Fine-tuned SAM for Glioma Segmentation on Multimodal Brain MR Images

Xiaoyu Shi, Shurong Chai, Rahul Kumar Jain, Yinhao Li, Jingliang Cheng, Jie Bai, Guohua Zhao, Lanfen Lin and Yen-Wei Chen*

Brain tumors are classified as either primary or secondary brain tumors. Glioma is the most prevalent primary brain tumor. Glioblastoma (GBM) is the most aggressive glioma. Note that <5% of patients survive five years after diagnosis. According to the 2016 World Health Organization (WHO) Classification scheme for gliomas, glioma segmentation is a very important basis for genotype prediction and diagnosis. Therefore, it is necessary to predict glioma region for treating glioma. Magnetic resonance imaging (MRI) is a common method for helping doctors to do glioma diagnosis. Usually, MRI produces images in four modalities, which include T1, T2, T1CE, and Flair. Each modality has unique characteristics, which is useful in glioma region segmentation. Although it is difficult to segment the glioma region using MRI images by an inexperienced person, it can be predicted with deep learning, which could help doctors save time in medical diagnosis. There have been many improvements achieved in the last few years in medical image segmentation using machine learning, particularly deep learning. Olaf et al. proposed a deep learning method using convolutional neural network (CNNs) with skip connection part to make a medical image segmentation. Furthermore, with the development of Vision Transformer (ViT) in deep learning, Chen et al. proposed a Tran-Unet. With the development of devices, more and more foundation models such as GPT and Segment Anything (SAM) are starting to emerge, which are trained using very large data sets and therefore have good results for different tasks. Because of the limited of difference between medical image data sets and real-world data sets, the deep learning-based method for medical data cannot just fine-tune the whole pre-trained foundation model. Therefore, there are already some methods for fine-tuning the foundation model for medical images. These methods fine-tune the SAM by adding different adapters and achieve good results. However, due to the specificity of multimodal brain MRI, we cannot directly use these methods to process multimodal data. In this study, we proposed a multi-modal fusion adapter for fine-tuning the foundation model (SAM) to improve the performance of glioma region segmentation task. We classify the modalities into T1-based modal and T2-based modal based on the modal characteristics, and then use two pre-trained foundation models to process these two modalities. We added the multi-modal fusion adapter module to interact with the two modalities to fine-tune the model. During the training process, the backbone network is frozen and the parameters of the adapter as well as the mask decoder are trained. We proposed two kinds of fusion adapters and did ablation studies for them. When comparing our model with the current state-of-the-art methods, our proposed method has better performance.