Segmentation of Brain MRI tumors by MedSAM with prompts generated by object detection

Kai Zhou*
Faculty of Data Science
City University of Macau
Macau, China
D22091101004@cityu.edu.mo

Dan Yu
Faculty of Data Science
City University of Macau
Macau, China
D22091101257@cityu.edu.mo

***Abstract*—This research paper explores the use of deep learning technology in the segmentation of multimodal brain tumor medical images, particularly emphasizing the application of the customized and fine-tuned MedSAM (Segment Anything in Medical Images) model. MedSAM, an adaptation of the original SAM (Segment Anything Model), is specifically tailored for the nuances of medical imaging. In the model's training and fine-tuning phase, it retains the robust image feature extraction capabilities of the pre-trained model's image and prompt encoders. At the same time, the mask decoder is specifically modified to boost the precision of tumor segmentation. The results of the study indicate that the refined MedSAM model surpasses both the standard SAM model and the initial version of MedSAM in Dice coefficient scores, confirming the method's efficacy in enhancing brain tumor image segmentation accuracy. This research demonstrates that a general model, when appropriately modified, can be effectively utilized for specialized medical image segmentation tasks, not only elevating the accuracy of brain tumor segmentation but also offering innovative approaches and insights for future clinical diagnostic applications. Furthermore, the study sheds light on the model's versatility and its potential applicability to diverse dataset types, providing a foundation for subsequent research avenues.**

Keywords-component; deep learning; detection of brain tumors; medical image segmentation

I. INTRODUCTION

Medical image segmentation is a crucial component of computer-assisted diagnosis, aiming to automatically pinpoint and extract specific organs, tissues, or disease areas from medical images. These images often feature, low signal-to-noise ratios, and multimodal properties, posing significant challenges to the segmentation process. In the lase decade, the rapid advancements in deep learning technology have led to notable progress in medical image segmentation techniques[1]. These models are adept at autonomously learning intricate features within the images, consequently improving segmentation accuracy[2]. Recently, large-scale deep learning models, which have garnered significant success in areas like natural language processing and computer vision, are now being applied to medical image segmentation tasks[15]. With their extensive parameter counts, these models can assimilate a wealth of information from vast datasets. Though successful, these large models are still in formative research stage in medical image segmentation.

Regarding the optimization of large-scale models, researchers are actively seeking methods to boost the effectiveness of these models in the field of medical image segmentation. These efforts involve expanding the parameter count of the models, enhancing their network architectures, and developing innovative training techniques. In terms of clinical applications, there is ongoing research into how these segmentation methods, powered by large models, can be applicated to clinical settings. This includes their application in supporting diagnostic processes and in surgical planning.

The evolution of image segmentation technology has transitioned from traditional to deep learning methods. Early segmentation techniques, such as threshold processing[3][4], region growing[5], and edge detection[6], worked well in certain conditions but often depended on manual tuning and were sensitive to noise and image quality fluctuations. The development of deep learning[7] has positioned brain tumor segmentation methods based on deep neural networks at the forefront. These methods enhance segmentation accuracy by learning tumor characteristics within images. Convolutional Neural Networks (CNNs)[8], widely used in deep learning, effectively capture local features in images. Fully Convolutional Networks (FCNs), a variation of CNNs, can segment images of any size. Nonetheless, CNNs might cause computational redundancy in processing dense imagery[9]. To address this issue, CNN-based algorithms like FCNs[10] and U-Net[11] were proposed. The U-Net architecture and its variants have gained extensive research and application in brain tumor segmentation[12]. For instance, the SCU-Net, an enhanced U-Net model developed by Zheng et al., has achieved notable advancements in segmentation efficiency[13]. The SCU-Net enhances segmentation precision through its decoding module, which combines multi-layer convolution and transposed convolution.

Meta Research has recently achieved notable advancements in their SAM (Segment Anything Model) model[14]. A key attribute of the SAM model is its zero-shot generalization ability, which enables it to segment objects in images effectively without needing further training. This capability broadens the SAM model's applications across various fields, including augmented reality, biomedical image segmentation, and integration with diffusion models[14]. MedSAM (Segment Anything in Medical Images) is an adaptation of the original SAM model, specifically tailored and enhanced for medical image segmentation. This model has been fine-tuned to cater to the unique aspects of medical images, fulfilling the specific demands of medical image segmentation.

