

# SamDSK: Combining Segment Anything Model with Domain-Specific Knowledge for Semi-Supervised Learning in Medical Image Segmentation

Yizhe Zhang, Tao Zhou, Shuo Wang, Ye Wu, Pengfei Gu, Danny Z. Chen

**Abstract**—The Segment Anything Model (SAM) exhibits a capability to segment a wide array of objects in natural images, serving as a versatile perceptual tool for various downstream image segmentation tasks. In contrast, medical image segmentation tasks often rely on domain-specific knowledge (DSK). In this paper, we propose a novel method that combines the segmentation foundation model (i.e., SAM) with domain-specific knowledge for reliable utilization of unlabeled images in building a medical image segmentation model. Our new method is iterative and consists of two main stages: (1) segmentation model training; (2) expanding the labeled set by using the trained segmentation model, an unlabeled set, SAM, and domain-specific knowledge. These two stages are repeated until no more samples are added to the labeled set. A novel optimal-matching-based method is developed for combining the SAM-generated segmentation proposals and pixel-level and image-level DSK for constructing annotations of unlabeled images in the iterative stage (2). In experiments, we demonstrate the effectiveness of our proposed method for breast cancer segmentation in ultrasound images, polyp segmentation in endoscopic images, and skin lesion segmentation in dermoscopic images. Our work initiates a new direction of semi-supervised learning for medical image segmentation: the segmentation foundation model can be harnessed as a valuable tool for label-efficient segmentation learning in medical image segmentation.

**Index Terms**—Medical Image Segmentation, Segment Anything Model, Semi-Supervised Learning, Domain-Specific Knowledge, Optimal Matching, Pseudo-label Selection and Refinement.

## 1 INTRODUCTION

Segmentation foundational models, such as the Segment Anything Model (SAM) [13], have opened up new opportunities in medical image segmentation studies. Many recent endeavors have focused on fine-tuning and adapting SAM for medical image segmentation tasks. In cases where SAM generates high-quality segmentation proposals for medical image segmentation datasets, domain-specific knowledge can be integrated with SAM to annotate unlabeled images. This approach can be beneficial and enhance medical image segmentation, especially in scenarios where plentiful unlabeled images are accessible while manually annotated images are scarce.

Leveraging unlabeled samples in training medical image segmentation models is a semi-supervised learning (SSL) problem, and has been extensively studied in the last decade. Zhang et al. [31] proposed to use adversarial training when using unlabeled samples. Adversarial training encourages a segmentation model to produce outputs for unlabeled samples that are indistinguishable from outputs for labeled samples in segmentation quality. Peiris et al. [17] added multi-view segmentation components to the

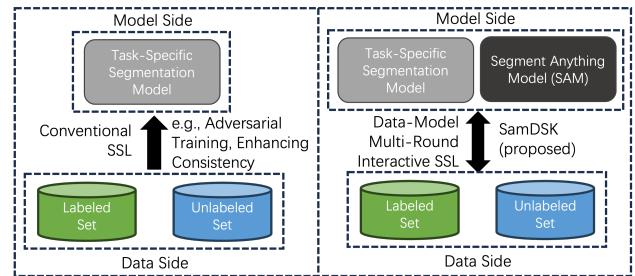


Fig. 1: A high-level comparison between a conventional approach (left) and our proposed SamDSK (right) for semi-supervised learning (SSL) in medical image segmentation.

adversarial training framework. Recently, Wang et al. [25] proposed to incorporate anatomical constraints such as connectivity, convexity, and symmetry in adversarial training for learning from unlabeled images. Luo et al. [15] combined a consistency criterion with adversarial training for semi-supervised learning of medical image segmentation models. Designing losses and network structures that enhance segmentation consistencies is another widely used approach for semi-supervised learning in medical image segmentation [27]. Another track of research for semi-supervised learning is based on pseudo-labels. Introduced in [14], pseudo-label techniques have been widely utilized for semi-supervised learning in medical image segmentation (e.g., [19]).

Adversarial training for semi-supervised learning is typically built on the assumption that the unlabeled samples are drawn from

• Y. Zhang, T. Zhou, and Y. Wu are with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. S. Wang is with the Digital Medical Research Center, School of Basic Medical Sciences, Fudan University, and Shanghai Key Laboratory of MICCAI, Shanghai, China. P. Gu and D. Chen are with the Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, USA.

E-mail: yizhe.zhang.cs@gmail.com, dchen@nd.edu

the same distribution of the labeled samples. The effectiveness of the min-max optimization in improving the segmentation quality depends on factors such as the architecture of the discriminator, the initial model parameters, the distribution changes (if any) between the unlabeled and labeled samples. Contrastive learning has shown great performance in representation learning, but, its effectiveness depends on the design of pre-tasks and the training process. Similarly, consistency loss requires identifying correct notation of consistency between unlabeled images and multiple views of the images. Incorrect or inappropriate design of the contrastive loss and consistency loss could make the learning ineffective and sometimes even harmful to the segmentation performance. Pseudo-label (PL) based techniques are often a go-to approach for utilizing unlabeled images in training a prediction model (e.g., classification model, segmentation model). Despite its effectiveness, if a large portion of the pseudo-labels is incorrect, then using PLs to train the model would suffer the risk of lower performance.

Previous work of semi-supervised learning commonly concentrated on network training and architecture design, but fewer attempts were made on constructing and expanding the (labeled) training set. In this paper, we cast the task of utilizing unlabeled images in training a medical image segmentation model as an iterative and gradual process of annotation acquisition of unlabeled images together with model training (and re-training). Both the training set and segmentation model evolve during the learning process. We develop a novel method that combines the emerging SAM with Domain-Specific Knowledge (DSK) for annotation acquisition and labeled set expansion. More specifically, suppose a labeled image set is initially given to our method. We train an initial medical image segmentation model using the labeled images. We then apply the trained model to the unlabeled images and obtain segmentation probability maps for all the unlabeled images. We refer to these maps as pixel-level domain-specific knowledge (pixel-level DSK). We further introduce a high-level DSK called image-level DSK, which includes prior knowledge of the segmentation task, such as the potential number of RoIs (Regions of Interests) for each segmentation class in an image. For an unlabeled image, an optimization process is performed to construct its segmentation annotation by matching the segmentation proposals (generated by SAM) with the pixel-level DSK, constrained by the image-level DSK. Unlabeled images with segmentation annotations attained by the above matching process are then examined according to the optimal matching scores, and those images (with the constructed annotations) with high matching scores will be added to the labeled set as machine-labeled samples for the next round of segmentation model training. The segmentation model receives a further improvement in each round of training since more labeled data become available, and through multiple rounds of such annotation acquisition and model training/retraining, we expect more unlabeled samples with high-quality segmentation to be added to the labeled set, and thus a better medical image segmentation model will be obtained. A high-level overview of our proposed method is illustrated in Fig. 1.

