How SAM Perceives Different mp-MRI Brain Tumor Domains?

Cecilia Diana-Albelda
Universidad Autónoma de Madrid
cecilia.diana@uam.es

Roberto Alcover-Couso
roberto.alcover@uam.es

Álvaro García-Martín
alvaro.garcia@uam.es

Jesus Bescos
j.bescos@uam.es

Abstract

Gliomas, among the deadliest forms of cancer, are brain tumors that present a significant challenge due to their rapid progression and resistance to treatment. Effective and early diagnosis is critical for improving patient prognosis. Deep learning, particularly through large-scale vision models like Segment Anything Model (SAM), offers a new pathway for tumor segmentation. This study seeks to address the primary challenge of adapting SAM for mp-MRI brain scans, which typically encompass multiple imaging modalities not fully utilized by standard three-channel vision models. We demonstrate that leveraging all available MRI modalities achieves superior performance compared to the standard mechanism of repeating a MRI scan to fit the input embedding. Our research also focuses on parameter-efficient tuning of SAM to effectively train the model while minimizing resource usage, showcasing significant improvements when evaluated across multiple datasets. Finally, we expose how SAM perceives differences across varied brain tumor domains by visually analyzing the features extracted on each of them. Our code and models are available at github.com/vpulab/med-sam-brain.

1. Introduction

Brain tumors, specially gliomas, stand as some of the deadliest cancers, in contrast with other types in terms of survival rates [1–4]. While advancements in medical science have significantly improved the prognosis for many cancers, gliomas remain particularly lethal, with approximately 80% of individuals diagnosed succumbing within two years [5–7]. This grim statistic is set against a backdrop where over 90% of individuals diagnosed with breast or prostate cancer can expect to survive beyond five years, regardless of the cancer's malignancy [8–10]. The diagnosis and treatment of brain tumors is a multifaceted and challenging task, primarily due to the complex nature of brain anatomy. Gliomas are specially notorious for their resistance to conventional therapies, making fast and early diagnosis crucial for the patient survival [4, 11–14].

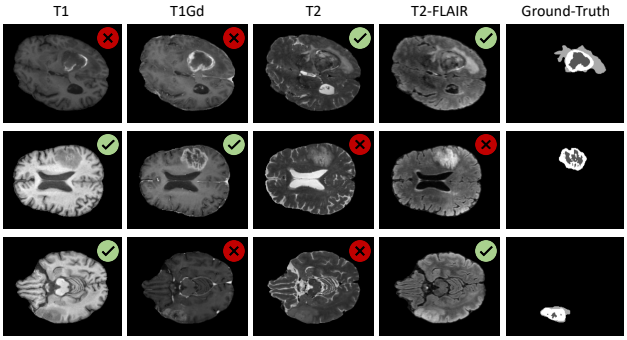

Figure 1. **Adult glioma images from the RSNA-ASNR-MICCAI BraTS challenge [30].** Each row illustrates an instance where the complete tumor structure is discernible in certain imaging modalities but not in others. This emphasizes the importance of evaluating each case using all four MRI modalities with automatic glioma segmentation models.

In this context, automated segmentation of tumors from medical imaging data emerges as a pivotal step by speeding the diagnosis process [11, 12, 15–24]. Nevertheless, the pace of progress in this area is inextricably linked to the advances in computer vision [11, 22–25]. The recent spotlight on Large-scale Vision Models (LVM) has captured the attention of researchers due to their exceptional versatility and performance across diverse domains [26–28]. Notably, the Segment Anything Model (SAM) [29] emerges as the foundational model designed for image segmentation tasks.

However, there are three main challenges in applying these methods directly to the specialized medical domain:

1. **Conventional models build on three-channel (RGB) color systems**, hence, its application involves the conversion of medical images (grayscale) into a compatible format: This workaround falls short when applied to multi-parametric Magnetic Resonance Imaging (mp-MRI) brain

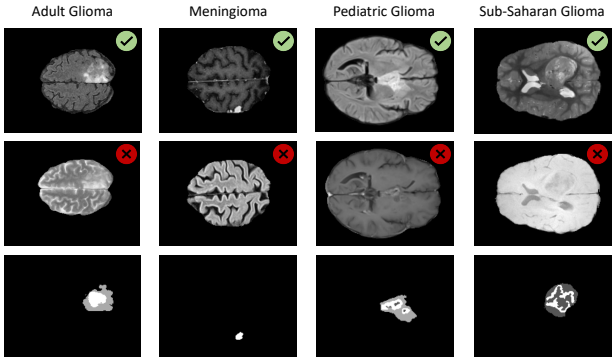

Figure 2. **Different brain tumor images.** Visual examples of patients where one MRI modality contains information to effectively segment the complete tumor contour and another does not.

scans, which encompass four distinct image modalities: T1-weighted (T1), post-gadolinium (Gd) contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) [30–33]. This gap underscores the need for a tailored approach that respects mp-MRI brain data, to fully harness its diagnostic potential. Figure 1 and 2 exemplify the need of combining all mp-MRI modalities. To address this challenge, this study modifies SAM architecture to entirely utilize each mp-MRI data entry.

2. **Pre-trained LVMs are usually not equipped with medical-specific knowledge**, therefore, training or fine-tuning with medical images is needed, a process computationally demanding which often results in catastrophic forgetting [34–36]. To deal with this problem, we propose to employ Parameter Eficcient Fine-Tuning (PEFT) of SAM through Low Rank Adaptation (LoRA) [37], where the weight update is restricted to a small rank hyper-plane.

3. **Once these methods are adapted to glioma images, how general are the features they extract?**. There is a risk that these features might become overly specialized, limiting the model's ability to interpret broader aspects within other brain tumor domains. To tackle this issue, we propose an evaluation setup composed of four domains [30–33, 38–40]. This setup allows us to observe the features extracted by the tuned model providing insight into the model perception across domains. Our finding suggests that LVMs present effective segmentation regardless of the dataset, but they still perceive significant discrepancies across domains.

Our study yield models with enhancements of up to 5% in dice scores by leveraging all four MRI modalities. Moreover, our introduced PEFT training leads to models adept at segmenting across various domains. Nonetheless, despite their broad utility, we observe discrepancies at the feature levels of the model when applied inter-domain.

2. Related work

Medical Image Segmentation (MIS) involves identifying and delineating anatomical structures of interest in medical images [11, 12, 15]. Given that medical professionals often spend considerable time manually conducting this task [16–20], numerous automated models have emerged to alleviate the workload of doctors [21–24].

2.1. SAM Zero-Shot on MIS

LVMs have attracted considerable interest from researchers due to their remarkable ability to generalize and effectively adapt across a wide range of domains [26–28]. Among these models, SAM [29] has been presented as a foundational architecture designed for image segmentation tasks. SAM is built upon the Vision Transformer (ViT) [41] and it has been trained on a large scale dataset composed of over 11 million images and 1 billion masks. The dense and varied training data allows SAM to present state-of-the-art zero-shot segmentation across multiples domains [29].

While SAM excels in segmenting natural images, recent studies have evidenced its sub-optimal performance in the segmentation of medical images [42–45], casting doubt on its zero-shot capability within this specialized context. This deficiency is due to the lack of domain-specific knowledge pertinent to medical imaging [42, 46–49], which presents unique challenges such as low image contrast, indistinct tissue boundaries, and tiny lesion regions [15–17, 21, 22, 49]. To analyze these challenges, we propose to gather multiple sources of MRI data to tune SAM with one specific domain and analyze if the gap is found only to the medical domain, or if the gap is also present across sources of tumor data, such as patient age or ethnicity.

