Input Augmentation with SAM: Boosting Medical Image Segmentation with Segmentation Foundation Model

Yizhe Zhang[1], Tao Zhou[1], Shuo Wang[2,3], Peixian Liang[4], Yejia Zhang[4], Danny Z. Chen[4]

[1] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China
yizhe.zhang.cs@gmail.com, taozhou.ai@gmail.com
[2] Digital Medical Research Center, School of Basic Medical Sciences, Fudan University, Shanghai 200032, China
[3] Shanghai Key Laboratory of MICCAI, Shanghai, Shanghai 200032, China
shuowang@fudan.edu.cn
[4] Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, USA
pliang@nd.edu, yzhang46@nd.edu, dchen@nd.edu

Abstract. The Segment Anything Model (SAM) is a recently developed large model for general-purpose segmentation for computer vision tasks. SAM was trained using 11 million images with over 1 billion masks and can produce segmentation results for a wide range of objects in natural scene images. SAM can be viewed as a general perception model for segmentation (partitioning images into semantically meaningful regions). Thus, how to utilize such a large foundation model for medical image segmentation is an emerging research target. This paper shows that although SAM does not immediately give high-quality segmentation for medical image data, its generated masks, features, and stability scores are useful for building and training better medical image segmentation models. In particular, we demonstrate how to use SAM to augment image input for commonly-used medical image segmentation models (e.g., U-Net). Experiments on three segmentation tasks show the effectiveness of our proposed SAMAug method.

1 Introduction

The Segment Anything Model (SAM) [10] is a remarkable recent advance in foundation models for computer vision tasks. SAM was trained using 11 million images and over 1 billion masks. Despite its strong capability in producing segmentation for a wide variety of objects, several studies [8,4,28] showed that SAM is not powerful enough for segmentation tasks that require domain expert knowledge (e.g., medical image segmentation).

For a given medical image segmentation task with image and annotation pairs, we aim to build and train a medical image segmentation model, denoted

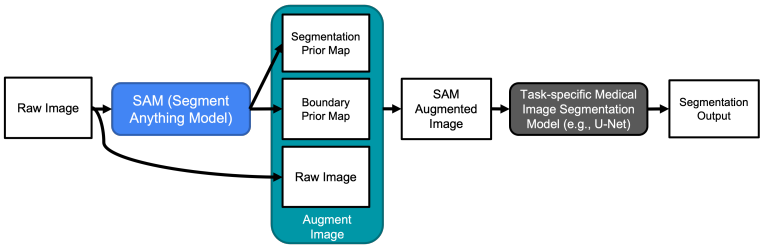

Fig. 1. Input augmentation with SAM for boosting medical image segmentation.

by $\mathcal{M}$, on top of the segmentation foundation model SAM. We propose a new method called SAMAug that directly utilizes the segmentation masks (with stability scores) generated by SAM to augment the raw inputs of the medical image segmentation model $\mathcal{M}$. The input augmentation is performed by a fusion function. The inference process (with SAMAug) for a given image is illustrated in Fig. 1. The task-specific medical image segmentation model $\mathcal{M}$ is trainable using a specific dataset[5] (e.g., MoNuSeg [11]). The parameters of SAM remain fixed, the fusion (augmentation) function is a parameter-free module, and the learning process aims to update the parameters of $\mathcal{M}$ with respect to the given foundation model SAM, the fusion function, and the training data.

Our main contributions can be summarized as follows. (1) We identify that the emerging segmentation foundation model SAM can provide attention (prior) maps for downstream segmentation tasks. (2) With a simple and novel method (SAMAug), we combine segmentation outputs of SAM with raw image inputs, generating SAM-augmented input images for building downstream medical image segmentation models. (3) We conduct comprehensive experiments to demonstrate that our proposed method is effective for both CNN and Transformer segmentation models in three medical image segmentation tasks.

2 Related Work

Data Augmentation. Data augmentation (DA) has been widely used in training medical image segmentation models [27,3]. A main aim of DA is to synthesize new views of existing samples in training data. Our SAMAug can be viewed as a type of DA technique. Unlike previous DA methods which often use hand-designed transformations (e.g., rotation, cropping), SAMAug utilizes a segmentation foundation model to augment raw images, aiming to impose semantically useful structures to the input of a medical image segmentation model.