In summary, the main contributions of this work are four-fold:

- We propose a schematic shift of utilizing unlabeled images in training a medical image segmentation model. Our proposition is that once unlabeled data can be “transformed” into labeled data, the training of the segmentation model becomes similar to the model that is trained under

the fully supervised learning setup.

- We develop a novel optimization-based method that combines the segmentation foundation model (SAM) with domain-specific knowledge (DSK) for semi-supervised learning in medical image segmentation. Our approach integrates the strengths of SAM and pseudo-label based learning, enhancing the controllability and reliability of the process of utilizing unlabeled images to train a segmentation model.
- Given that many medical image segmentation tasks involve and benefit from image-level domain-specific knowledge, we extend our method by developing an image-level DSK within the optimization process. More specifically, we directly apply and utilize the count of regions of interest in our proposed method, aiming to regularize the process of constructing annotations for unlabeled images.
- Analyses and experiments confirm the effectiveness of our proposed method. We also illustrate and discuss current limitations and future research directions for utilizing SAM to leverage unlabeled images in medical image segmentation.

## 2 RELATED WORK

### 2.1 Pseudo-labels

Our proposed SamDSK uses probability maps generated by a medical image segmentation network for unlabeled images. Although probability maps are not exactly taken as labels or pseudo-labels, these maps are often referred to as “soft” pseudo-labels. Hence, below we give a review of the methods that use pseudo-labels for semi-supervised learning in medical image segmentation.

Wu et al. [27] proposed Mutual Consistency Learning (MCL) for utilizing unlabeled samples in segmentation model training. MCL creates multiple decoders which all share one encoder for generating segmentations. Because of the random initialization, these decoders generate possibly different segmentation predictions, and the differences between these predictions are used to compute prediction uncertainties. In addition, mutual learning [30] is performed during model training to enforce the output from each decoder to be close (consistent) to the pseudo-labels generated by the other decoders. At the end of the training, only one decoder is used for model deployment. Since pseudo-labels inherently contain errors, Thompson et al. [23] proposed to use super-pixels to refine pseudo-labels in semi-supervised learning. The proposed process largely follows the classical approach of using pseudo-labels for SSL [14], with additional steps which refine pseudo-labels according to super-pixels. The main motivation of this method is to impose a spatial structure induced from the super-pixels for regularizing and refining pseudo-labels so that pseudo-labels would yield better quality and benefit the subsequent segmentation model training. Seibold et al. [20] proposed to use reference images (with annotations) for generating pseudo-labels. This approach works for medical image segmentation tasks with well-defined and stable object structures (e.g., chest radiographic anatomy segmentation). For tasks under more dynamic scenes where objects can be at arbitrary locations and with large variations of shapes (e.g., polyp segmentation in endoscopic images), reference-based pseudo-label generation could become less reliable and effective. Recently, Basak et al. [4] proposed to utilize pseudo-labels in conjunction with contrastive learning.

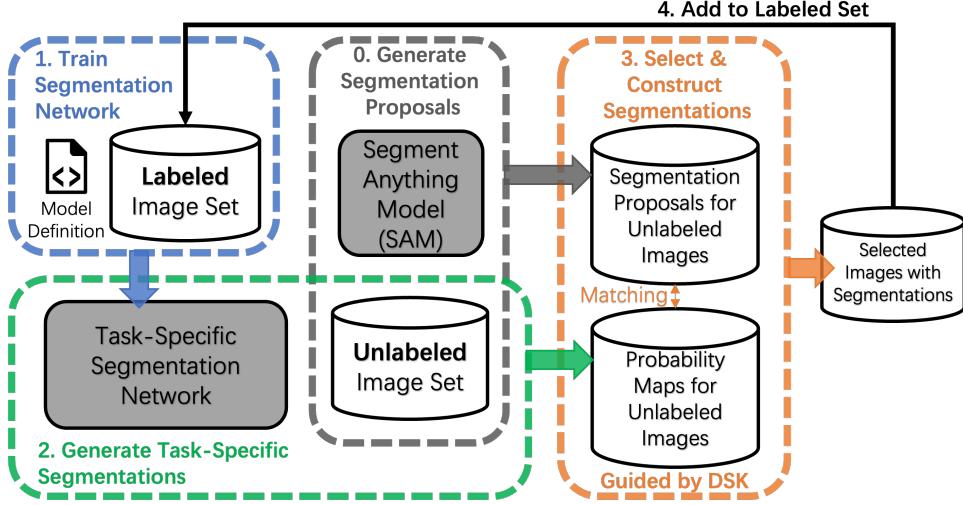


Fig. 2: The main steps of our proposed SamDSK method. Step-0 to step-4 are repeated iteratively until a certain criterion is met (e.g., no more samples are added to the labeled set). Step-0 may need to run only once if SAM remains unchanged during the process.

Our proposed method utilizes probability maps generated by a task-specific medical image segmentation model. Similar to [23], we employ an external component (i.e., SAM in our case) to work with the medical image segmentation model. Instead of refining pseudo-labels using super-pixels, we aim to use DSK to match and select SAM-generated segmentation proposals and construct segmentations for unlabeled images for model re-training. Comparing to the super-pixel based methods, SAM is a much more advanced model and using SAM allows us to design a more controllable optimization process (e.g., multiple rounds of optimal matching) in the usage of domain-specific knowledge both at the pixel level and image level for generating segmentations of unlabeled images.