2.2. Training Foundational Models on Medical Domains

To tailor foundational models for the medical domain, one approach involves fine-tuning the entire model with a medical database [46]. However, this process is highly computational and memory demanding [43, 44, 47, 50], and often results in catastrophic forgetting [34–36], where the model forgets previously learned patterns [51].

To overcome such drawback, recent techniques aim at regularizing the modification of the weights to retain the open-world recognition capabilities of the model. Specifically, PEFT emerges as a compelling solution [52]. PEFT typically modulates the weight update as a combination of low rank matrices reducing the computational costs and the impact of the weight update [37, 53, 54]. This approach accelerates adaptation to the new domain, allowing efficient tuning of the model given the immense parameter count in foundational models. Moreover, PEFT has not only shown efficiency but has also surpassed performance scores

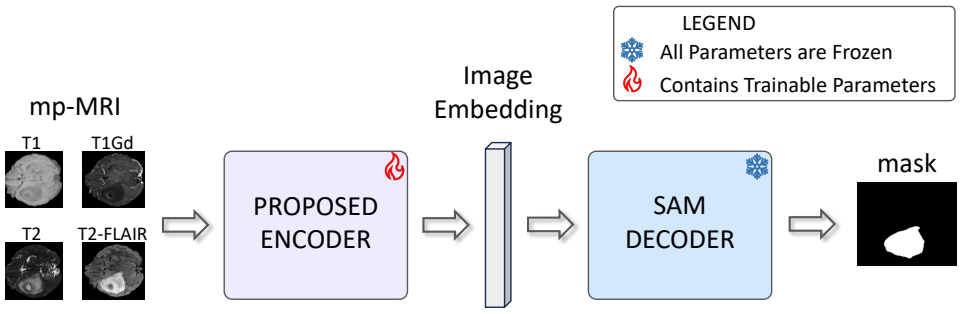

Figure 3. **Pipeline overview**. We propose to adapt the encoder by: 1) accounting for all the mp-MRI volumetric image modalities; and 2) specifically tuning of the encoder to retain the open-world segmentation capabilities of SAM.

achieved by fully fine-tuning the entire model [49], while mitigating catastrophic forgetting [52, 53].

2.3. SAM on Volumetric Medical Data

Medical imaging modalities such as MRI [40, 55, 56] or Computed Tomography (CT) [25, 57–59] possess an additional dimension compared to RGB images. The dimensions of an RGB image are (c, H, W), with c denoting the number of channels, and H and W the height and width of the image, respectively. However, the dimensions of an MRI are (c, H, W, Z), where typically $c = 1$ and Z is the depth of the volumetric image, i.e., the number of slices.

To train SAM with these images, most works do not account for the correlation of slices [60, 61]. On the other hand, some studies capture 3D spatial information by adapting 2D to 3D, often by freezing 2D layers while training 3D adapters, enabling the model to learn from 3D images [47, 49, 50]. Alternatively, [48] introduces alterations to the entire network architecture (Image Encoder, Prompt Encoder, and Mask Decoder) to transition it to 3D and train it completely with these modalities of medical data. Certain medical images such as mp-MRI brain scans encompass several volumetric image modalities associated with a single patient and diagnosis [38, 39, 62–64]: T1, T1Gd, T2, and T2-FLAIR. In this study, we focus on adapting SAM to segment mp-MRI medical images in various domains: Adult Glioma [38], Meningioma [31], Pediatric Glioma [32], and Africa Sub-Sahara Glioma [33]. Thus, the objective is to understand how SAM perceives the differences across these diverse domains.

3. Adaptation Method

We propose to adapt SAM model for mp-MRI medical images following the pipeline illustrated in Figure 3 and focusing on three aspects. First, we tune the patch embedding layer of the encoder so that the model accounts for each volumetric image modality (T1, T1Gd, T2 and T2-FLAIR). Second, we propose an efficient tuning of the spatial layers of the encoder. Third, we leverage SAM decoder and propose a novel mp-MRI setting for evaluating how the model generalizes to different domains in the context of brain tumor segmentation.

3.1. Segmentation framework

SAM Architecture. SAM [29] is divided into three primary components: an image encoder, a prompt encoder, and a mask decoder. The image encoder is in charge of extracting image representations, for which it combines a patch embedding layer (16x16 patches) and transformer blocks based on a Masked Auto-Encoder (MAE) [65] pretrained ViT [41]. Prompt encoder can manage both sparse and dense prompts, sparse being represented by 2D position encoddings, and dense encoded via a convolution algorithm. On the other hand, the mask decoder employs bidirectional cross-attention to capture the interactions between prompts and images. SAM then samples the image encoding, and an MLP maps the output token into a dynamic linear classifier that predicts the target mask.

The adaptation of SAM for processing mp-MRI is illustrated in Figure 4. It comprises two key aspects: 1) Modifying the encoder architecture to correctly handle the correlation between various MRI modalities at the patient level. 2) Implementing PEFT on the modified model to efficiently adapt it to the domain of brain images.

4-Channel Encoder. Since SAM is designed for RGB images, the encoder is configured with an input composed of 3 channels. This input is encoded by the patch embedding which consists only of a single layer. Therefore, we propose to modify the initial patch embedding as means to analyze the four mp-MRI modalities: T1, 1Gd, T2, and T2-FLAIR. We follow this protocol as we have noticed a single image modality does not always present complete tumor information (Figures 1 and 2), therefore, making the model

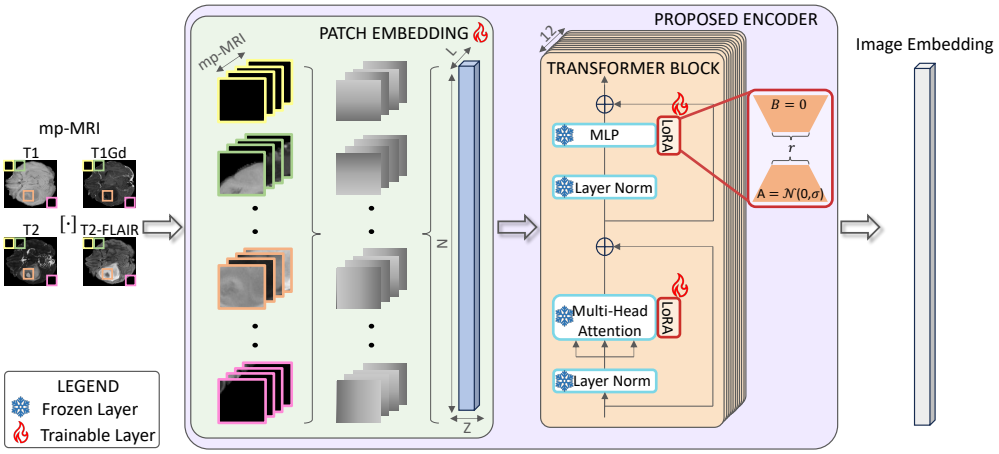

Figure 4. **Proposed Encoder.** We propose to modify the patch embedding layer, so that it accounts for the all the MRI modalities, allowing for a seamless integration of the information. Then, we employ LoRAs to tune Multi Layer Perceptron blocks (MLP) and Attention (Q,K,V embedding) layers of the transformer blocks.

unable to segment it regardless of the efficacy.

