Colorectal Tumor Segmentation using SAM-Adapter: A Fine-tuning Approach

Sony Annem, Dr.Khalid Niazi, Ziyu Su, Dr.Metin Gurcan, Dr.Mostafa Rezapour, UsamaSajjad, Dr.Bayram

ABSTRACT

In this project we mostly focused on applying a cutting-edge technique known as SAM-Adapter to identify and segment tumor buds present in the histopathology images of Colorectal cancer. Colorectal cancer (CRC) is the 3rd most common form of cancer preceding lung and breast cancers, respectively. Segmentation is a crucial step in determining Colorectal Cancer (CRC) staging through the analysis of digital pathology images. To analyze tumors, present in the images we finetuned the Segment Anything Model (SAM, released in April-2023 by Meta AI), rather than training the entire model to minimize the cost we just inserted Adapter module (feedforward layers) in between the layers of model and only trained the Adapters by freezing the weights of SAM's original architecture. We compared the outcomes of training a basic adapter based on a random click prompt versus training the adapter using task-specific information, training adapters by feeding them task-specific knowledge led to significantly improved performance in segmenting tumor buds.

Keywords: Colorectal Cancer pathology images, Segment Anything, Adapters, task-specific information.

1. INTRODUCTION

On April 5th, Meta AI gained attention by releasing the first foundational model on images called Segment Anything Model [1], it mainly focuses on the promptable segmentation tasks. The performance of SAM is awesome on natural segmentation tasks, but when it comes to medical images, especially histopathological images, it is very hard to come up with consistent segmentation criteria and, we get to know that SAM is based on prompts, prompting histopathology images is challenging task, so we came up to fine tune the SAM model suitable for histopathology images to make it work without prompting. In this research work we concentrated primarily on the colorectal cancer pathology dataset. Colorectal cancer is one of the most worrying types of cancer since it can affect either the colon or the rectum[2], which are parts of the digestive system. It is also commonly referred to as colon cancer or rectal cancer, depending on the location of the tumor. Segmenting colorectal cancer histopathology images offers numerous benefits, ranging from improving diagnostic accuracy and treatment planning to advancing research and drug development in the fight against colorectal cancer. With the help of SAM-Adapter [3] by passing task specific information into adapters, we were able to detect and segment the majority of tumor buds present in the images, which gave us very positive results.

Related Work

Semantic Segmentation

In recent years, tremendous progress in the medical area regarding the analysis of tumor buds has been accomplished using semantic segmentation[4]. It is a method of computer vision that involves the process of dividing an image into many segments or regions and then assigning a meaningful label to each pixel that is contained within those segments. The tumors that are present in the image are given a meaningful label when using semantic segmentation.

SAM Architecture

SAM simplified the process of semantic segmentation by introducing its first foundational model on images especially for the segmentation tasks. SAM architecture has mainly three components Image Encoder, Prompt Encoder, Mask Decoder. At the Image Encoder part, it generates image embeddings by using masked auto-encoder, MAE[5] , pre-trained Vision Transformer (ViT). The prompt encoder encodes background points, masks, bounding boxes, or texts into an embedding vector in real time. The research considers two sets of prompts: sparse (points, boxes, text) and dense (masks). A lightweight mask decoder predicts the segmentation masks based on the embeddings from both the image and prompt encoder.

Fine-Tuning

When we already have a deep learning model that has been pre-trained on a huge dataset, but if we want to adapt it to a new task, we may apply a transfer learning approach called fine-tuning. We may avoid training the model from scratch by beginning with the weights of a pre-trained model and then continuing training it on the new data set to finetune its parameters for the particular task. Coming back to the SAM model, however, we decided not to mess with the original architectural weights and instead used the idea of adapters to train for our downstream task.

Adapters

Adapters are nothing but typically small neural networks that are inserted between the pre-trained model's existing layers. It consists of fully connected layers with activation function when pretraining a model using adapters, the process only updates the parameters in the adapter layers while keeping the rest of the pre-trained model unchanged.

Dataset Details

The CRC digital pathology images were obtained from The Ohio State University Wexner Medical Center (OSUMC) located in Ohio, United States. To train the model, we employed mountaged images, which prevents the model from being trained primarily on memorizing the position of tumor buds, we used 10k mountaged images with a resolution of 1024 * 1024 for the training part and 1,243 images for testing.