## 2.2 SAM for Medical Image Segmentation

Since the introduction of the Segment Anything Model (SAM), there has been a wave of attempts to develop a more accurate and robust medical image segmentation system with the utilization of SAM [11, 16, 18, 26, 28, 32]. Recent work has shown that SAM alone, without further fine-tuning and/or adaptation, often delivers unsatisfactory results for medical image segmentation tasks [11, 32]. To utilize SAM more effectively, Ma et al. [16] proposed to fine-tune SAM using labeled images. Wu et al. [26] proposed to add additional layers to adapt SAM for a medical image segmentation task. Our work aims at using SAM for semi-supervised learning of a medical image segmentation model.

## 2.3 SAM and Pseudo-labels

Recently, Chen et al. [7] proposed to use SAM to refine pseudo-labels in weakly-supervised learning. SAM is employed in three places: (1) refining CAM-generated attention maps, (2) refining segmentation generated by post-processing, and (3) refining segmentations generated by a segmentation model (DeepLab). In this paper, we utilize probability maps as pixel-level DSK for selecting segmentation proposals produced by SAM. To make the selection more reliable, we further employ an image-level DSK (e.g., the potential number of regions of interests) during the selection process to constrain the optimization process as a way of realizing the idea of combining SAM with domain-specific knowledge (at both the pixel level and image level).

## 3 METHODOLOGY

The input to our SamDSK model consists of four components: a labeled image set  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ , an unlabeled image set  $U = \{u_1, u_2, \dots, u_m\}$ , Segment Anything Model SAM, and a medical image segmentation model  $M$  with an encoder initialized using ImageNet [9] pre-trained weights and a decoder initialized randomly. SamDSK performs five main steps (step-0 to step-4). An overview of the five steps and their relations are illustrated in Fig. 2. In Sections 3.2, 3.3, and 3.4, we describe the key step (step-3) for generating annotations of unlabeled images. In Section 3.5, we present the process of segmentation model training using both human-labeled and machine-labeled images. In Section 3.6, we analyze the dynamics of our proposed multi-round SSL method. Finally, in Section 3.7, we give discussions of several medical image segmentation tasks that SamDSK is applicable to.

### 3.1 Applying SAM to Unlabeled Medical Images

Given a set  $U$  of unlabeled medical images, we apply the state-of-the-art SAM to these images to obtain segmentation proposals in each image. A key consideration in this step is to use relatively lower threshold settings in SAM to ensure the inclusion of all potential Regions of Interest (RoIs) in the generated segmentation proposals. By default, we use the following parameter values when employing SAM: “`crop_nms_threshold = 0.5, box_nms_thresh = 0.5, pred_iou_thresh = 0.5, stability_score_thresh = 0.5`”. With the ongoing development of the segmentation foundation model, future SAM variants are expected to provide improved coverage of RoIs for various medical image segmentation tasks.

### 3.2 Matching Segmentation Proposals with Pixel-level DSK

Suppose we already train a segmentation model  $M$  using the currently available labeled data. Also, suppose for an unlabeled image  $u$  with width  $w$  and height  $h$ , SAM provides a set of  $K$  segmentation proposals,  $S = \{s_1, s_2, \dots, s_K\}$ , where  $s_i \in \{0, 1\}^{h \times w}$  for  $i = 1, 2, \dots, K$ . The segmentation model  $M$  gives  $C$  distinct probability maps  $\{p_1, p_2, \dots, p_C\}$ , where  $p_c \in \mathbb{R}^{h \times w}$  for  $c = 1, 2, \dots, C$ .  $C$  is the number of classes

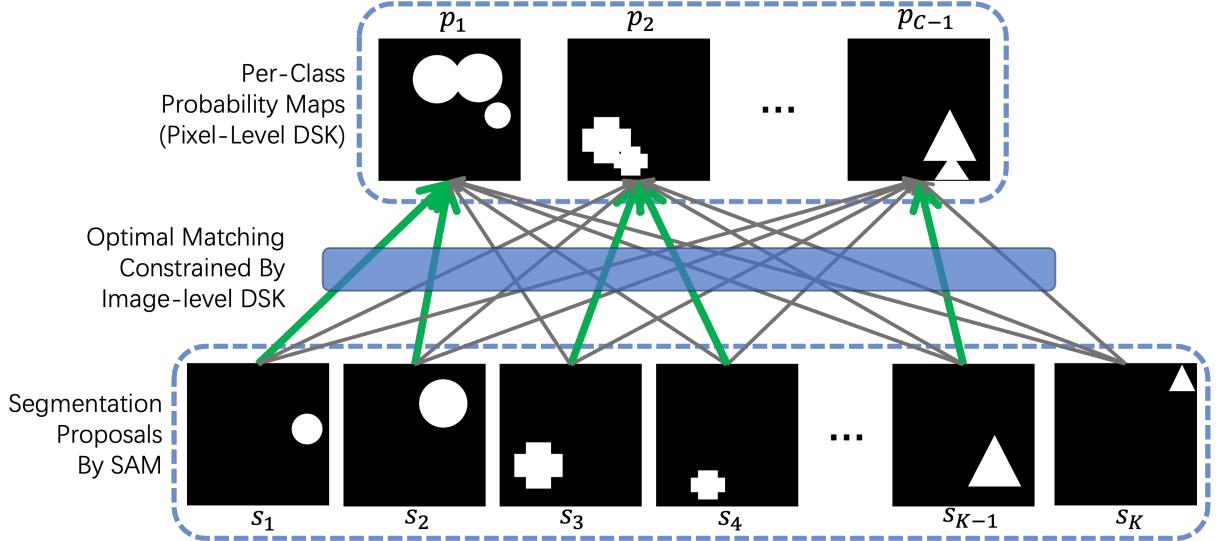


Fig. 3: Matching segmentation proposals generated by SAM with pixel-level DSK generated by the current trained segmentation model. The matching is constrained by additional image-level DSK, i.e., potential counts of RoIs. Green arrows mark the obtained solution.