Cross-slice attention. Another significant issue we find in medical images is the relationship across the different slices of the mp-MRI image. To account for such relationships, we modify the original SAM architecture following [49]. To that end, we feed contiguous slices to the model and provide a cross-slice mechanism into each transformer block. This cross-slice mechanism aims at modulating the relationships between slices. Specifically, we set a residual cross-slice attention module (depth branch) $\mathcal{Z}$ into each transformer block $\mathcal{T}$:

$$x_z(i,:) = Attn(x(\cdot,i,:)) + x(\cdot,i,:), \tag{1}$$

where $Attn$ is the multi-head attention, $x_z \in \mathcal{R}^{N,L}$, $x \in \mathcal{R}^{Z,N,L}$, Z is the number of slices per mp-MRI image analyzed, N is the number of features per image and L the dimension of the features. Empirically, this is efficiently computed as an attention block on the transposed inputs. Then these depth features x_z are combined in the usual transformer block (spatial branch) by:

$$x_s(z,i,:) = Attn(x(z,i,:)) + x(z,i,:) + x_z(i,:). \tag{2}$$

This combinations are computed as: first, for the depth branch in Equation 1, transposing the input feature x. Second, for the combination with spatial features in Equation 2, x_z transposed back to its original shape. This procedure is illustrated in Figure 5.

Efficient tuning of SAM to medical images. We employ a PEFT strategy for tuning SAM to medical images. Specifically, we employ LoRAs of the SAM weights (see Section 2):

$$\Phi_l = \phi_l + \Delta\phi_l, \tag{3}$$

where $\phi \in R^{N\times L}$ are the original SAM weights for a given layer l , Φ the weights after tuning and $\Delta\phi$ the weight update obtained through fine-tuning (Without loss of generality, we explain the process for a single layer). As $\Delta\Phi_l$ should be of the same dimension than Φ, such update typically leads to the model downgrade if the tuning is not performed with large scale data. To overcome this, LoRA update presents a framework where $\Delta\Phi$ is constrained by low rank matrices. Therefore, the weight update is restricted to a hyperplane of rank r, where r is a hyper-parameter to define the low rank matrices. Empirically, this is defined as:

$$\Delta\phi_l = A \times B, \tag{4}$$

where $B \in R^{r\times L}$ and $A \in R^{N\times r}$ are the matrix decomposition. This approach allows us to train with exponentially less resources, while also acting as a regularization measure to reduce forgetting of the previously learnt patterns. We initialize $A = \mathcal{N}(0,\sigma)$ and $B = \mathbf{0}$ following [37]. The workflow with LoRAs follows:

$$\Phi_l(x) = \phi_l(x) + BAx \tag{5}$$

3.2. mp-MRI evaluation setting

In this study, we assess the performance of SAM across various brain tumor domains: Adult Glioma [30, 38, 39], Meningioma [31], Pediatric Glioma [32], and Sub-Saharan Glioma [33, 40]. Through this analysis, we aimed to gauge its generalization ability after training on one type of brain tumor and testing on others. Detailed images for each specified domain are presented in Figure 6.

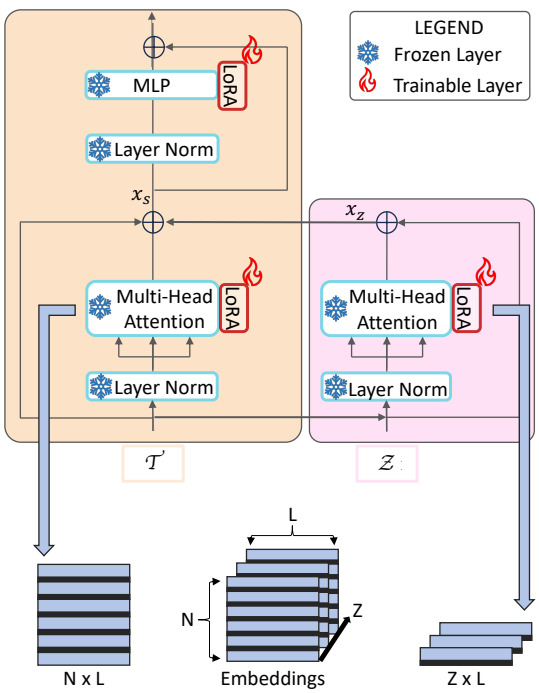

Figure 5. **Cross-slice attention mechanism**. We employ cross-slice attention to account for both space and depth relationships in the mp-MRI images through two linked branches in each transformer block, one of them receiving transposed inputs.

BioPhysic Domain Differences. Gliomas arise from neoplastic glial cells [62–64]. On the other hand, meningiomas [66–68] are typically more circumscribed than gliomas, but they pose additional segmentation challenges due to their extra-axial location, multiplicity, and tendency for skull-base involvement. Unlike gliomas, meningiomas arise from the arachnoid layer of the meninges, situated between the dura mater and the pia mater [31, 68].

While adult gliomas are predominantly found in the frontal or temporal lobes, pediatric gliomas, Diffuse Midline Gliomas (DMG) are commonly located in the pons [32, 69, 70]. Moreover, characteristic imaging features such as enhancing tumor regions and necrotic tissues are less distinct in DMGs compared to adult gliomas. Therefore, dedicated imaging tools are necessary for accurate characterization and diagnosis of such pediatric brain tumors [32, 71].

Despite the significant progress in glioma diagnosis facilitated by deep learning methods, the applicability of models trained on american/european glioma data to clinical practice in Sub-Saharan Africa (SSA) remains uncertain, particularly given the extensive use of lower quality MRI technology in the region [40, 72]. Consequently, further Machine Learning (ML) methods may be required for accurate tumor segmentation in SSA [33].

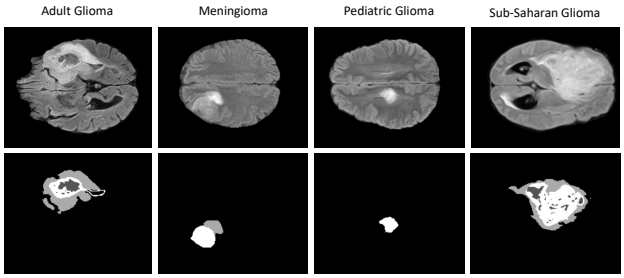

Figure 6. **Four selected brain tumor domains.** The top row represents samples from the T2-FLAIR MRI modality and the bottom one is the ground-truth. In the ground-truth images, white pixels are part of the enhancing tumor, light grey pixels represent an edema, and dark grey pixels are necrotic/non enhancing cells. However, we will only focus on the ground-truth complete extension of the tumor.