2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT) | 979-8-3503-8555-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/AINIT61980.2024.10581523

This article concentrates on utilizing the MedSAM model for segmenting brain tumor medical images, specifically fine-tuning it for the distinct features of multimodal MRI of brain tumors. We aim to showcase the superiority of MedSAM in brain tumor image segmentation by contrasting it with the original SAM and the unmodified MedSAM. Our expectation is that the refined MedSAM model will deliver enhanced accuracy in the identification and segmentation of brain tumors, thus offering a more effective diagnostic aid to physicians.

In the following chapters, we will comprehensively describe the architecture and optimization strategies of the MedSAM model, along with its practical applications and analysis of results in the context of brain tumor medical image segmentation. Our research aims to contribute new insights and approaches to the evolution of the medical image segmentation domain, with a particular focus on the utilization and refinement of large-scale models.

II. Methods

The overall model framework proposed in this study is shown in Figure 1. In this framework, the YOLO model is first used for object detection on the images in the training set to recognize the tumor. The coordinates of the located tumor region are then used as the prompt of the fine-tuned MedSAM model. During the process of fine-tuning the MedSAM model, some weights of the image encoder and cue encoder were frozen to take advantage of the existing image feature extraction capabilities in the pre-trained model. Finally, the tuned mask decoder is responsible for generating an accurate tumor mask output.

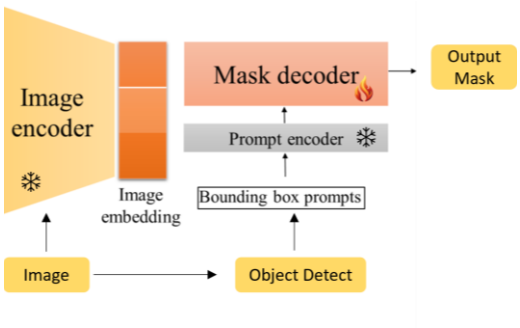

Figure 1. Output display of fine-tuned model.

A. A teardown analysis of YOLOv5s model

YOLOv5 maintains a foundational structure akin to its YOLO predecessors, yet it incorporates several enhancements and optimizations [16]. It employs CSPDarknet53, as introduced by Alexey Bochkovskiy and others in the YOLOv4, serving as its backbone for proficient feature extraction [16]. This architecture, which divides and then intersects feature maps at different network stages, significantly lowers computational complexity while preserving effective feature extraction capabilities. The model's 'Neck' consists of a Feature Pyramid Network (FPN) and a Path Aggregation Network (PAN). The FPN amalgamates features from various levels, facilitating the model's ability to detect objects of diverse sizes. Meanwhile, the PAN further betters the flow of information across different feature layers, ensuring that ample contextual information is leveraged during object detection. In its Detection Head, YOLOv5 features detection heads on three distinct scales, each tasked with predicting the position, size, confidence, and category of bounding boxes. This approach of multi-scale detection makes YOLOv5 adept at identifying tumors of varying sizes.

B. A teardown analysis of SAM model

The SAM model is an end-to-end image semantic segmentation model that can segment any object in an image based on prompts. The SAM model consists of three main parts: an image encoder, a prompt encoder, and a mask decoder. The image encoder uses a Transformer-based ViT (Vision Transformer) model [17], while the prompt encoder uses an RNN model. The mask decoder merges outputs of the encoder and the prompt encoder to generate a segmentation mask. The SAM model, as a foundational model for image segmentation, was trained using the SA-1B dataset and performs well in segmenting everyday life images; however, its performance is not as effective on medical images like multimodal MRI of the brain, where boundaries are more ambiguous.

C. A teardown analysis of MedSAM model

The MedSAM model, based on the SAM model, has been fine-tuned for medical image segmentation using a comprehensive medical image dataset that includes various medical imaging modalities. In the MedSAM model, the image encoder and prompt encoder of the SAM model are frozen, with only the mask decoder being fine-tuned. Medical image segmentation is a diverse task, as different medical images have different characteristics. Using a comprehensive medical image dataset helps the model learn these features, thereby enhancing the model's versatility.