in the segmentation task. The background class is taken always as the class  $C$ . For each class  $c \in \{1, 2, \dots, C\}$ , we create a set of binary scalar indicators,  $Z_c = \{z_{c,1}, z_{c,2}, \dots, z_{c,K}\}$ , where  $z_{c,k} \in \{0, 1\}$  for  $k = 1, 2, \dots, K$ . Each indicator  $z_{c,k}$  is multiplied with the corresponding segmentation proposal  $s_k$  in constructing the segmentation map for class  $c$ . We aim to find an optimal configuration of  $z_{1,1}, z_{1,2}, \dots, z_{C,K}$  which maximizes the total sum of IoU scores between the segmentation proposals and the segmentation probability maps, as:

$$z_{1,1}^*, \dots, z_{C,K}^* = \underset{z_{1,1}, z_{1,2}, \dots, z_{C,K}}{\operatorname{argmax}} \sum_{c=1}^{C-1} \text{IoU}(p_c, \tau(\sum_{k=1}^K z_{c,k} s_k)), \quad (1)$$

$$\text{s.t., } \sum_{c=1}^{C-1} z_{c,k} \leq 1, \quad (2)$$

$$z_{c,k} \in \{0, 1\}, c = 1, 2, \dots, C, k = 1, 2, \dots, K,$$

where  $\tau(\cdot)$  converts all values which are larger than 1 to 1, and  $\text{IoU}(\cdot, \cdot)$  computes the Intersection over Union score between a probability map and a segmentation proposal. The constraints in Eq. (2) ensure that each segmentation proposal can be assigned to only one class label. Note that there can be segmentation proposals not being selected in forming the optimal matching.

In Fig. 3, we provide a high-level illustration of the mechanism that we use for this optimal matching formulation. The optimization problem in Eq. (1) (with the constraints in Eq. (2)) is a maximum bipartite matching problem which can be efficiently solved in polynomial time [3].

### 3.3 Incorporating Image-level DSK

In this section, we aim to improve the effectiveness and reliability of the above proposed segmentation annotation construction by adding image-level DSK as new constraints to the optimal matching formulation. Suppose the number of RoIs for an image of a given medical image segmentation task ranges from  $v_c^{\text{lower}}$  to  $v_c^{\text{upper}}$  for class  $c$ . We add these constraints to Eq. (2). The optimization objective remains unchanged, but the constraints are updated with additional constraints imposed by  $v_c^{\text{lower}}$  and

$v_c^{\text{upper}}$ . Formally, we aim to optimize the objective in Eq. (1), with the following constraints:

$$\begin{aligned} & \text{s.t., } \sum_{c=1}^{C-1} z_{c,k} \leq 1, \\ & v_c^{\text{lower}} \leq \sum_{k=1}^K z_{c,k} \leq v_c^{\text{upper}}, c \in \{1, 2, \dots, C-1\}, \\ & z_{c,k} \in \{0, 1\}, c = 1, 2, \dots, C, k = 1, 2, \dots, K. \end{aligned} \quad (3)$$

In each round of the model training and labeled set expansion, we control the  $v_c^{\text{lower}}$  and  $v_c^{\text{upper}}$  values to ensure a stable way of gradually including more samples with more RoIs into the labeled set. For the first round, we can set  $v_c^{\text{lower}}$  and  $v_c^{\text{upper}}$  as both equal to 1. This indicates that the optimal matching is only a one-to-one matching between the segmentation proposals and the probability maps. In the second round, we can set  $v_c^{\text{lower}}$  as 1 and  $v_c^{\text{upper}}$  as 2 to allow a two-to-one matching. Depending on the specific segmentation task, we can use a larger increment of  $v_c^{\text{upper}}$  to allow a more effective inclusion of unlabeled samples into the labeled set.

### 3.4 Segmentation Annotation Construction

With the identified values  $z_{1,1}^*, z_{1,2}^*, \dots, z_{C,K}^*$  through optimizing Eq. (1) and Eq. (3), we compute the overall score  $\beta$  for the image, as follows:

$$\beta = \frac{1}{C-1} \sum_{c=1}^{C-1} \text{IoU}(p_c, \tau(\sum_{k=1}^K z_{c,k}^* s_k)). \quad (4)$$

A sample with a  $\beta$  score higher than a pre-defined threshold (denoted as  $\beta^*$ ; its default value is 0.9) is selected to advance to the next round of segmentation model training. For a selected sample, its annotation map is constructed as  $q_c = \tau(\sum_{k=1}^K z_{c,k}^* s_k)$ , for  $c = 1, 2, \dots, C-1$ . Pixel areas not covered by any segmentation proposals selected by the optimal matching are considered as belonging to the background class (class  $C$ ).

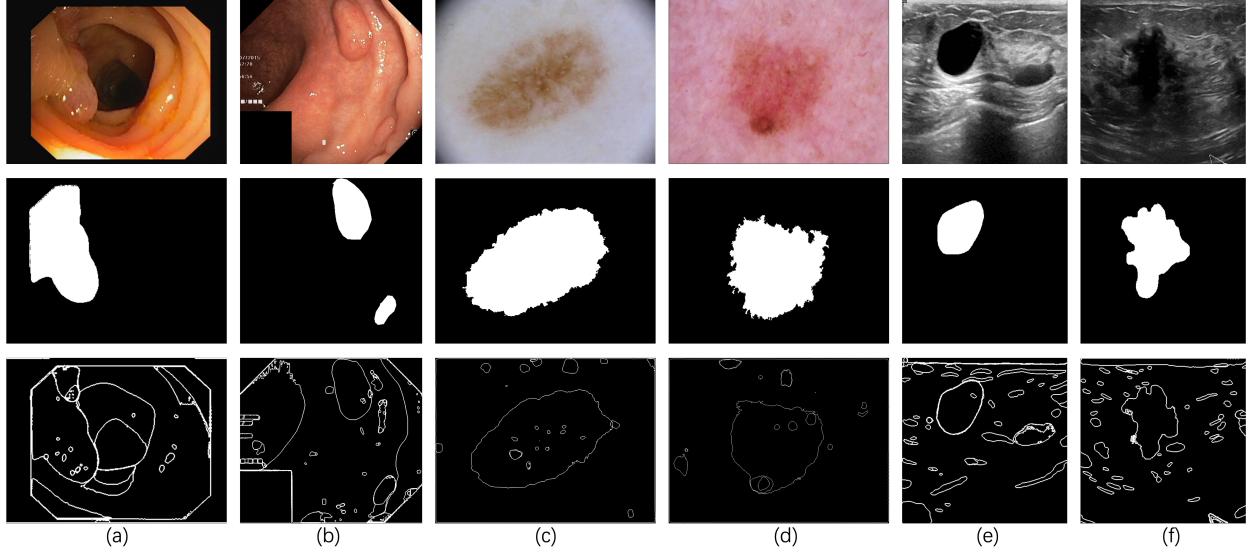


Fig. 4: Image examples of polyp segmentation ((a) and (b)), skin lesion segmentation ((c) and (d)), and breast cancer segmentation ((e) and (f)). Row 1: raw images. Row 2: ground truth annotations. Row 3: Boundary maps of the SAM-generated segmentation proposals.