3.3. Comparison of different models

Comparing LoRAs Matrixes to assest distribution shift. Our training protocol provides us a comparable weight update from an initial set of parameters ϕ to the tuned parameters on each evaluated dataset Φ_D. We propose to use the singular value decomposition (svd) of each LoRA:

$$\Delta\phi_l = U\Sigma V, \tag{6}$$

where $U \in R^{Z\times Z}$ and $V \in R^{M\times M}$ are real orthogonal matrices and $\Sigma \in R^{Z\times M}$ is a matrix non-negative real numbers on the diagonal denoted as σ_i are the singular values. To measure the distance from one model tuning to another, we follow [73] corollary:

$$|\sigma_i(E) - \sigma_i(F)| \leq ||E - F||\,, \forall E, F. \tag{7}$$

Providing us with r estimations of the discrepancy:

$$||\Delta\phi_{D1} - \Delta\phi_{D2}|| \geq |\sigma_i(\Delta\phi_{D1}) - \sigma_i(\Delta\phi_{D2})|. \tag{8}$$

We follow this approximation as we account for the translation of the weights rather than the element-wise distance.

4. Experiments and Results

In this section, we present the selected datasets, experimental setup and results of our study on adapting SAM for processing mp-MRI images. First, we describe our training strategy, which also includes details on hardware specifications and dataset splitting. Subsequently, we demonstrate the effectiveness of utilizing all four mp-MRI modalities with SAM compared to using fewer modalities under identical conditions. Finally, we provide comprehensive results from our experiments, including comparisons with existing models and variations in training approaches.

4.1. Dataset

The dataset used in this study has been extracted from the public RSNA-ASNR-MICCAI BraTS 2023 challenge [30, 38, 39], which aims to promote research on automatic methods for the diagnosis of gliomas. Our dataset groups four sets of data corresponding to different brain tumor domains, in order to observe how SAM perceives the existing differences among them under identical training conditions and evaluate its performance changes. Therefore, our dataset $D = \{D_1, D_2, D_3, D_4\}$, where:

- D_1 (Adult Glioma) $= \{n_0, n_1, ..., n_{1251}\}$ [38]
- D_2 (Meningioma) $= \{n_0, n_1, ..., n_{1000}\}$ [31]
- D_3 (Pediatric Glioma) $= \{n_0, n_1, ..., n_{99}\}$ [32]
- D_4 (Sub-Sahara Glioma) $= \{n_0, n_1, ..., n_{60}\}$ [33]

Considering the number of samples for each dataset, D_1 and D_2 will serve as both training and evaluation sets, with a train-test split applied within each dataset, being TD_1 and TD_2 the training splits for D_1 and D_2 respectively, and V_1, V_2 its corresponding validation sets. Meanwhile, D_3 and D_4 will solely be used as evaluation sets to assess generalization of the trained models to these domains.

4.2. Training Setup

Each training session was executed on a single 48GB NVIDIA A40 GPU, utilizing a batch size of 1. We employed an 80-20 split of the dataset to validate the learning outcomes for each case.

Our objective does not involve surpassing the state-of-the-art performance in brain tumor segmentation. Rather, we aim to showcase the enhancement achieved by utilizing all four mp-MRI images with SAM in comparison to using only one or three of them, under identical conditions. Consequently, all trainings were conducted using the same optimizer (Adam), $r = 4$ for every LoRA block, and over the same number of epochs, 25. As performance metric, we employ the Dice score [74], as it is the most standard utilized metric in medical image segmentation [15, 21, 75].

Point Prompt. We train the network utilizing a single point as a prompt generated following the iterative procedure outlined in [76]. Here, each subsequent click is determined based on the previously selected set of clicks, with the initial one being randomly generated [77].

4.3. Analysis of the image employed

In the ablation study, we aim to observe the benefits of employing all four mp-MRI image modalities compared to using just one (triplicated to match SAM's original input channels) or a concatenation of three modalities. To achieve this, we conducted various experiments with different setups explained as follows, all trained on TD_1 for 25 epochs.

	MRI modality	$VD_1 \uparrow$	$VD_2 \uparrow$	$D_3 \uparrow$	$D_4 \uparrow$
Zero-Shot	T1	4.33	1.55	3.43	7.59
LoRA		**58.45**	**74.16**	**65.68**	**52.34**
Zero-Shot	T1Gd	6.51	2.39	4.24	6.51
LoRA		**58.54**	**74.71**	**65.59**	**52.61**
Zero-Shot	T2	4.51	2.87	1.76	7.75
LoRA		**58.19**	**74.51**	**65.36**	**52.46**
Zero-Shot	T2-FLAIR	3.85	2.19	3.08	7.85
LoRA		**59.64**	**74.72**	**65.46**	**52.73**
Zero-Shot	T1, T2, T2-FLAIR	5.14	2.91	2.99	6.03
LoRA		**57.87**	**74.44**	**65.37**	**52.67**

Table 1. **Dice scores of SAM Zero-Shot vs LoRA Adaptation.** These results showcase the low performance of foundational models applied directly to medical images, motivating the research of techniques to effectively adapt to the medical domain. (KEY: VD_1 = Validation split for Adult Glioma, VD_2 = Validation split for Meningioma, D_3 = Pediatric Glioma, D_4 = Sub-Saharan Glioma.)

As SAM focuses solely on segmenting the tumor as a whole, when concatenating three modalities we excluded T1Gd. Clinicians typically use T1Gd to delineate only the tumor core and active regions, rather than capturing its entire extent, according to the BraTS annotation protocol [30]. Consequently, our evaluation included concatenating T1, T2, and T2-FLAIR images to represent three different stacked modalities when testing SAM's performance.

Performance impact of scan modality in SAM. Initially, we assessed the zero-shot transferability of SAM with its pre-trained weights across D_1, D_2, D_3, and D_4, and compare it with the model's performance after incorporating LoRA blocks through PEFT. This preliminary evaluation is performed without any modifications to the patch embedding, hence, each run involved the utilization of either a single MRI modality or a concatenation of three modalities. As depicted in Table 1, pre-trained SAM model performs poorly on our dataset, with all dice scores below 10.0. Subset D_4 shows slightly better performance compared to others for each MRI modality. However, applying PEFT with LoRA leads to a significant improvement in results. Within each domain, the dice scores converge to approximately the same value regardless of the chosen MRI modality using LoRA.

How do we combine MRI modalities? In Table 2 we present the dice scores obtained by training SAM model from its pre-trained weights and applying PEFT method described in Section 3 without modifying the patch embedding. This experiment was conducted for each different mp-MRI modality, with each selected image repeated three times to input the net encoder a three-channel image. Additionally, we included in the comparison table the same test performed by concatenating T1, T2, and T2-FLAIR, along

MRI modality	$VD_1 \uparrow$	$VD_2 \uparrow$	$D_3 \uparrow$	$D_4 \uparrow$
$T1 \times 3$	58.45	74.16	65.62	52.34
$T1Gd \times 3$	58.54	74.71	65.59	52.61
$T2 \times 3$	58.19	74.51	65.36	52.46
$T2$-FLAIR$\times 3$	59.64	74.72	65.46	52.73
$[T1, T2, T2\text{-FLAIR}]$	57.87	74.44	65.37	52.67
Ours(TD_1)	**61.85**	65.68	50.65	**56.20**
Ours(TD_2)	58.38	**75.27**	**65.62**	52.35

Table 2. **Performance comparison of models tuned with replicating each MRI modality and our proposed combination of the four of them.** Our method achieves performance improvements across all datasets, highlighting the benefits of the proposed combination. (KEY: TD_1,VD_1 = Training, Validation splits for Adult Glioma; TD_2,VD_2 = Training, Validation splits for Meningioma; D_3 = Pediatric Glioma; D_4 = Sub-Saharan Glioma.)

with the evaluation of our proposed method using all 4 MRI modalities. As observed, the proposed approach leveraging complete mp-MRIs exhibits superior performance compared to others in D_1 and D_2, which are the domains used to train these models. This suggests that the architecture can more effectively delineate tumor structures by utilizing correlations across all 4 image modalities.