Implementation

As part of implementation, we initially tried to identify tumor buds by using the Segment Anything Model[6] the results were very disappointing, and since SAM is based on prompts and prompting histopathology images for normal people is a difficult task, we came up with an idea to fine tune the SAM model so that it is suitable for histopathology images and enables it to function without prompting[7]. This was accomplished by inserting simple Adapters in between the layers of the image encoder and at Image Decoder. At first during the training phase, we made use of a random click prompt[8]; positive clicks indicated foreground regions, while negative clicks indicated background regions. This simple implementation results in satisfiable performance to segment tumor regions in the images. Rather than having a click prompt during the training process, we input Task-Specific Information into the Adapters, the task specific information is element wise addition of the patch embeddings and High frequency Component of image following the same setting in[9]; by training the model using this strategy improves performance compared to click prompt.

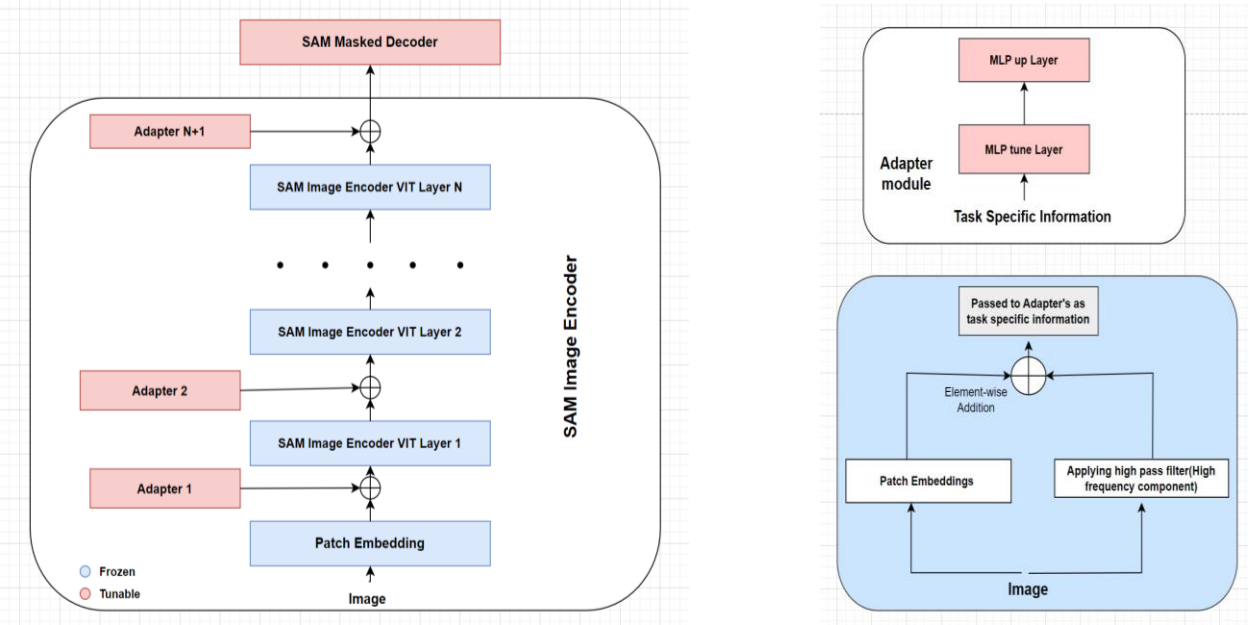

Figure 1: Architecture of SAM-Adapter and about Task-specific Details that passed to Adapter's

By using SAM as a backbone having VIT-H/16 as image encoder with the help of SAM-adapter we implemented segmentation model for colorectal pathology images, we followed the same architecture as SAM-Adapter.

Task Specific Information

Task specific information is the input to Adapters, where task specific info is generated by adding patch embeddings and high frequency component. Here patch embeddings mean dividing images into patch's then collecting embedding in vectors and High frequency component means applying high pass filter on the images in order to achieve edges, texture and fine details of the image, same as SAM-Adapter paper we used Fast Fourier transform to get the high pass component, then converting high frequency component image into feature vectors of the same dimension as the patch embedding's feature vectors. Once both sources have the same dimensionality, perform element-wise addition to combine them.