### 3.5 Segmentation Model Training

Suppose the unlabeled samples with their annotation maps generated by the above optimization procedure are denoted as  $D^{machine} = \{(u_j, q_{j,c})\}$ , for  $j = 1, 2, \dots, m'$ ,  $m' \leq m$ , and  $c = 1, 2, \dots, C$ . The original human-labeled set is denoted as  $D = \{(x_i, y_{i,c})\}$ , for  $i = 1, 2, \dots, n$ ,  $c = 1, 2, \dots, C$ . We aim to optimize the following objective function with respect to the parameters  $\theta$  of a medical image segmentation model  $M$ , as:

$$\operatorname{argmin}_{\theta} \left( \sum_{c=1}^C \sum_{i=1}^n L(M(x_i, y_{i,c})) + \lambda \sum_{c=1}^C \sum_{j=1}^{m'} L(M(u_j, q_{j,c})) \right), \quad (5)$$

where  $\lambda$  is set as 1 by default. Assuming that the medical image segmentation model  $M$  generates probability maps for each class, the loss function  $L$  can take the form of a region-based loss (e.g., Dice loss), a pixel-level entropy-based loss, or a combination of these two types of losses. A mini-batch based stochastic gradient descent method (e.g., Adam) can be applied to optimize the loss via updating the parameters  $\theta$  of the model  $M$ .

### 3.6 Analyses

The SamDSK method combines SAM-generated segmentation proposals with domain-specific knowledge both at the pixel-level and image-level in order to generate annotations for unlabeled images. Pixel-level DSK is provided by the segmentation model  $M$  trained on the current labeled image set, and optimal matching is performed with constraints imposed by image-level DSK as well as the base constraints that only one segmentation proposal can be assigned to one class label. There exist clearly two cases after the optimal matching process for constructing annotations of the unlabeled images.

- Case-1: Prediction maps produced by the current segmentation model  $M$  match well with a subset of the segmentation proposals (generated by SAM). In this scenario, the optimal matching gives a high  $\beta$  matching score (i.e., higher than  $\beta^*$ ), and this image with the constructed

annotations is added to the labeled image set for the subsequent rounds of segmentation model retraining.

- Case-2: Prediction maps produced by the current segmentation model  $M$  do not match well with any subsets of the segmentation proposals generated by SAM. In this situation, optimal matching gives a low  $\beta$  matching score, and this image is not added to the labeled set for the next round of model retraining.

For simplicity, we use a binary segmentation task to illustrate, and focus on its segmentation class 1 (the foreground class) to provide the analyses below. The same logic applies to multi-class segmentation tasks with multiple classes of foreground objects.

**Assumption 1:** For an unlabeled image  $u$ , there exists a subset of segmentation proposals,  $\Gamma_1$ , generated by SAM, such that the union of its elements closely approximates the ground truth of segmentation class 1 in  $u$ .<sup>1</sup> More formally, this assumption can be described as:

$$\text{IoU}\left(\bigcup_{s \in \Gamma_1} s, gt_1\right) > 1 - \epsilon, \quad (6)$$

where  $gt_1$  represents the ground truth annotation map of class 1 for the image  $u$ , and  $\epsilon$  is a small positive value (e.g.,  $\epsilon = 0.02$ ). It is important to note that we do not have access to the ground truth annotations of the unlabeled images.

**Proposition 1:** If an unlabeled image falls into the Case-2 category described above, then its probability maps generated by the segmentation model  $M$  are not closely aligned with their corresponding ground truth annotation maps.

**Proof:** For an unlabeled image in Case-2, the optimal matching identifies no subset of the segmentation proposals that yields a sufficiently high Intersection over Union (IoU) score (e.g., a score higher than 0.9) for the probability map of class 1 (i.e.,  $p_1$ ) generated by the segmentation model  $M$ . Given that  $\Gamma_1$  is a feasible solution attained when seeking an optimal matching, it

1. SamDSK is not yet ready to be applied to those medical image segmentation tasks for which SAM fails to generate sufficiently good segmentation proposals.

TABLE 1: Breast cancer segmentation in ultrasound images (in Dice coefficient). The best results are marked in **bold**, and the second best results are underlined. The red color highlights performance degradation compared to the baseline. The same markings are used in the other tables.

Method	Network	Labeled	Unlabeled	Dice
Baseline	TransUNet	30%	0%	52.7
PL [14]			70%	<u>50.8</u>
SP-Refine [23]				53.2
BCP [2]				53.7
UPS [19]				<u>54.3</u>
SamDSK (ours)				<b>59.4</b>
Baseline			0%	64.3
PL [14]	HSNet	30%	70%	69.5
SP-Refine [23]				69.7
BCP [2]				<u>64.1</u>
UPS [19]				70.8
SamDSK (ours)				<b>73.6</b>

follows that the union of elements in  $\Gamma_1$  is not well-matched with the probability map  $p_1$ . This can be formally expressed as:

$$\text{IoU}(\bigcup_{s \in \Gamma_1} s, p_1) < \beta^*. \quad (7)$$

Since the union of elements in  $\Gamma_1$  closely approximates the ground truth and  $p_1$  is not well aligned with the union of elements in  $\Gamma_1$ , it follows that  $p_1$  is NOT closely aligned with the ground truth annotation map  $gt_1$ . Formally, an upper bound of the Intersection over Union (IoU) between the ground truth annotation map and the prediction map of class 1 is given by:

$$\text{IoU}(gt_1, p_1) \lesssim \beta^* + \epsilon. \quad (8)$$

With the above argument, we have demonstrated that the unlabeled samples which are not added to the labeled set (Case-2) do not yet contain sufficiently accurate annotations (e.g.,  $p_1$  is not sufficiently accurate with respect to the ground truth). On the other hand, for the unlabeled images with annotations that are added to the labeled set (Case-1), it is still possible that although their annotations match well with a subset of the segmentation proposals, denoted as  $\Gamma_1$ , the ground truth annotation map is actually closer to another subset  $\Gamma'_1$  of the segmentation proposals, and  $\Gamma_1$  may or may not be equal to  $\Gamma'_1$ . Consequently, we cannot provide an explicit guarantee on the correctness of  $p_1$  and its corresponding  $\bigcup_{s \in \Gamma_1} s$  for the samples in Case-1. Nevertheless, in our experiments, we present empirical evidence to demonstrate that the annotations of samples in Case-1 (those added to the labeled set) are closer to their corresponding ground truth annotations than those in Case-2.