Figure 7 presents segmentation examples where our model is capable of effectively segmenting the tumor, while models trained with fewer image modalities do not detect it. This example illustrates the benefits of employing complete mp-MRI images, as a single modality may not contain enough information. Additionally, it is included the ground-truth extension of the tumor, which was provided by doctors for the BraTS challenge using all four mp-MRI modalities, adhering to the annotation protocol outlined in [30].

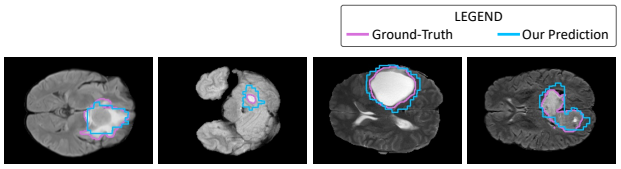

Figure 7. **Visual results.** Illustrated in this image are four cases where our proposed method accurately identifies the tumor areas, contrasting with models trained on fewer MRI modalities, which failed to detect them, hence evidencing the importance of utilizing the entire mp-MRI data.

How should we train the model? As we are removing the weights of the first layer of the SAM model, we analyze how we should train it. Is it better to first only tune the randomly initialized patch embedding, or should the training be performed jointly?

Table 3 compares the performance of two training strategies: first training the patch embedding to accommodate all four mp-MRI images simultaneously alongside the LoRA blocks. In contrast, the second method comprises training the patch embedding initially and then utilizing the resultant weights as a starting point to train the LoRA blocks, while keeping the patch embedding frozen. As depicted in Table 3, our one-step proposed approach significantly outperforms the two-step adaptation process across all datasets. This results highlight the need of specifically tuning the patch embedding with the PEFT strategy to effectivelly adapt to the medical domain.

	$VD_1 \uparrow$	$VD_2 \uparrow$	$D_3 \uparrow$	$D_4 \uparrow$
2 Step	30.36	32.77	19.36	26.84
Ours (TD_1)	**61.85**	**65.68**	**50.65**	**56.20**

Table 3. **Adaptation using complete mp-MRI: 1-step vs 2-step.** (KEY: VD_1 = Validation split for Adult Glioma, VD_2 = Validation split for Meningioma, D_3 = Pediatric Glioma, D_4 = Sub-Saharan Glioma.)

4.4. Effectiveness of our PEFT strategy

Table 4 presents a comparison between fine-tuning the entire SAM model and our proposed method outlined in Section 3. Our approach involves utilizing the 4 mp-MRI images as input, training the patch embedding, and applying PEFT on the LoRA blocks. Notably, our method achieves results up to 5 times better than fully fine-tuning the model across all validation sets, despite training only a small fraction of the model parameters. This significant improvement can be attributed to two factors: first, fully fine-tuning foundational models is not straightforward and typically leads to catastrophic forgetting. Second, the inherent generalization ability of foundational models allows them to adjust to new domains through slight weight shifts provided by LoRA.

	$VD_1 \uparrow$	$VD_2 \uparrow$	$D_3 \uparrow$	$D_4 \uparrow$
SAM FT	17.19	14.92	11.96	11.71
Ours	**61.85**	**65.68**	**50.65**	**56.20**

Table 4. **Fine-Tuning SAM vs LoRA Adaptation.** (KEY: VD_1 = Validation split for Adult Glioma, VD_2 = Validation split for Meningioma, D_3 = Pediatric Glioma, D_4 = Sub-Saharan Glioma.)

Experimental analysis of our proposed method We find that our model presents slightly better performance across the evaluated domains compared with the state-of-the-art SAM-based methods, as shown in Table 5. Our method is specially effective compared with the state-of-the-art on Pediatric and Sub-Saharan patients. Both sets are notoriously hard to segment due to their intrinsic discrepancies with the original training data (D1 or D2), see Section 3.2.

4.5. Explainability

Discussion of the adaptation to medical domain Figure 8 depicts the upper bound obtained from Eq 8 from models trained with different scans. We notice that the models

Model	$VD_1 \uparrow$	$VD_2 \uparrow$	$D_3 \uparrow$	$D_4 \uparrow$
SAM-Med [46]	47.2	74.7	52.8	42.3
Ranem et.al., [78]	61.9	-	-	-
Ours(TD_1)	**61.9**	65.7	50.7	**56.2**
Ours(TD_2)	58.4	**75.3**	**65.7**	52.4

Table 5. **SAM methods performance comparison.** (KEY: TD_1,VD_1 = Training, Validation splits for Adult Glioma; TD_2,VD_2 = Training, Validation splits for Meningioma; D_3 = Pediatric Glioma; D_4 = Sub-Saharan Glioma.) Not reported results indicated with "-".

present small discrepancy for their updates in the early attention layers despite being faced with different sources of input. We believe this indicates that the core gap between SAM and brain tumor scans lays in understanding the medical domain, rather than the discrepancies found on each of the MRI modalities. Additionally, we find that T2-FLAIR presents slight discrepancies with the alternatives, as evidenced by the greater gaps from these models with the alternative modalities. These results motivate the usage of T2-FLAIR in future frameworks. We believe these results can be specially useful for future ensemble frameworks which should incorporate at least one model employing T2-FLAIR with other modality to promote diversity across models.

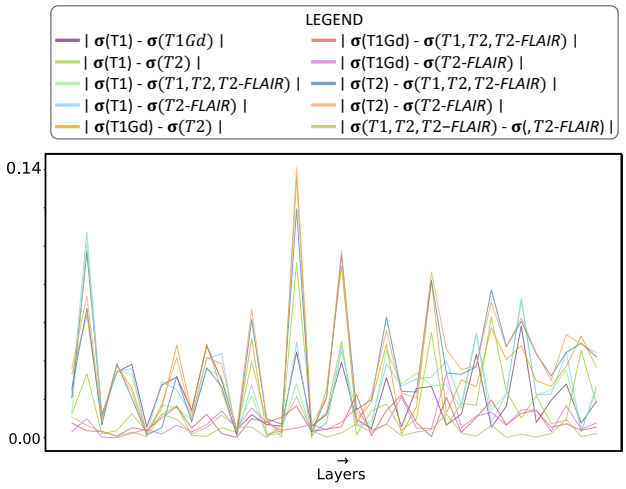

Figure 8. **Measuring distances of models trained with each MRI image.** Despite having different inputs, the process of tuning SAM to MRI is targeted mainly towards adapting the model to the medical domain regardless of the scan employed. This conclusion is observed by having relatively similar weight updates for the initial layers (which are presented with different types of scans), while presenting larger discrepancies for the last layers.

Discussion of the domain gap across datasets Figure 9 illustrates the feature plot derived from different tumor domains. While performance exhibits minor variations across these domains, the model discerns discrepancies within each, as it is not able to extract generic features to the broader brain tumor domain. This underscores SAM's susceptibility to domain gaps among brain tumor types, but its open-world recognition allows effective segmentation despite the existing disparities.

