features from input images using the ViT architecture.
- Prompt Encoder: Encodes auxiliary inputs like bounding boxes to align with the vision encoder's outputs.

Mask Decoder

Generates segmentation masks by decoding features from the vision and prompt encoders. Only the mask decoder's parameters are updated during training.

Evaluation Metrics

- IoU (Intersection over Union): Evaluates segmentation performance, averaged across batches and epochs.

Implementation Challenges and Adjustments

Initial Project and Dataset

Our initial project focused on enhancing image segmentation using the MedSAM model, a derivative of the Segment Anything Model (SAM) tailored for medical imaging. We aimed to achieve high segmentation accuracy and consistency across various imaging modalities, integrating FeatUp for improved feature resolution and GAFL for optimized frequency content. We planned to use the Harvard FairSeg dataset with 10,000 samples, which is significant for studying fairness in medical segmentation.

Challenges Faced

Dataset Size and Processing Constraints

The substantial size of the Harvard-FairSeg10k dataset (45 GB) posed significant challenges for local processing and on platforms like Google Colab. Attempts to upload and process the dataset on available servers also failed due to hardware limitations and resource constraints.

Computational Resource Limitations

Even after reducing the dataset size from 10,000 to 1,000 samples, server failures during model training highlighted the need for a more feasible approach.

That's why we decided to change the dataset on LGG Segmentation Dataset.

Adjustments and New Approach

Dataset Reduction and Alternative Dataset

We pivoted to a more manageable dataset (LGG Segmentation Dataset) due to ongoing computational constraints. This change allowed us to focus on achieving our goals without compromising on the quality of our results.

Implementation of Multiple Models

To adapt, we implemented and evaluated multiple models, exploring various architectures to ensure robust performance on the new dataset.

Preliminary Results

Segmentation

Implemented baseline models DeepLabV3+, Unet, Unet++, Deeplabv3, PAN.

Model	Train loss	Valid Loss	IOU
DeepLabV3+	1.6634	1.6662	0.8741
UNET	1.6691	1.6712	0.8517
UNET++	1.6730	1.6750	0.8268
DeepLabV3	1.6646	1.6700	0.7986
PAN	1.6656	1.6683	0.7970

Implemented Sam model training on 5 epochs. (Mean loss: 0.0054, Mean IoU: 0.644)

Contributions

- Iana Kulichenko: Data preparation, segmentation models development, fine-tuning models, writing reports.
- Hasaan Maqsood: Segmentation models development, writing reports, fine-tuning models, evaluating results.
- Daniil Volkov: FeatUp, GAFL implementation.

Remaining Work

- Fine-tuning Med SAM model to increase the performances.
- Enhancing segmentation mask accuracy, comparison analysis, to combine with Featup and GAFL.
- Conducting extensive testing and validation to ensure robustness and reliability of the proposed methods.

Conclusion

In our quest to automate brain tumor segmentation through deep learning, we've made significant strides by deploying and assessing a range of foundational models such as DeepLabV3+, U-Net, U-Net++, DeepLabV3, and Pyramid Attention Network (PAN). Building on this foundation, we've taken a significant step forward by integrating the SAM model into our framework. After training it for 5 epochs, we've observed encouraging initial results, achieving a mean loss of 0.0054 and a Mean Intersection over Union (IoU) of 0.644. Moving forward, we plan to enhance our MedSAM model and incorporate advanced technologies like FeatUp and GAFL. This integration aims to refine feature resolution and optimize frequency content, respectively, further boosting our model's performance.

References

1. Ma, J., He, Y., Li, F., Han, L., You, C., & Wang, B. (2024). Segment Anything in Medical Images. *Nature Communications*, 15(1), 654. https://doi.org/10.1038/s41467-024-44824-z

2. Xu, D., Zhou, X., Niu, X., & Wang, J. (2020, December). Automatic Segmentation of Low-Grade Glioma in MRI Image Based on UNet++ Model. *Journal of Physics: Conference Series*, 1693(1), 012135. https://dx.doi.org/10.1088/1742-6596/1693/1/012135

3. Hafeez, H. A., Elmagzoub, M. A., Abdullah, N. A. B., Reshan, M. S. A., Gilanie, G., Alyami, S., Hassan, M. U., & Shaikh, A. (2023). A CNN-Model to Classify Low-Grade and High-Grade Glioma from MRI Images. *IEEE Access*, 11, 46283–46296.

4. Khan, F., Ayoub, S., Gulzar, Y., Majid, M., Reegu, F. A., Mir, M. S., Soomro, A. B., & Elwasila, O. (2023). MRI-Based Effective Ensemble Frameworks for Predicting Human Brain Tumor. *Journal of Imaging*, 9(8), 163.

5. Saeedi, S., Rezayi, S., Keshavarz, H., & Niakan Kalhori, S. R. (2023). MRI-Based Brain Tumor Detection Using Convolutional Deep Learning Methods and Chosen Machine Learning Techniques. *BMC Medical Informatics and Decision Making*, 23(1), 16. https://doi.org/10.1186/s12911-023-02114-6

6. Chatterjee, S., Nizamani, F. A., Nürnberger, A., & Speck, O. (2022). Classification of Brain Tumours in MR Images Using Deep Spatiospatial Models. *Scientific Reports*, 12(1), 1505. https://doi.org/10.1038/s41598-022-05572-6

7. Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D., Hawkins, C., Ng, H. K., Pfister, S. M., Reifenberger, G., Soffietti, R., von Deimling, A., & Ellison, D. W. (2021). The 2021 WHO Classification of Tumors of the Central Nervous System: A Summary. *Neuro-Oncology*, 23(8), 1231–1251. https://doi.org/10.1093/neuonc/noab106

8. Aggarwal, M., Tiwari, A. K., Sarathi, M. P., & Bijalwan, A. (2023). An Early Detection and Segmentation of Brain Tumor Using Deep Neural Network. *BMC Medical Informatics and Decision Making*, 23(1), 78. https://doi.org/10.1186/s12911-023-02174-8

9. Fu, S., Hamilton, M., Brandt, L., Feldmann, A., Zhang, Z., & Freeman, W. T. (2024). FeatUp: A Model-Agnostic Framework for Features at Any Resolution. *MIT & Microsoft & Adobe Research*.

10. Shipitsin, V., Bespalov, I., & Dylov, D. V. (2022). GAFL: Global Adaptive Filtering Layer for Computer Vision. *Skolkovo Institute of Science and Technology*. Received 4 August 2021, Revised 20 April 2022, Accepted 27 July 2022, Available online 1 August 2022, Version of Record 10 August 2022.