### 3.7 Applicable Segmentation Tasks

Medical image segmentation encompasses a multitude of imaging modalities (e.g., MRI, CT, microscopy), covering a wide range of diverse objects of interest and clinical tasks. SamDSK relies on SAM to provide segmentation proposals. Therefore, SamDSK is especially applicable to those medical image segmentation tasks in which the regions of interest can be adequately addressed by segmentation proposals generated by SAM. It is important to note that our optimal matching formulation allows for the utilization of multiple segmentation proposals to cover one or more regions of interest. Consequently, SamDSK accommodates segmentation

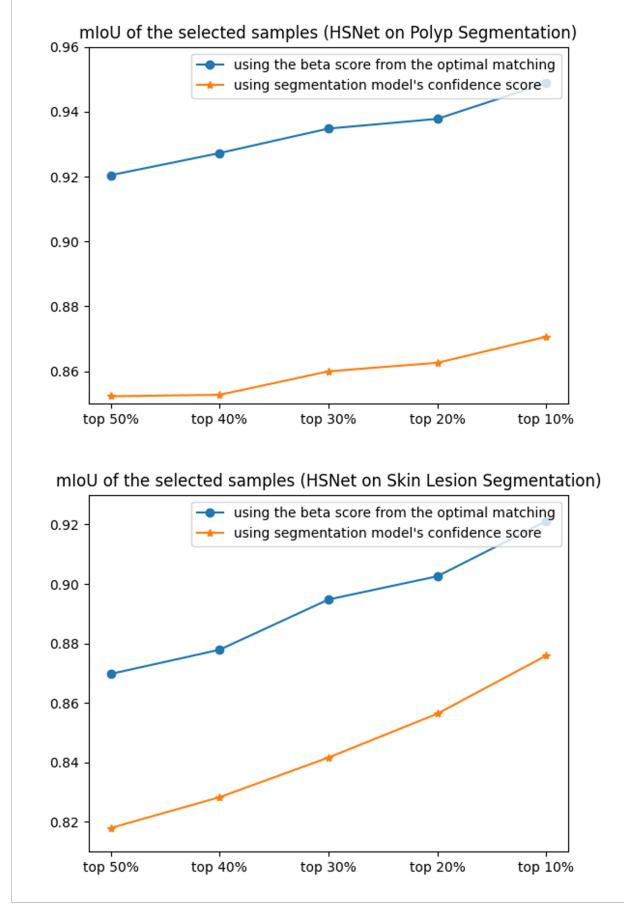


Fig. 5: Mean IoU (mIoU) of segmentation predictions and ground truth annotations for selected samples. The  $\beta$  score (in Eq. (4)) enables SamDSK to select samples with higher segmentation quality (higher mIoU) compared to samples selected by using the model’s own uncertainty and confidence measures (as conducted in UPS [19]).

proposals that consist of over-segmented regions with respect to the regions of interest.

For illustration, we highlight several specific medical image segmentation tasks for which SamDSK is well-suited: (1) polyp segmentation in endoscopic images, (2) skin lesion segmentation in dermoscopic images, and (3) tumor region segmentation in ultrasound images. Fig. 4 presents visual examples of these tasks and the corresponding segmentation proposals generated by SAM. We conduct comprehensive experiments to demonstrate the effectiveness of SamDSK in tackling these segmentation tasks, as shown in the next section.

## 4 EXPERIMENTS

We utilize three public datasets to demonstrate the effectiveness of our proposed SamDSK method. We compare SamDSK with several classical methods as well as state-of-the-art methods: (1) Pseudo-labels (PL) [14]; (2) SP-Refine [23]: utilizing superpixels for refining pseudo-annotations; (3) Bidirectional Copy-Paste (BCP) [2]; (4) Uncertainty-aware Pseudo-label Selection (UPS) [19]: a multi-round SSL method based on uncertainty-aware pseudo-label selection.

TABLE 2: Segmentation performances of different methods on five polyp segmentation datasets (in Dice coefficient).

Method	Network	Labeled	Unlabeled	CVC-300	CVC-ClinicDB	Kvasir	CVC-ColonDB	ETIS
Baseline	PraNet	10%	0%	82.3	76.2	83.1	63.7	60.2
PL [14]			90%	83.8	77.7	85.1	63.4	63.3
BCP [2]			90%	83.4	78.3	85.3	65.6	64.3
UPS [19]			90%	84.3	81.7	86.6	62.3	58.2
SamDSK (ours)			90%	<b>89.4</b>	<b>84.4</b>	<b>88.0</b>	<b>67.6</b>	63.5
Baseline			0%	85.1	81.7	86.9	66.4	68.9
PL [14]	HSNet	10%	90%	87.5	84.3	89.6	71.5	72.9
BCP [2]			90%	85.8	84.9	88.1	68.5	68.3
UPS [19]			90%	<b>88.3</b>	<b>86.3</b>	<b>89.8</b>	72.9	<b>76.7</b>
SamDSK (ours)			90%	<b>88.3</b>	<b>85.2</b>	<b>90.2</b>	<b>76.6</b>	72.7
Baseline	PraNet	30%	0%	89.7	83.2	86.5	63.2	64.7
PL [14]			70%	<b>90.8</b>	<b>86.7</b>	<b>88.3</b>	68.5	<b>61.1</b>
BCP [2]			70%	<b>88.3</b>	84.3	<b>89.1</b>	69.0	<b>69.7</b>
UPS [19]			70%	<b>87.8</b>	84.2	87.1	64.8	<b>58.3</b>
SamDSK (ours)			70%	<b>91.7</b>	85.2	87.6	<b>71.4</b>	67.4
Baseline	HSNet	30%	0%	87.1	86.5	90.9	74.9	76.4
PL [14]			70%	87.8	88.9	91.2	77.3	76.9
BCP [2]			70%	88.3	88.4	90.9	75.5	<b>74.0</b>
UPS [19]			70%	87.6	<b>89.4</b>	<b>91.8</b>	77.8	<b>78.5</b>
SamDSK (ours)			70%	<b>88.9</b>	<b>89.6</b>	<b>92.0</b>	<b>79.8</b>	77.5

TABLE 3: Skin lesion segmentation on the ISIC 2018 dataset (in Dice coefficient).