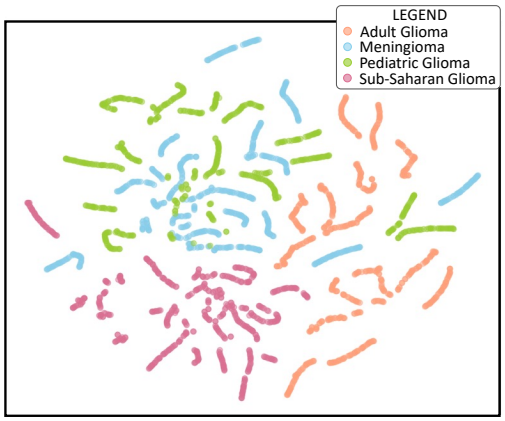

Figure 9. **Image Embedding TSNE Representation**. This image displays the features extracted by SAM encoder for each brain tumor type. Features of different classes are separated in the feature space, highlighting that the model does not recognize the broader brain tumor domain.

5. Conclusions

In this study, we have adapted SAM to process mp-MRI brain data, optimizing its performance by PEFT and also including attention to slice correlation. Additionally, we have examined the performance of SAM across various brain tumor domains and visualized how it discerns the biophysical inherent differences in each case through the extracted features. Our findings demonstrate that:

- When trained under identical conditions, using all modalities of mp-MRI brain images with SAM yields superior results compared to selecting only one or three modalities.
- Implementing PEFT on SAM using LoRA blocks enables the model to adapt to the domain instead of the specific MRI modality employed in each instance.
- The features extracted by SAM are not generic, reflecting variations across domains.

As future work, our direction is to further enhance the capability of SAM to generate domain agnostic features to the brain tumor domain. This direction could potentially improve its performance across a wider range of brain tumor types and facilitate its applicability in clinical settings.

Acknowledgement This work has been partially supported by the SEGA-CV (TED2021-131643A-I00) and the HVD (PID2021-125051OB-I00) projects of the Ministerio de Ciencia e Innovación of the Spanish Government.

References

[1] Q. Ostrom, D. Cote, M. Ascha, C. Kruchko, and J. Barnholtz-Sloan. Adult glioma incidence and survival by race or ethnicity in the united states from 2000 to 2014. *JAMA Oncology*, 4:1254–1262, 2018. 1

[2] Khadije Maajani, A. Jalali, S. Alipour, M. Khodadost, H. Tohidinik, and K. Yazdani. The global and regional survival rate of women with breast cancer: A systematic review and meta-analysis. *Clinical breast cancer*, 2019.

[3] Soheil Hassanipour, Hamed Delam, M. Arab-Zozani, E. Abdzadeh, S. Hosseini, Hossein-Ali Nikbakht, Mahdi Malakoutikhah, M. T. Ashoobi, M. Fathalipour, H. Salehiniya, and Shirin Riahi. Survival rate of prostate cancer in asian countries: A systematic review and meta-analysis. *Annals of Global Health*, 86, 2020.

[4] Ajay Aggarwal, Naomi Herz, Philip Campbell, Leo Arkush, Susan Short, and Jeremy Rees. Diagnostic delay and survival in high-grade gliomas–evidence of the 'waiting time paradox'? *British journal of neurosurgery*, 29(4):520–523, 2015. 1

[5] S. Boisclair, Shenae Samuels, L. Raez, D. Wietecha, A. Cohen, A. Khunger, K. King, Candice Sareli, Jorges Adames, J. Jacques, M. Salzberg, B. Hunis, M. Vulfovich, and A. Hussein. Ncog-34. an assessment of overall survival and palliative care use among patients with gliosarcoma diagnosed over a twelve-year span. *Neuro-oncology*, 22, 2020. 1

[6] Tomasz Tykocki and Mohamed Eltayeb. Ten-year survival in glioblastoma. a systematic review. *Journal of Clinical Neuroscience*, 54:7–13, 2018.

[7] PD Delgado-López and EM Corrales-García. Survival in glioblastoma: a review on the impact of treatment modalities. *Clinical and Translational Oncology*, 18(11):1062–1071, 2016. 1

[8] Claudia Allemani, Hannah K Weir, Helena Carreira, Rhea Harewood, Devon Spika, Xiao-Si Wang, Finian Bannon, Jane V Ahn, Christopher J Johnson, Audrey Bonaventure, et al. Global surveillance of cancer survival 1995–2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (concord-2). *The lancet*, 385(9972):977–1010, 2015. 1

[9] Michel P Coleman, Manuela Quaresma, Franco Berrino, Jean-Michel Lutz, Roberta De Angelis, Riccardo Capocaccia, Paolo Baili, Bernard Rachet, Gemma Gatta, Timo Hakulinen, et al. Cancer survival in five continents: a worldwide population-based study (concord). *The lancet oncology*, 9(8):730–756, 2008.

[10] Khadije Maajani, Arash Jalali, Sadaf Alipour, Mahmoud Khodadost, Hamid Reza Tohidinik, and Kamran Yazdani. The global and regional survival rate of women with breast cancer: a systematic review and meta-analysis. *Clinical Breast Cancer*, 19(3):165–177, 2019. 1

[11] Jiefeng Luo, Mika Pan, Ke Mo, Yingwei Mao, and Donghua Zou. Emerging role of artificial intelligence in diagnosis, classification and clinical management of glioma. In *Seminars in Cancer Biology*, 2023. 1, 2

[12] Chen Chen, Wei Wu, Cheng Chen, Fangfang Chen, Xiaogang Dong, Mingrui Ma, Ziwei Yan, Xiaoyi Lv, Yuhua Ma, and Min Zhu. Rapid diagnosis of lung cancer and glioma based on serum raman spectroscopy combined with deep learning. *Journal of Raman Spectroscopy*, 52(11):1798–1809, 2021. 1, 2

[13] Chris McKinnon, Meera Nandhabalan, Scott A Murray, and Puneet Plaha. Glioblastoma: clinical presentation, diagnosis, and management. *Bmj*, 374, 2021.

[14] Ron Batash, Noam Asna, Pamela Schaffer, Nicole Francis, and Moshe Schaffer. Glioblastoma multiforme, diagnosis and treatment; recent literature review. *Current medicinal chemistry*, 24(27):3002–3009, 2017. 1

[15] Ahmed Elnakib, Georgy Gimel'farb, Jasjit S Suri, and Ayman El-Baz. Medical image segmentation: a brief survey. *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies: Volume II*, pages 1–39, 2011. 1, 2, 6

[16] Dinesh D Patil and Sonal G Deore. Medical image segmentation: a review. *International Journal of Computer Science and Mobile Computing*, 2(1):22–27, 2013. 2

[17] KKD Ramesh, G Kiran Kumar, K Swapna, Debabrata Datta, and S Suman Rajest. A review of medical image segmentation algorithms. *EAI Endorsed Transactions on Pervasive Health and Technology*, 7(27), 2021. 2

[18] Gregory Sharp, Karl D Fritscher, Vladimir Pekar, Marta Peroni, Nadya Shusharina, Harini Veeraraghavan, and Jinzhong Yang. Vision 20/20: perspectives on automated image segmentation for radiotherapy. *Medical physics*, 41(5):050902, 2014.