Image Enhancement. From the image enhancement (IE) view point, SAMAug enhances images by adding semantic structures from a segmentation foundation model. A critical difference between SAMAug and the previous enhancement methods [16,5] is that traditional IE often works at a low level, e.g., de-blurring

[5] SAMAug performs on all images, including training images and testing images.

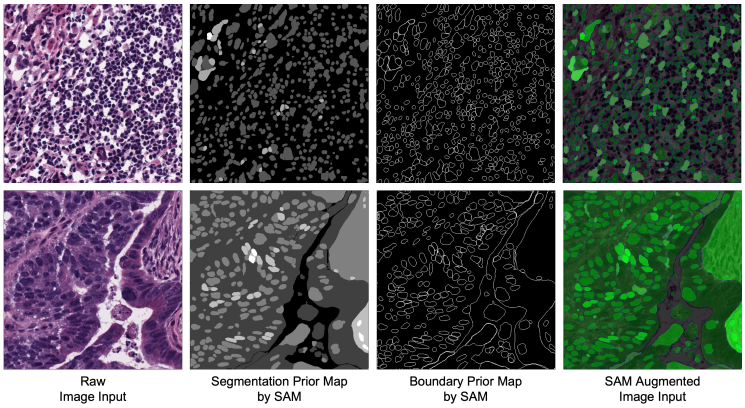

Fig. 2. Visual examples of a raw input image, its segmentation prior map by SAM, boundary prior map by SAM, and SAM-augmented image input (illustrated in Fig. 1). The image sample is from the MonuSeg dataset [11].

and noise reduction, and the purpose of enhancement is to reconstruct and recover. In contrast, SAMAug aims to add high-level structures to raw images, providing better semantics for the subsequent medical image segmentation model.

Recent SAM-related Methods. Since the introduction of SAM, many attempts have been made to understand and utilize SAM for medical image analysis (e.g., [6,28,12,24,25,14]). Recent work has shown that SAM alone, without further fine-tuning and/or adaptation, often delivers unsatisfied results for medical image segmentation tasks [6,28]. In order to utilize SAM more effectively, Ma et al. [12] proposed to fine-tune SAM using labeled images. Wu et al. [24] proposed to add additional layers to adapt SAM for a medical image segmentation task. Compared with these fine-tuning and adaptation methods, our method is more efficient in computation and memory costs during model training. In test time, these fine-tuning, adapting, and augmentation methods all require performing forward propagation of test images through SAM.

3 Methodology

In Section 3.1, we describe the two key image representations obtained by applying SAM to a medical image, a segmentation prior map and a boundary prior map. In Section 3.2, we show how to augment a medical image using the two obtained prior maps. In Section 3.3, we present the details of using augmented images in training a medical image segmentation model. Finally, in Section 3.4, we show how to use the trained model in model deployment (model testing).

3.1 Segmentation and Boundary Prior Maps

In the grid prompt setting, SAM uses a grid prompt to generate segmentation masks for a given image. That is, segmentation masks are generated at all plau-

sible locations in the image. The generated segmentation masks are then stored in a list. For each segmentation mask in the list, we draw the mask on a newly created segmentation prior map using the value suggested by the mask's corresponding stability score (generated by SAM). In addition to the segmentation prior map, we further generate a boundary prior map according to the masks provided by SAM. We draw the exterior boundary of each segmentation mask in the mask list and put all the boundaries together to form a boundary prior map. For a given image x, we generate two prior maps, $\text{prior}_{\text{seg}}$ and $\text{prior}_{\text{boundary}}$, using the process discussed above. In Fig. 2 (the second and third columns), we give visual examples of these two prior maps thus generated.

3.2 Augmenting Input Images

With the prior maps generated, our next step is to augment the input image x with the generated prior maps. We choose a simple method for this augmentation: adding the prior maps to the raw image. Note that many medical image segmentation tasks can be reduced to a three-class segmentation task in which the 1st class corresponds to the background, the 2nd class corresponds to the regions of interest (ROIs), and the 3rd class corresponds to the boundaries between the ROIs and background. We add the segmentation prior map to the second channel of the raw image and the boundary prior map to the third channel of the raw image. If the raw image is in gray-scale, we create a 3-channel image with the first channel consisting of the gray-scale raw image, the second channel consisting of its segmentation prior map (only), and the third channel consisting of its boundary prior map (only). For each image x in the training set, we generate its augmented version $x^{aug} = \text{Aug}(\text{prior}_{\text{seg}}, \text{prior}_{\text{boundary}}, x)$. Fig. 2 (the fourth column) gives a visual example of the SAM-augmented image input.