III. Experiments and Results

In this section, we present the experiments conducted and the results obtained during the course of our study. Our research focuses on the segmentation of adult glioma tumors in multi-modal magnetic resonance imaging (mpMRI) data. We begin by describing the data sorting and preprocessing steps, followed by the training of YOLOv5 models for tumor detection. Subsequently, we fine-tune the MedSAM model for tumor segmentation and provide a detailed analysis of the results. Our evaluation metrics include precision, mean average precision (mAP), and Dice coefficient, which assess the performance of the models in detecting and segmenting glioma tumors. The outcomes demonstrate the effectiveness of our approach and the significant improvement achieved through fine-tuning.

A. Data Sorting and Preprocessing

The data set used in this study comes from the public training data of the Segmentation - Adult Glioma task of BraTS 2023[18][19][20], including 802 cases of multi-modal magnetic resonance imaging (mpMRI). These data sets cover four types of mpMRI images, namely: pre-contrast T1 weighted (T1n), post-contrast T1-weighted (T1c), T2-weighted (T2w), and T2 fluid attenuated inversion recovery (T2-FLAIR, T2f for short), accompanied by manual annotation by experts. All BraTS mpMRI scans are provided in NIfTI file format.

In this study, we focus on the analysis of three types of magnetic resonance (MR) data: T1c, T2w, and T2f. The

advantage of selecting these three modalities is that they provide richer tumor-related contrast information compared to T1n. For example, T1c modality helps to highlight areas with dense vasculature or compromised blood-brain barriers; T2w makes fluid and edematous areas appear brighter in the images, which is beneficial for detecting edematous brain lesions; T2f, through special pulse sequences, suppresses fluid signals, making it suitable for detecting small lesions. we process these modalities by slicing, normalizing the pixel values to a range of 0 to 255, and saving them as 24-bit depth PNG images with three channels. For slices with labeled data, we calculate the bounding boxes of tumor locations and save them in text files in YOLO format, while unlabeled slices are also saved in the same manner as background data.

B. Training YOLOv5 Model and Training Results

The preprocessed dataset was divided into training, testing and validation sets in the ratio of 6:2:2. Using the official YOLOv5 code repository and pretrained weights, two different sizes of YOLOv5s and YOLOv5l models were trained under the conditions of 100 epochs, an initial learning rate of 0.01, and the SGD optimizer. As shown in Table 1, the larger model outperforms the smaller model in all metrics when considering performance under different thresholds of precision and mean accuracy (mAP). Specifically, the large number of parameters advantage of the larger model provides better results in single-task recognition.

Table 1 Comparison of YOLOv5 Model Detection Results.

Model Name:	Model Performance Metrics		
	Precision	*mAP_0.5*	*mAP_0.5:0.95*
YOLOv5s	0.91669	0.86116	0.63171
YOLOv5l	0.93441	0.90303	0.71470

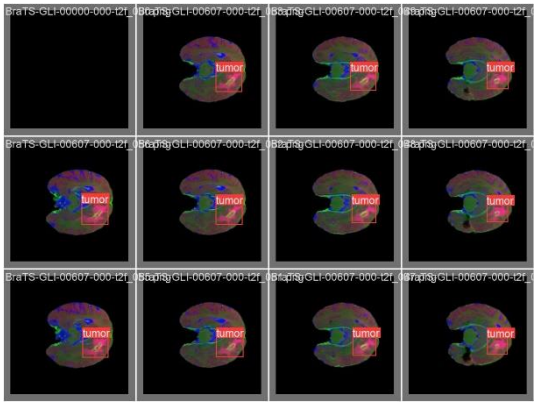

Figure 2. Example of verification set labeling.

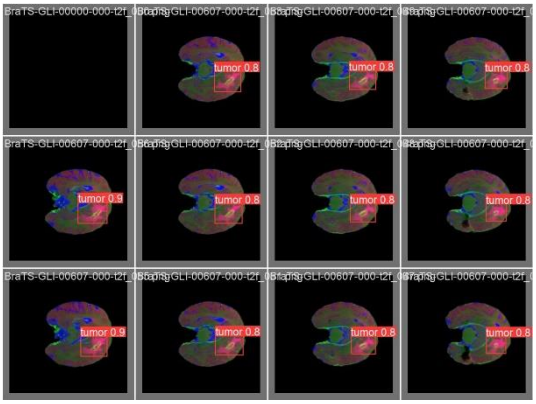

Figure 3. Prediction results of YOLOv5l model on the validation set.