[19] Xiuping Nie, Lilu Liu, Lifeng He, Liang Zhao, Haojian Lu, Songmei Lou, Rong Xiong, and Yue Wang. Weakly-interactive-mixed learning: less labelling cost for better medical image segmentation. *IEEE Journal of Biomedical and Health Informatics*, 2023.

[20] Li-Qiang Zhou, Jia-Yu Wang, Song-Yuan Yu, Ge-Ge Wu, Qi Wei, You-Bin Deng, Xing-Long Wu, Xin-Wu Cui, and Christoph F Dietrich. Artificial intelligence in medical imaging of the liver. *World journal of gastroenterology*, 25(6):672, 2019. 2

[21] Risheng Wang, Tao Lei, Ruixia Cui, Bingtao Zhang, Hongying Meng, and Asoke K Nandi. Medical image segmentation using deep learning: A survey. *IET Image Processing*, 16(5):1243–1267, 2022. 2, 6

[22] Mohammad Hesam Hesamian, Wenjing Jia, Xiangjian He, and Paul Kennedy. Deep learning techniques for medical image segmentation: achievements and challenges. *Journal of digital imaging*, 32:582–596, 2019. 1, 2

[23] Yichi Zhang, Zhenrong Shen, and Rushi Jiao. Segment anything model for medical image segmentation: Current applications and future directions. *Computers in Biology and Medicine*, page 108238, 2024.

[24] Qiumei Pu, Zuoxin Xi, Shuai Yin, Zhe Zhao, and Lina Zhao. Advantages of transformer and its application for medical image segmentation: a survey. *BioMedical Engineering OnLine*, 23(1):14, 2024. 1, 2

[25] Jie Liu, Yixiao Zhang, Jie-Neng Chen, Junfei Xiao, Yongyi Lu, Bennett A Landman, Yixuan Yuan, Alan Yuille, Yucheng Tang, and Zongwei Zhou. Clip-driven universal model for

organ segmentation and tumor detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21152–21164, 2023. 1, 3

[26] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning (ICML)*, pages 8748–8763. PMLR, 2021. 1, 2

[27] Chao Jia, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc Le, Yun-Hsuan Sung, Zhen Li, and Tom Duerig. Scaling up visual and vision-language representation learning with noisy text supervision. In *International Conference on Machine Learning (ICML)*, pages 4904–4916. PMLR, 2021.

[28] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *ICCV*, pages 9650–9660, 2021. 1, 2

[29] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *ICCV*, pages 4015–4026, 2023. 1, 2, 3

[30] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014. 1, 2, 4, 6, 7

[31] Dominic LaBella, Maruf Adewole, Michelle Alonso-Basanta, Talissa Altes, Syed Muhammad Anwar, Ujjwal Baid, Timothy Bergquist, Radhika Bhalerao, Sully Chen, Verena Chung, et al. The asnr-miccai brain tumor segmentation (brats) challenge 2023: Intracranial meningioma. *arXiv preprint arXiv:2305.07642*, 2023. 3, 4, 5, 6

[32] Anahita Fathi Kazerooni, Nastaran Khalili, Xinyang Liu, Debanjan Haldar, Zhifan Jiang, Syed Muhammed Anwar, Jake Albrecht, Maruf Adewole, Udunna Anazodo, Hannah Anderson, et al. The brain tumor segmentation (brats) challenge 2023: Focus on pediatrics (cbtn-connect-dipgr-asnr-miccai brats-peds). *ArXiv*, 2023. 3, 4, 5, 6

[33] Maruf Adewole, Jeffrey D Rudie, Anu Gbdamosi, Oluyemisi Toyobo, Confidence Raymond, Dong Zhang, Olubukola Omidiji, Rachel Akinola, Mohammad Abba Suwaid, Adaobi Emegoakor, et al. The brain tumor segmentation (brats) challenge 2023: Glioma segmentation in sub-saharan africa patient population (brats-africa). *ArXiv*, 2023. 2, 3, 4, 5, 6

[34] Yuxuan Ding, Lingqiao Liu, Chunna Tian, Jingyuan Yang, and Haoxuan Ding. Don't stop learning: Towards continual learning for the clip model. *ArXiv*, abs/2207.09248, 2022. 2

[35] Shipeng Yan, Lanqing Hong, Hang Xu, Jianhua Han, T. Tuytelaars, Zhenguo Li, and Xuming He. Generative negative text replay for continual vision-language pretraining. *ArXiv*, abs/2210.17322, 2022.

[36] Chong Zhou, Chen Change Loy, and Bo Dai. Extract free dense labels from clip. In *ECCV*, 2022. 2

[37] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021. 2, 4

[38] Ujjwal Baid, Satyam Ghodasara, Suyash Mohan, Michel Bilello, Evan Calabrese, Errol Colak, Keyvan Farahani, Jayashree Kalpathy-Cramer, Felipe C Kitamura, Sarthak Pati, et al. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv preprint arXiv:2107.02314*, 2021. 2, 3, 4, 6

[39] Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint arXiv:1811.02629*, 2018. 3, 4, 6

[40] Udunna C Anazodo, Jinggang J Ng, Boaz Ehiogu, Johnes Obungoloch, Abiodun Fatade, Henk JMM Mutsaerts, Mario Forjaz Secca, Mamadou Diop, Abayomi Opadele, Daniel C Alexander, et al. A framework for advancing sustainable magnetic resonance imaging access in africa. *NMR in Biomedicine*, 36(3):e4846, 2023. 2, 3, 4, 5

[41] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 2, 3

[42] Sheng He, Rina Bao, Jingpeng Li, P Ellen Grant, and Yangming Ou. Accuracy of segment-anything model (sam) in medical image segmentation tasks. *arXiv preprint arXiv:2304.09324*, 2023. 2

[43] Yuhao Huang, Xin Yang, Lian Liu, Han Zhou, Ao Chang, Xinrui Zhou, Rusi Chen, Junxuan Yu, Jiongquan Chen, Chaoyu Chen, et al. Segment anything model for medical images? *Medical Image Analysis*, 92:103061, 2024. 2

[44] Guoyao Deng, Ke Zou, Kai Ren, Meng Wang, Xuedong Yuan, Sancong Ying, and Huazhu Fu. Sam-u: Multi-box prompts triggered uncertainty estimation for reliable sam in medical image. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 368–377. Springer, 2023. 2

[45] Tassilo Wald, Saikat Roy, Gregor Koehler, Nico Disch, Maximilian Rouven Rokuss, Julius Holzschuh, David Zimmerer, and Klaus Maier-Hein. Sam. md: Zero-shot medical image segmentation capabilities of the segment anything model. In *Medical Imaging with Deep Learning, short paper track*, 2023. 2

[46] Junlong Cheng, Jin Ye, Zhongying Deng, Jianpin Chen, Tianbin Li, Haoyu Wang, Yanzhou Su, Ziyan Huang, Jilong Chen, Lei Jiang, et al. Sam-med2d. *arXiv preprint arXiv:2308.16184*, 2023. 2, 8

[47] Shizhan Gong, Yuan Zhong, Wenao Ma, Jinpeng Li, Zhao Wang, Jingyang Zhang, Pheng-Ann Heng, and Qi Dou. 3dsam-adapter: Holistic adaptation of sam from 2d to 3d for promptable medical image segmentation. *arXiv preprint arXiv:2306.13465*, 2023. 2, 3