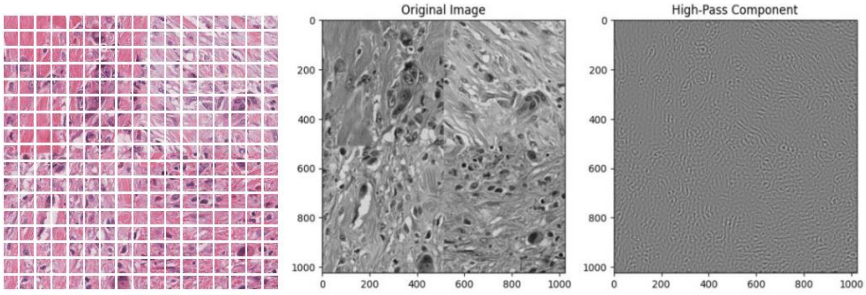

Figure 2: Image Patches and High pass Component

Adapter modules consist of two layers such as MLP^{i}_{tune} which is used to generate task specific and MLP_{up} is used to adjust the dimensions to make it equal with the VITH layer. We were able to obtain successful outcomes by utilizing Adapters and providing task related information.

Results

The SAM model was loaded with three distinct encoders: the ViT-B encoder, which had 91 M parameters; the ViT-L encoder, which had 308 M parameters; and the ViT-H encoder, which had 636 M parameters. We trained our dataset on VIT-B and VIT-H, and we observed that the results obtained by VIT-H were much superior to those obtained by VIT-B. The model was trained and evaluated using regular images as well as mountaged images with a resolution of 1024 by 1024 pixels for each image, it taken two days to run the model on Nvidia GPU for 35 epochs. We compared the results of SAM-Adapter training with task-specific information with those of Automatic SAM model and with the results of basic SAM-Adapter training with random clicks.

Vanilla SAM: We used the SAM website that was made accessible by Meta AI; nonetheless, it was quite challenging to locate the tumor bud.

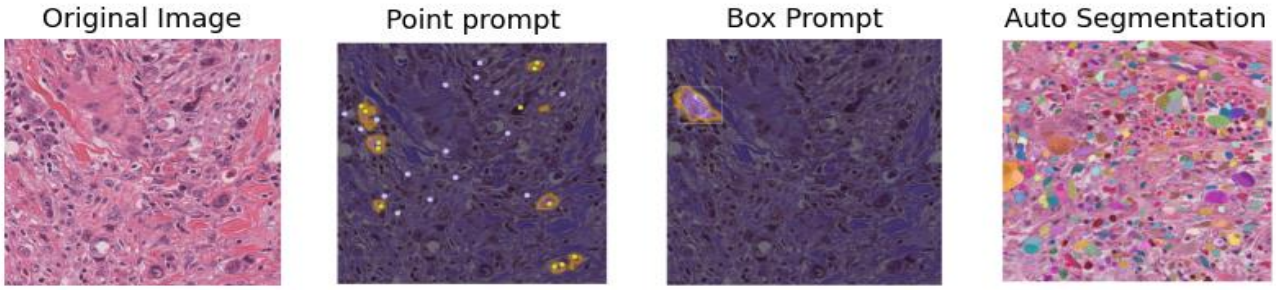

Figure3: Vanilla SAM

SAM-Adapter training using random click prompt: Compared to SAM model, the results of the SAM-Adapter when training with random click prompt get better results but it missed some of the True Positives.

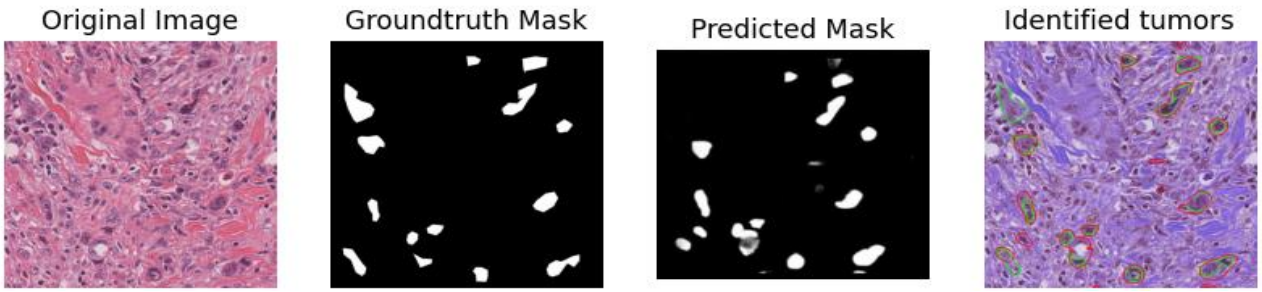

Figure 4: SAM- Adapter, training image Encoder with random click prompt [Green contours:GT, Red contours:Predicted]

SAM-Adapter training using Task-Specific Information: Training by passing task specific information into the Adapters achieves best results, this method perfectly predicted each and every tumor bud without prompt.