As shown in Figure 2, the annotation results on multiple slice images are displayed. Figure 3 shows the prediction results of the YOLOv5l model on the validation set. The area within the red bounding box identifies the tumor location predicted by the model and gives the corresponding confidence score.

Subsequently, the trained YOLOv5l model is used to predict the test set data, identify the tumor location, size, center point and confidence score, and save it as text data for use by the SAM model and MedSAM model to output the tumor mask.

C. Fine-tuning the MedSAM Model and Comparing Results

After completing the YOLO model training, we further fine-tuned the pre-trained weights of the MedSAM model. The fine-tuning process follows a similar approach to the original training of MedSAM, but is specifically optimized for our dataset. During 100 training epochs, we set an initial learning rate of 0.0001 and selected AdamW as the optimizer. This is because the AdamW optimizer combines momentum and weight decay and has better performance for sparse gradients or noisy data.

When fine-tuning the model, the choice of loss function is crucial to the training process and final performance. In order to achieve more accurate tumor segmentation, we adopt a strategy of combining loss functions, combining Dice loss and cross-entropy loss. Specifically, we use the Dice loss provided by the MONAI framework, which applies a Sigmoid activation function as well as squared predictions to reduce the imbalance between predictions and targets, and adopts mean as a reduction method, aiming to optimize the similarity between model output and real labels. At the same time, we also introduced binary cross-entropy loss (BCEWithLogitsLoss), which has built-in Logit activation and is suitable for binary classification tasks. It helps to deal with pixel-level classification problems and further balances the accuracy of segmentation boundaries.

The mathematical expression of Dice loss is:

$$L_{Dice} = 1 - \frac{2 \cdot \sum_i^N p_i \cdot g_i}{\sum_i^N {p_i}^2 + \sum_i^N {g_i}^2} \quad (1)$$

p_i is the probability predicted by the model, processed by the Sigmoid activation function; and g_i is the real label.

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\big(g_i\, log\big(\sigma(p_i)\big) + (1-g_i) \cdot log\big(1-\sigma(p_i)\big)\big) \quad (2)$$

$\sigma(\,)$ represents the Sigmoid function, p_i is the original output of the model, and g_i is the corresponding real label.

The final total loss L is the sum of these two losses:

$$L = L_{Dice} + L_{BCE} \quad (3)$$

To adapt to the task of identifying and segmenting adult glioma, we made customized adjustments to the structure of MedSAM. Specifically, we freeze the weights of the Image Encoder and Prompt Encoder to maintain their ability to encode multi-modal MRI data. Instead, we fine-tuned the Mask Decoder to accommodate the need for precise mask output for specific types of brain tumors. This approach allows the model to leverage pre-trained feature extraction capabilities while focusing on improving tumor localization and segmentation accuracy.

The fine-tuned MedSAM model is capable of generating more detailed tumor mask images, demonstrating its application potential on complex brain MRI data.

Table 2 Comparison of Model Segmentation Results.

Model name:	SAM	MedSAM	Fine-tuned MedSAM
Dice score:	0.692	0.723	0.889

In our experiments, we evaluate the performance of SAM and MedSAM models on the tumor segmentation task. As shown in Table 2, using the Dice coefficient as the evaluation criterion, the SAM model achieved a score of 0.692, while the MedSAM model without fine-tuning achieved a score of 0.723. Remarkably, after fine-tuning for this task, the performance of the MedSAM model improved to a Dice score of 0.889, indicating that the fine-tuning process significantly improved the model's ability to accurately identify and segment tumor regions. This result highlights the importance of fine-tuning to improve the performance of deep learning models in specific medical image analysis tasks. The fine-tuned MedSAM model can better understand and adapt to the morphological characteristics of tumors, providing more accurate predictions when generating tumor masks.