3.3 Model Training with SAM-Augmented Images

With the input augmentation on each image sample in the training set, we obtain a new augmented training set $\{(x_1^{aug}, y_1), (x_2^{aug}, y_2), \ldots, (x_n^{aug}, y_n)\}$, where $x_i^{aug} \in \mathbb{R}^{w \times h \times 3}$, $y_i \in \{0,1\}^{w \times h \times C}$ is the annotation of the input image x_i, and C is the number of classes for the segmentation task. A common medical image segmentation model $\mathcal{M}$ (e.g., a U-Net) can be directly utilized for learning from the augmented training set. A simple way to learn from SAM-augmented images is to use the following learning objective with respect to the parameters of $\mathcal{M}$:

$$\sum_{i=1}^{n} loss(\mathcal{M}(x_i^{aug}), y_i). \tag{1}$$

The above objective only uses SAM-augmented images for model training. Consequently, in model testing, the trained model accepts only images augmented by SAM. In situations where SAM fails to give plausible prior maps, we consider training a segmentation model using both raw images and images with

SAM augmentation. The new learning objective is to minimize the following objective with respect to the parameters of $\mathcal{M}$:

$$\sum_{i=1}^{n} \beta loss(\mathcal{M}(x_i), y_i) + \lambda loss(\mathcal{M}(x_i^{aug}), y_i), \tag{2}$$

where β and λ control the importance of the training loss for samples with raw images and samples with augmented images. When setting $\beta = 0$ and $\lambda = 1$, the objective function in Eq. (2) is reduced to Eq. (1). By default, we set both β and λ equal to 1. The spatial cross-entropy loss or Dice loss can be used for constructing the loss function in Eq. (1) and Eq. (2). An SGD-based optimizer (e.g., Adam [9]) can be applied to reduce the values of the loss function.

3.4 Model Deployment with SAM-Augmented Images

When the segmentation model is trained using only SAM-augmented images, the model deployment (testing) requires the input also to be SAM-augmented images. The model deployment can be written as:

$$\hat{y} = \tau(\mathcal{M}(x^{aug})), \tag{3}$$

where τ is an output activation function (e.g., a sigmoid function, a softmax function), and x^{aug} is a SAM-augmented image (as described in Section 3.2). When the segmentation model $\mathcal{M}$ is trained using both raw images and SAM-augmented images, we identify new opportunities in inference time to fully realize the potential of the trained model. A simple way of using $\mathcal{M}$ would be to apply sample inference twice for each test sample: The first time inference uses the raw image x as input and the second time inference uses its SAM augmented image as input. The final segmentation output can be generated by the average ensemble of the two outputs. Formally, this inference process can be written as:

$$\hat{y} = \tau(\mathcal{M}(x) + \mathcal{M}(x^{aug})). \tag{4}$$

Another way of utilizing the two output candidates $\mathcal{M}(x)$ and $\mathcal{M}(x^{aug})$ is to select a plausible segmentation output from these two candidates:

$$\hat{y} = \tau(\mathcal{M}(x^*)), \tag{5}$$

where x^* is obtained via solving the following optimization:

$$x^* = \text{argmin}_{x' \in \{x, x^{aug}\}} Entropy(\tau(\mathcal{M}(x'))). \tag{6}$$

Namely, we choose an input version out of the two input candidates (x and x^{aug}) according to the entropy (prediction certainty) of the segmentation output. Segmentation output with a lower entropy means that the model is more certain in its prediction, and a higher certainty in prediction often positively correlates to higher segmentation accuracy [21].

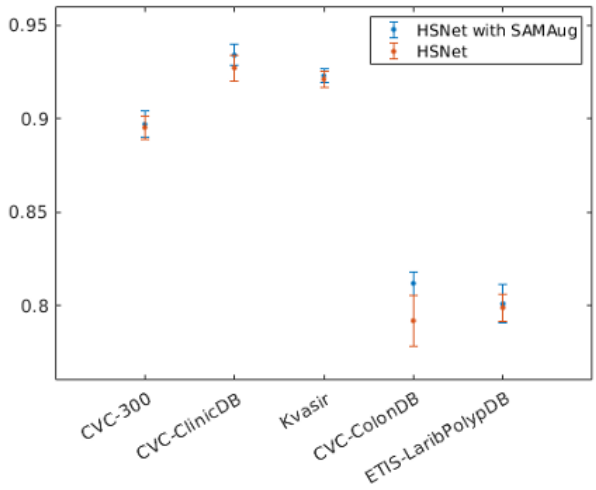

Fig. 3. Polyp segmentation results of the vanilla HSNet and SAMAug-enhanced HSNet.