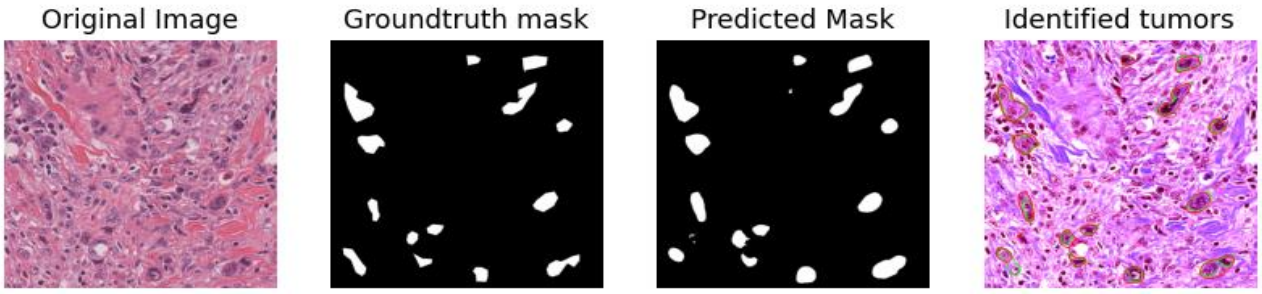

Figure 5: SAM- Adapter, training using task specific information, [Green contours:GT, Red contours:Predicted]

Some sample results:

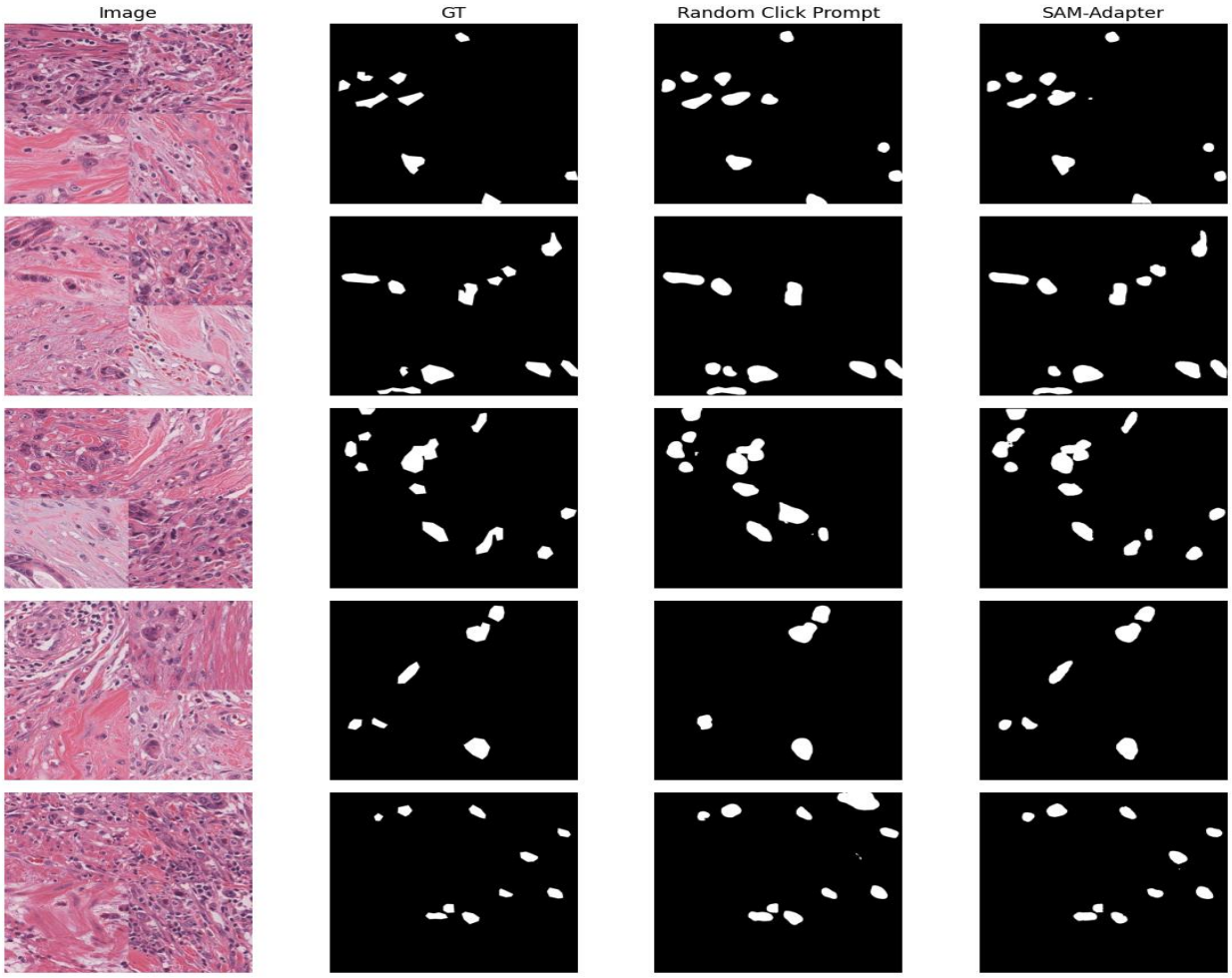

Figure 6: Comparison with the results of training Adapter using random click prompt versus using task specific information. When we clearly observe the results, While the Basic SAM Adapter (random click prompt) failed to identify some of the True Positive tumor buds, the SAM-Adapter was able to accurately forecast all of the tumorbuds.

We used two types of quantitative measures such as Intersection Over Union (IOU) and Dice Coefficient, SAM-Adapter training with task-specific outcomes came near UNET++ on 512 size images. UNET++ is not good for the resolution of 1024, because UNET has an input size limit of 572*572, it is not suitable for segmentation tasks with an input image size of 1024*1024. The size of medical images may be exceedingly vast, taking smaller size images or resizing them are not the viable options coming to histopathology images. When you resize histopathological images, you will notice a decrease in the image's overall quality.

Architecture	IOU	DICE
UNET ++ on 512 resolution	0.60	0.755
Basic SAM- Adapter	0.59	0.678
SAM-Adapter with task specific information	0.65	0.78

When it comes to the tumor bud identification measure, the typical overlap dice coefficient is not a good option because creating accurate and consistent ground truth annotations can be challenging. So, we modified the dice Coefficient instead of taking the intersection match we counted the predicted tumors, first we labelled all the connected components on both true_masks and pred_masks. This labeling assigns a unique integer label to each connected region in the binary masks. Connected regions are regions where pixels are connected and have the same value (in this case, 1 for tumor regions) if there is match in connected component then it is taken as count for true positive, otherwise considered as false negative, finally the dice is calculated based on the count of tumors predicted.

Architecture	IOU	DICE
Basic SAM- Adapter	0.59	0.87
SAM-Adapter with task specific information	0.65	0.9160

Conclusion

Segmenting Colorectal tumor bud is still one of the challenging tasks, finetuning SAM model using Adapters by passing task specific information provides remarkable results and using mountaged images helps the model to learn features instead of memorizing the location of tumorbuds. In the future we are going to implement segmenting on the whole slide colorectal tumor bud and also, we are going to finetune entire SAM model on histopathology images without adapters for sole purpose of segmenting histopathology images.

References

[1] A. Kirillov *et al.*, "Segment anything," *arXiv preprint arXiv:2304.02643*, 2023.

[2] R. L. Siegel, A. Jemal, and E. M. Ward, "Increase in incidence of colorectal cancer among young men and women in the United States," *Cancer Epidemiology Biomarkers & Prevention*, vol. 18, no. 6, pp. 1695-1698, 2009.

[3] T. Chen *et al.*, "SAM Fails to Segment Anything?--SAM-Adapter: Adapting SAM in Underperformed Scenes: Camouflage, Shadow, and More," *arXiv preprint arXiv:2304.09148*, 2023.

[4] J. Wang, J. D. MacKenzie, R. Ramachandran, and D. Z. Chen, "A deep learning approach for semantic segmentation in histology tissue images," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, 2016: Springer, pp. 176-184.

[5] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 16000-16009.

[6] S. He, R. Bao, J. Li, P. E. Grant, and Y. Ou, "Accuracy of segment-anything model (sam) in medical image segmentation tasks," *arXiv preprint arXiv:2304.09324*, 2023.

[7] R. Deng *et al.*, "Segment anything model (sam) for digital pathology: Assess zero-shot segmentation on whole slide imaging," *arXiv preprint arXiv:2304.04155*, 2023.

[8] J. Wu *et al.*, "Medical sam adapter: Adapting segment anything model for medical image segmentation," *arXiv preprint arXiv:2304.12620*, 2023.

[9] W. Liu, X. Shen, C.-M. Pun, and X. Cun, "Explicit visual prompting for low-level structure segmentations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 19434-19445.