Method	Network	Labeled	Unlabeled	Dice
Baseline	PraNet	10%	0%	86.1
PL [14]			90%	87.3
SP-Refine [23]			90%	87.2
BCP [2]			90%	87.5
SamDSK (ours)			90%	<b>88.1</b>
Baseline	HSNet	10%	0%	86.5
PL [14]			90%	87.7
SP-Refine [23]			90%	88.2
BCP [2]			90%	88.4
SamDSK (ours)			90%	<b>89.9</b>

In addition to comparing with the known methods, we conduct ablation and additional studies to validate the effectiveness of the key components in SamDSK. In Section 4.4.1, we validate the effectiveness of utilizing SAM for sample selection. In Section 4.4.2, we validate the effectiveness of multi-round SSL with SAM. In Section 4.4.3, we demonstrate empirical coverage of the SAM-generated segmentation proposals with respect to ground truth annotations.

#### 4.1 Breast Cancer Segmentation in Ultrasound Images

We first utilize the breast cancer segmentation task in ultrasound images [1] to validate the effectiveness of our proposed SamDSK. This task is a binary segmentation problem of segmenting breast cancer regions in ultrasound images. We randomly split the original dataset into two equally-sized sets, one set for model training and the other set for testing. In the training set, we randomly select 30% of samples and treat them as labeled images, and treat the remaining 70% of samples as unlabeled images.<sup>2</sup> Two state-of-the-art medical image segmentation models, TransUNet [6] and HSNet [29], are taken as the segmentation model  $M$  for the experiments. We perform SamDSK for three rounds of processing (step-0 to step-4 as illustrated in Fig. 2 are one round of processing). For Eq. (3), both  $v_1^{lower}$  and  $v_1^{upper}$  are set as 1. In Table 1, the results show that SamDSK outperforms the state-of-the-art

2. All the data split information will be published together with the code release.

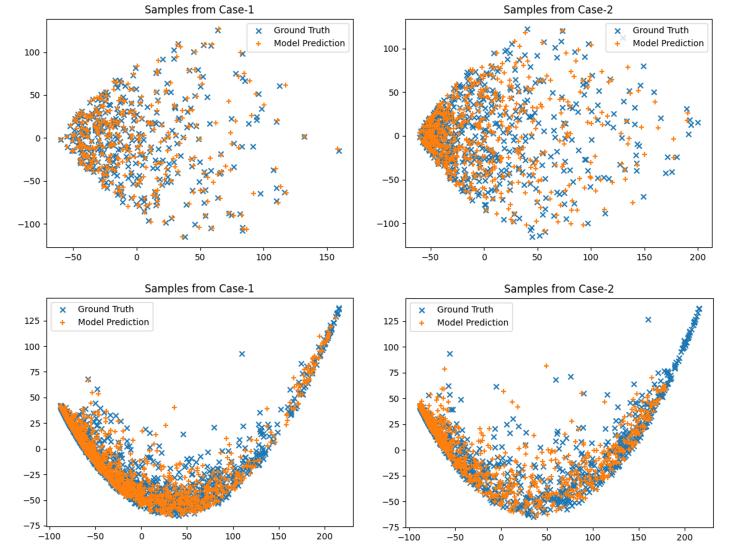


Fig. 6: Ground truth annotations and segmentation predictions after dimensionality reduction (PCA). The segmentation predictions for samples in Case-1 are closer to the ground truth annotations than those in Case-2. Top row: samples of the polyp segmentation dataset. Bottom row: samples of the skin lesion segmentation dataset.

methods for SSL. SamDSK yields considerably better segmentation performance than the closely related BPC method, indicating that SAM plays a critical role in sample selection and annotation construction of unlabeled images. In addition, the results show that the PL method may degrade the segmentation performance (see the TransUNet case) when the initial segmentation model incurs too many errors in pseudo-labels.

#### 4.2 Polyp Segmentation in Endoscopic Images

Automatic polyp segmentation in endoscopic images can help improve the efficiency and accuracy of clinical screenings and tests for gastrointestinal diseases. Many deep learning (DL) based

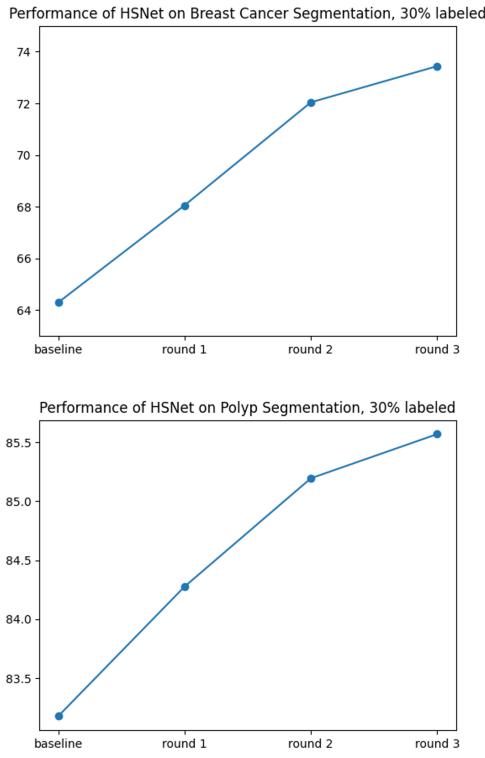


Fig. 7: Segmentation performances (in Dice coefficient) across different rounds of processing in SamDSK.

methods have been proposed for robust and automatic segmentation of polyps. Here, we utilize the SOTA polyp segmentation model HSNet [29] and the widely-used polyp segmentation model PraNet [10] for evaluating our proposed SamDSK method. HSNet uses the PVT backbone (Transformer-based) and PraNet uses the Res2Net backbone (CNN-based). We perform SamDSK for three rounds of processing. For Eq. (3),  $v_1^{lower}$  is set as 1, 1, and 1 for the first, second, and third rounds, respectively, and  $v_1^{upper}$  is set as 1, 2, and 3 for the first, second, and third rounds, respectively (the polyp class is set as class 1). That is, the constraints on the number of RoIs for the polyp class are gradually changed when we perform more rounds of model training and labeled set expansion, and we allow an increase in the number of segmentation proposals for the foreground-class segmentation. Following the data split settings in [10, 32], 900 images from Kvasir [12] and 550 images from CVC-ClinicDB [5] are randomly selected to form the training set. The remaining samples from these two datasets (i.e., Kvasir and CVC-ClinicDB) and the samples from CVC-ColonDB [22], ETIS [21], and CVC-300 [24] form the test set.