4 Experiments and Results

4.1 Datasets and Setups

We perform experiments on the Polyp [28], MoNuSeg [11], and GlaS [18] benchmarks to demonstrate the effectiveness of our proposed SAMAug method. For the polyp segmentation experiments, we follow the training setup used in training the state-of-the-art (SOTA) model HSNet [26][6]. For the MoNuSeg and GlaS segmentation, the training of a medical image segmentation model uses the Adam optimizer [9], with batch size = 8, image cropping window size = 256×256, and learning rate = $5e-4$. The total number of training iterations is 50K. The spatial cross entropy loss is used for the model training.

4.2 Polyp Segmentation on Five Datasets

Automatic polyp segmentation in endoscopic images can help improve the efficiency and accuracy in clinical screenings and tests for gastrointestinal diseases. Many deep learning (DL) based models have been proposed for robust and automatic segmentation of polyps. Here, we utilize the SOTA model HSNet [26] for evaluating our proposed SAMAug method. We use the objective function described in Eq. (2) in model training. In test time, we use the model deployment strategy given in Eq. (6). In Fig. 3, we show the segmentation performance (in Dice score) of the vanilla HSNet and SAMAug-enhanced HSNet on the test sets of CVC-300 [20], CVC-ClinicDB [1], Kvasir [7], CVC-ColonDB [19], and

[6] https://github.com/baiboat/HSNet

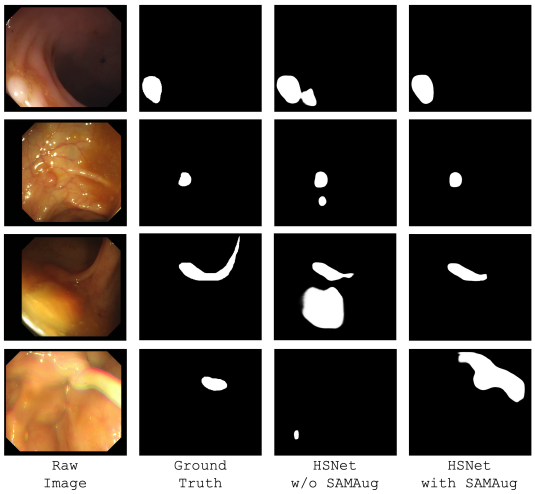

Fig. 4. Visual results of the HSNet and SAMAug-enhanced HSNet in polyp segmentation.

ETIS [17]. All the model training sessions were run ten times with different random seeds for reporting the means and standard deviations of the segmentation performance. In Fig. 3, we observe that SAMAug improves HSNet on the CVC-ClinicDB and CVC-ColonDB datasets significantly, and remains at the same level of performance on the other three datasets (all validated by t-test). Furthermore, we give visual result comparisons in Fig. 4.

4.3 Cell Segmentation on the MoNuSeg Dataset

The MoNuSeg dataset [11] was constructed using H&E stained tissue images (at 40x magnification) from the TCGA archive [23]. The training set consists of 30 images with about 22000 cell nuclear annotations. The test set contains 14 images with about 7000 cell nuclear annotations. We use the objective function described in Eq. (1) in model training. In test time, we use the model deployment strategy given in Eq. (3). In Table 1, we show clear advantages of our proposed method in improving segmentation results for the U-Net, P-Net, and Attention U-Net models. AJI (Aggregated Jaccard Index) is a standard segmentation evaluation metric[7] used on MoNuSeg which evaluates segmentation performance on the object level. F-score evaluates the cell segmentation performance on the pixel level. In addition, we give visual result comparisons in Fig. 5. Note that, although the segmentation generated by SAM (e.g., see the 3rd column of Fig. 5) does not immediately give accurate cell segmentation, SAM provides a general segmentation perceptual prior for the subsequent DL models to generate much more accurate task-specific segmentation results.

[7] https://monuseg.grand-challenge.org/Evaluation/

Table 1. Cell segmentation results on the MoNuSeg dataset.