IV. Conclusions

This study demonstrates the effectiveness of utilizing deep learning models, specifically the custom-modified and fine-tuned MedSAM model, in the task of MRI image segmentation of brain tumors. By introducing the YOLOv5 model for preliminary identification of lesion areas, and combining it with the advanced segmentation capabilities of the MedSAM model, we are able to achieve high-precision brain tumor segmentation in multi-modal MRI images. Experimental results show that compared with the SAM model, the performance of the fine-tuned MedSAM model on the Dice coefficient is significantly improved, thus verifying the effectiveness of our method in improving segmentation accuracy.

It's noteworthy that the SAM model, as a foundational model, has shown certain versatility in medical image segmentation tasks. Although the SAM model was not initially designed specifically for medical images, its zero-shot generalization capability offers new possibilities for handling complex and ambiguous boundaries in medical imaging. Through the fine-tuning methods in this study, we have further extended the applicability of the SAM model in medical image segmentation, proving that even a general model can be appropriately adjusted to meet specific medical needs.

However, it should also be noted that although our model performs well on the BraTS 2023 dataset, the model's generalization ability and performance on other types of datasets still need to be further verified. Future research can explore the application of our method on more diverse medical imaging datasets, as well as further improve the generality and robustness of the model.

In summary, our study not only improves the accuracy of brain tumor image segmentation, but also provides useful insights into using general deep learning models to solve specialized medical tasks. It is expected that these findings can promote future applications in clinical diagnosis.

References

[1] Chen, X., Wang, X., Zhang, K., Fung, K. M., Thai, T. C., Moore, K., Mannel, R. S., Liu, H., Zheng, B., & Qiu, Y. (2022). Recent advances and clinical applications of deep learning in medical image analysis. Medical image analysis, 79, 102444. https://doi.org/10.1016/j.media.2022.102444

[2] Roth, H.R., Shen, C., Oda, H., Oda, M., Hayashi, Y., Misawa, K., & Mori, K. (2018). Deep learning and its application to medical image segmentation. ArXiv, abs/1803.08691.

[3] Zhang, Y., Wang, Y., & Zhang, J. (2022). A threshold-based method for pulmonary nodule detection. Computerized Medical Imaging and Graphics, 89, 102021.

[4] Zhang, Y., Wang, Y., & Zhang, J. (2023). A level set-based method for pulmonary nodule detection. IEEE Transactions on Medical Imaging, 32, 579-591.

[5] Zhang, X., Li, X., & Feng, Y. (2015). A medical image segmentation algorithm based on bi-directional region growing. Optik, 126(20), 2398-2404.

[6] Liu, H., Feng, Y., Xu, H. et al. MEA-Net: multilayer edge attention network for medical image segmentation. Sci Rep 12, 7868 (2022).

[7] Rana, M., & Bhushan, M. (2023). Machine learning and deep learning approach for medical image analysis: diagnosis to detection. Multimedia Tools and Applications, 82(17), 26731-26769.

[8] Jyothi, P., & Singh, A. R. (2023). Deep learning models and traditional automated techniques for brain tumor segmentation in MRI: a review. Artificial intelligence review, 56(4), 2923-2969.

[9] Yang T, Song J, Li L, Tang Q. Improving brain tumor segmentation on MRI based on the deep U-net and residual units. J X-Ray Sci Technol. 2020;28(1):95–110.

[10] Long J, Shelhamer E, Darrell T. IEEE: fully convolutional networks for semantic segmentation. In: Proceedings of IEEE conference on computer vision and pattern recognition. Boston, MA: IEEE; 2015: 3431–3440.

[11] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: 18th Proceedings of international conference on medical image computing and computer-assisted intervention, vol. 9351.

[12] Munich, Germany. Springer International Publishing Ag; 2015: 234–241.

[13] Zheng, P., Zhu, X. & Guo, W. Brain tumour segmentation based on an improved U-Net. BMC Med Imaging 22, 199 (2022). https://doi.org/10.1186/s12880-022-00931-1

[14] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). Segment anything. arXiv preprint arXiv:2304.02643.

[15] Ma, J., He, Y., Li, F. et al. Segment anything in medical images. Nat Commun 15, 654 (2024). https://doi.org/10.1038/s41467-024-44824-z

[16] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

[17] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Uszkoreit, J. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv preprint arXiv:2010.11929.

[18] U.Baid, et al., The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification, arXiv:2107.02314, 2021.

[19] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694

[20] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, et al., "Advancing the Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117