[48] Haoyu Wang, Sizheng Guo, Jin Ye, Zhongying Deng, Junlong Cheng, Tianbin Li, Jianpin Chen, Yanzhou Su, Ziyan

Huang, Yiqing Shen, et al. Sam-med3d. *arXiv preprint arXiv:2310.15161*, 2023. 3

[49] Junde Wu, Rao Fu, Huihui Fang, Yuanpei Liu, Zhaowei Wang, Yanwu Xu, Yueming Jin, and Tal Arbel. Medical sam adapter: Adapting segment anything model for medical image segmentation. *arXiv preprint arXiv:2304.12620*, 2023. 2, 3, 4

[50] Cheng Chen, Juzheng Miao, Dufan Wu, Zhiling Yan, Sekeun Kim, Jiang Hu, Aoxiao Zhong, Zhengliang Liu, Lichao Sun, Xiang Li, et al. Modality-agnostic sam adaptation for 3d medical image segmentation. *arXiv preprint arXiv:2309.08842*, 3, 2023. 2, 3

[51] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017. 2

[52] Elad Ben Zaken, Shauli Ravfogel, and Yoav Goldberg. Bitfit: Simple parameter-efficient fine-tuning for transformer-based masked language-models. *arXiv preprint arXiv:2106.10199*, 2021. 2, 3

[53] Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. Qlora: Efficient finetuning of quantized llms. *Advances in Neural Information Processing Systems*, 36, 2024. 2, 3

[54] Ying Sheng, Shiyi Cao, Dacheng Li, Coleman Hooper, Nicholas Lee, Shuo Yang, Christopher Chou, Banghua Zhu, Lianmin Zheng, Kurt Keutzer, et al. S-lora: Serving thousands of concurrent lora adapters. *arXiv preprint arXiv:2311.03285*, 2023. 2

[55] Arnaud Gucciardi, Safouane El Ghazouali, Francesca Venturini, Vida Groznik, and Umberto Michelucci. Symbrain: A large-scale dataset of mri images for neonatal brain symmetry analysis. *arXiv preprint arXiv:2401.11814*, 2024. 3

[56] Samuel G Armato III, Henkjan Huisman, Karen Drukker, Lubomir Hadjiiski, Justin S Kirby, Nicholas Petrick, George Redmond, Maryellen L Giger, Kenny Cha, Artem Mamonov, et al. Prostatex challenges for computerized classification of prostate lesions from multiparametric magnetic resonance images. *Journal of Medical Imaging*, 5(4):044501–044501, 2018. 3

[57] Thorsten M Buzug. Computed tomography. In *Springer handbook of medical technology*, pages 311–342. Springer, 2011. 3

[58] Lee W Goldman. Principles of ct and ct technology. *Journal of nuclear medicine technology*, 35(3):115–128, 2007.

[59] Yucheng Tang, Dong Yang, Wenqi Li, Holger R Roth, Bennett Landman, Daguang Xu, Vishwesh Nath, and Ali Hatamizadeh. Self-supervised pre-training of swin transformers for 3d medical image analysis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20730–20740, 2022. 3

[60] Wenhui Lei, Xu Wei, Xiaofan Zhang, Kang Li, and Shaoting Zhang. Medlsam: Localize and segment anything model for 3d medical images. *arXiv preprint arXiv:2306.14752*, 2023. 3

[61] Chenglong Wang, Dexuan Li, Sucheng Wang, Chengxiu Zhang, Yida Wang, Yun Liu, and Guang Yang. Sammed: A medical image annotation framework based on large vision model. *arXiv preprint arXiv:2307.05617*, 2023. 3

[62] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4(1):1–13, 2017. 3, 5

[63] McKinsey L Goodenberger and Robert B Jenkins. Genetics of adult glioma. *Cancer genetics*, 205(12):613–621, 2012.

[64] Judith A Schwartzbaum, James L Fisher, Kenneth D Aldape, and Margaret Wrensch. Epidemiology and molecular pathology of glioma. *Nature clinical practice Neurology*, 2(9):494–503, 2006. 3, 5

[65] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *CVPR*, pages 16000–16009, 2022. 3

[66] Ali-Reza Fathi and Ulrich Roelcke. Meningioma. *Current neurology and neuroscience reports*, 13:1–8, 2013. 5

[67] Joseph Wiemels, Margaret Wrensch, and Elizabeth B Claus. Epidemiology and etiology of meningioma. *Journal of neuro-oncology*, 99:307–314, 2010.

[68] Elizabeth B Claus, Melissa L Bondy, Joellen M Schildkraut, Joseph L Wiemels, Margaret Wrensch, and Peter M Black. Epidemiology of intracranial meningioma. *Neurosurgery*, 57(6):1088–1095, 2005. 5

[69] Valentina Di Ruscio, Giada Del Baldo, Francesco Fabozzi, Maria Vinci, Antonella Cacchione, Emmanuel de Billy, Giacomina Megaro, Andrea Carai, and Angela Mastronuzzi. Pediatric diffuse midline gliomas: An unfinished puzzle. *Diagnostics*, 12(9):2064, 2022. 5

[70] Stefano Gabriele Vallero, Luca Bertero, Giovanni Morana, Paola Sciortino, Daniele Bertin, Anna Mussano, Federica Silvia Ricci, Paola Peretta, and Franca Fagioli. Pediatric diffuse midline glioma h3k27-altered: A complex clinical and biological landscape behind a neatly defined tumor type. *Frontiers in Oncology*, 12:7516, 2023. 5

[71] Tabitha M Cooney, Evan Lubanszky, Rachna Prasad, Cynthia Hawkins, and Sabine Mueller. Diffuse midline glioma: review of epigenetics. *Journal of Neuro-Oncology*, 150:27–34, 2020. 5

[72] Dong Zhang, Raymond Confidence, and Udunna Anazodo. Stroke lesion segmentation from low-quality and few-shot mris via similarity-weighted self-ensembling framework. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 87–96. Springer, 2022. 5

[73] Gene H Golub and Charles F Van Loan. *Matrix computations*. JHU press, 2013. 5

[74] Jeroen Bertels, Tom Eelbode, Maxim Berman, Dirk Vandermeulen, Frederik Maes, Raf Bisschops, and Matthew B Blaschko. Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen,*

China, October 13–17, 2019, Proceedings, Part II 22, pages 92–100. Springer, 2019. 6

[75] Annika Reinke, Minu D Tizabi, Carole H Sudre, Matthias Eisenmann, Tim Rädsch, Michael Baumgartner, Laura Acion, Michela Antonelli, Tal Arbel, Spyridon Bakas, et al. Common limitations of image processing metrics: A picture story. *arXiv preprint arXiv:2104.05642*, 2021. 6

[76] Sabarinath Mahadevan, Paul Voigtlaender, and Bastian Leibe. Iteratively trained interactive segmentation. *arXiv preprint arXiv:1805.04398*, 2018. 6

[77] Zheng Lin, Zhao Zhang, Lin-Zhuo Chen, Ming-Ming Cheng, and Shao-Ping Lu. Interactive image segmentation with first click attention. In *CVPR*, pages 13339–13348, 2020. 6

[78] Amin Ranem, Niklas Babendererde, Moritz Fuchs, and Anirban Mukhopadhyay. Exploring sam ablations for enhancing medical segmentation in radiology and pathology. *arXiv preprint arXiv:2310.00504*, 2023. 8