Model	SAMAug	AJI	F-score
Swin-UNet [2]	✗	61.66	80.57
U-Net [15]	✗	58.36	75.70
	✓	64.30	82.36
P-Net [22]	✗	59.46	77.09
	✓	63.98	82.56
Attention UNet [13]	✗	58.76	75.43
	✓	63.15	81.49

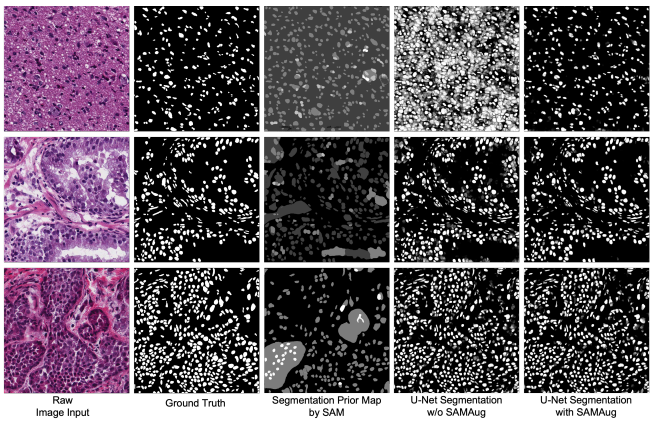

Fig. 5. Visual comparisons of segmentation results on the MoNuSeg dataset.

4.4 Gland Segmentation on the GlaS Dataset

The GlaS dataset [18] has 85 training images (37 benign (BN), 48 malignant (MT)), and 60 test images (33 BN, 27 MT) in part A and 20 test images (4 BN, 16 MT) in part B. We use the official evaluation code[8] for evaluating segmentation performance. For simplicity, we merge test set part A and test set part B, and perform segmentation evaluation at once for all the samples in the test set. We use the objective function described in Eq. (1) in model training. In test time, we use the model deployment strategy given in Eq. (3). From Table 2, one can see that U-Net with SAMAug augmentation performs considerably better than that without SAMAug augmentation.

Acknowledgement. This work was supported in part by National Natural Science Foundation of China (62201263) and Natural Science Foundation of Jiangsu Province (BK20220949). S.W. is supported by Shanghai Sailing Programs of Shanghai Municipal Science and Technology Committee (22YF1409300).

[8] https://warwick.ac.uk/fac/cross_fac/tia/data/glascontest/evaluation/

Table 2. Gland segmentation results on the GlaS dataset.

Model	SAMAug	F-score	Object Dice
U-Net [15]	✗	79.33	86.35
	✓	82.50	87.44

5 Conclusions

In this paper, we proposed a new method, SAMAug, for boosting medical image segmentation that uses the Segment Anything Model (SAM) to augment image input for commonly-used medical image segmentation models. Experiments on three segmentation tasks showed the effectiveness of our proposed method. Future work may consider conducting further research on: (1) designing a more robust and advanced augmentation function; (2) improving the efficiency of applying SAM in the SAMAug scheme; (3) utilizing SAMAug for uncertainty estimations and in other clinically-oriented applications.

References

1. Jorge Bernal, F Javier Sánchez, Gloria Fernández-Esparrach, Debora Gil, Cristina Rodríguez, and Fernando Vilariño. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics*, 43:99–111, 2015.
2. Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-Unet: Unet-like pure Transformer for medical image segmentation. In *Computer Vision–ECCV 2022 Workshops, Part III*, pages 205–218. Springer, 2023.
3. Phillip Chlap, Hang Min, Nym Vandenberg, Jason Dowling, Lois Holloway, and Annette Haworth. A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, 65(5):545–563, 2021.
4. Ruining Deng, Can Cui, Quan Liu, Tianyuan Yao, Lucas W Remedios, Shunxing Bao, Bennett A Landman, Lee E Wheless, Lori A Coburn, Keith T Wilson, et al. Segment anything model (SAM) for digital pathology: Assess zero-shot segmentation on whole slide imaging. *arXiv preprint arXiv:2304.04155*, 2023.
5. Phu-Hung Dinh and Nguyen Long Giang. A new medical image enhancement algorithm using adaptive parameters. *International Journal of Imaging Systems and Technology*, 32(6):2198–2218, 2022.
6. 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? *arXiv preprint arXiv:2304.14660*, 2023.
7. Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, and Håvard D Johansen. Kvasir-SEG: A segmented polyp dataset. In *MultiMedia Modeling: 26th International Conference, Part II*, pages 451–462. Springer, 2020.
8. Ge-Peng Ji, Deng-Ping Fan, Peng Xu, Ming-Ming Cheng, Bowen Zhou, and Luc Van Gool. SAM struggles in concealed scenes–empirical study on "segment anything". *arXiv preprint arXiv:2304.06022*, 2023.

9. Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
10. 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. *arXiv preprint arXiv:2304.02643*, 2023.
11. Neeraj Kumar, Ruchika Verma, Sanuj Sharma, Surabhi Bhargava, Abhishek Vahadane, and Amit Sethi. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Transactions on Medical Imaging*, 36(7):1550–1560, 2017.
12. Jun Ma and Bo Wang. Segment anything in medical images. *arXiv preprint arXiv:2304.12306*, 2023.
13. Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention U-Net: Learning where to look for the pancreas. In *International Conference on Medical Imaging with Deep Learning*, 2018.
14. Yu Qiao, Chaoning Zhang, Taegoo Kang, Donghun Kim, Shehbaz Tariq, Chenshuang Zhang, and Choong Seon Hong. Robustness of SAM: Segment anything under corruptions and beyond. *arXiv preprint arXiv:2306.07713*, 2023.
15. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Part III*, pages 234–241. Springer, 2015.
16. Leonardo Rundo, Andrea Tangherloni, Marco S Nobile, Carmelo Militello, Daniela Besozzi, Giancarlo Mauri, and Paolo Cazzaniga. MedGA: A novel evolutionary method for image enhancement in medical imaging systems. *Expert Systems with Applications*, 119:387–399, 2019.
17. Juan Silva, Aymeric Histace, Olivier Romain, Xavier Dray, and Bertrand Granado. Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer. *International Journal of Computer Assisted Radiology and Surgery*, 9:283–293, 2014.
18. Korsuk Sirinukunwattana, Josien PW Pluim, Hao Chen, Xiaojuan Qi, Pheng-Ann Heng, Yun Bo Guo, Li Yang Wang, Bogdan J Matuszewski, Elia Bruni, Urko Sanchez, et al. Gland segmentation in colon histology images: The GlaS challenge contest. *Medical Image Analysis*, 35:489–502, 2017.
19. Nima Tajbakhsh, Suryakanth R Gurudu, and Jianming Liang. Automated polyp detection in colonoscopy videos using shape and context information. *IEEE Transactions on Medical Imaging*, 35(2):630–644, 2015.
20. David Vázquez, Jorge Bernal, F Javier Sánchez, Gloria Fernández-Esparrach, Antonio M López, Adriana Romero, Michal Drozdzal, and Aaron Courville. A benchmark for endoluminal scene segmentation of colonoscopy images. *Journal of Healthcare Engineering*, 2017, 2017.
21. Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations*, 2021.
22. Guotai Wang, Maria A Zuluaga, Wenqi Li, Rosalind Pratt, Premal A Patel, Michael Aertsen, Tom Doel, Anna L David, Jan Deprest, Sébastien Ourselin, et al. DeepIGeoS: A deep interactive geodesic framework for medical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(7):1559–1572, 2018.

23. Zhining Wang, Mark A Jensen, and Jean Claude Zenklusen. A practical guide to the cancer genome atlas (TCGA). *Statistical Genomics: Methods and Protocols*, pages 111–141, 2016.
24. 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.
25. Chaoning Zhang, Sheng Zheng, Chenghao Li, Yu Qiao, Taegoo Kang, Xinru Shan, Chenshuang Zhang, Caiyan Qin, Francois Rameau, Sung-Ho Bae, et al. A survey on Segment Anything Model (SAM): Vision foundation model meets prompt engineering. *arXiv preprint arXiv:2306.06211*, 2023.
26. Wenchao Zhang, Chong Fu, Yu Zheng, Fangyuan Zhang, Yanli Zhao, and Chiu-Wing Sham. HSNet: A hybrid semantic network for polyp segmentation. *Computers in Biology and Medicine*, 150:106173, 2022.
27. Amy Zhao, Guha Balakrishnan, Fredo Durand, John V Guttag, and Adrian V Dalca. Data augmentation using learned transformations for one-shot medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8543–8553, 2019.
28. Tao Zhou, Yizhe Zhang, Yi Zhou, Ye Wu, and Chen Gong. Can SAM segment polyps? *arXiv preprint arXiv:2304.07583*, 2023.