The results are reported in Table 2, which show that SamDSK improves the segmentation performances of the HSNet and PraNet models for the settings where 10% and 30% of the total labeled samples from the training set are available. Compared to the classical pseudo-label method [14] and the recently proposed methods, SamDSK achieves similar or better segmentation performances. Notably, we observe that SamDSK is stable and reliable in providing segmentation improvement, while some of the known methods can encounter situations where worse segmentation performances might be yielded in some cases.

### 4.3 Skin Lesion Segmentation in Dermoscopic Images

The ISIC 2018 skin lesion segmentation dataset [8] contains 2594 training dermoscopic images and 1000 test images for melanoma segmentation. Once again, we employ HSNet and PraNet for the experiments. We are aware that the image-level DSK for skin lesion segmentation in most of these images is that they contain only one region of interest. Consequently, we set  $v_1^{lower}$  to 1 and  $v_1^{upper}$  to 1 in Eq. (3) (the melanoma class is designated as class 1) throughout all the rounds of processing. The results in Table 3 demonstrate that SamDSK enhances the segmentation performances of the state-of-the-art models when utilizing 10% of the total labeled samples. SamDSK achieves a higher performance gain in comparison to the competing methods.

### 4.4 Ablation and Additional Studies

#### 4.4.1 Effectiveness of Using SAM

A key question that we seek to answer in this section is whether using the  $\beta$  score obtained from optimal matching actually helps in selecting samples of better segmentation quality. Furthermore, we examine whether using the model’s own segmentation map uncertainty and confidence serves as a better criterion for sample selection. In Fig. 5, we show that by using the  $\beta$  score obtained from optimal matching, we can select samples with significantly better segmentation quality (higher mIoU) than using the model’s own segmentation map uncertainty and confidence measures. For polyp segmentation, samples with the top 10% of  $\beta$  scores achieve a 95% mIoU. For skin lesion segmentation, samples with the top 10% of  $\beta$  scores achieve a 92% mIoU. Additionally, in Fig. 6, we provide visualization plots for ground truth annotations and segmentation predictions for samples in Case-1 and Case-2, which are identified after the optimal matching process. We observe that the samples in Case-1 (selected and added to the labeled set) exhibit a tighter alignment with their corresponding ground truth annotations than the samples in Case-2 (not yet added to the labeled set).

#### 4.4.2 Effectiveness of Multi-round SSL

In Fig. 7, we show the segmentation performances with respect to the rounds of processing in SamDSK, along with that of the initial model (the baseline model). It is evident that the model’s performance is improved through multiple rounds of processing, as more samples are added to the labeled set for model training/retraining.

#### 4.4.3 Performances of SAM in Generating Segmentation Proposals

For the polyp segmentation task, we apply optimal matching between the segmentation proposals (generated by SAM) and the ground truth annotations. We use those segmentation proposals identified by optimal matching to construct segmentation annotation maps for each selected unlabeled sample. By comparing the generated annotation maps with the ground truths, we find that the segmentation annotations achieve an 89.1% Dice coefficient. We apply the same procedure to the breast cancer segmentation dataset, and find that the Dice coefficient is 75.3% for the benign cases and 64.4% for the malignant cases. These results suggest that SAM performs better on endoscopic images than on ultrasound images (note that ultrasound images in the breast cancer segmentation dataset are in grayscale). This is probably due to the fact that SAM was trained using natural scene color images, which in appearance are closer to endoscopic images than

ultrasound images. Additionally, in the case of ultrasound images, the benign cases consist of objects with more regular shapes and better contrast around the object boundaries, which may lead SAM to yield better coverage with its generated segmentation proposals.

## 5 CONCLUSIONS

In this paper, we proposed a new semi-supervised learning method, SamDSK, which utilizes the Segment Anything Model (SAM) for training a medical image segmentation model. SAM is employed in our proposed optimal matching system for selecting and refining segmentation predictions of unlabeled images (obtained from the current trained segmentation model). A multi-round iterative procedure of model training and labeled set expansion is performed on top of the optimal matching system to gradually improve the target segmentation model while enlarging the labeled training set. Experiments on three datasets, compared with several recently developed SSL methods, demonstrated the effectiveness and advantages of our proposed SamDSK method. Future work may study further expanding the applicable tasks of SamDSK to other medical imaging modalities and segmentation scenarios.

## REFERENCES

- [1] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. Dataset of breast ultrasound images. *Data in Brief*, 28:104863, 2020.
- [2] Yunhao Bai, Duowen Chen, Qingli Li, Wei Shen, and Yan Wang. Bidirectional copy-paste for semi-supervised medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11514–11524, 2023.
- [3] Jørgen Bang-Jensen and Gregory Gutin. *Section 2.3.4: The Bellman-Ford-Moore Algorithm. Digraphs: Theory, Algorithms and Applications*. Springer-Verlag, London, UK, 2000.
- [4] Hritam Basak and Zhaozheng Yin. Pseudo-label guided contrastive learning for semi-supervised medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19786–19797, 2023.
- [5] 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.
- [6] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. TransUNet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.
- [7] Tianle Chen, Zheda Mai, Ruiwen Li, and Wei-lun Chao. Segment Anything Model (SAM) enhanced pseudo labels for weakly supervised semantic segmentation. *arXiv preprint arXiv:2305.05803*, 2023.
- [8] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC). *arXiv preprint arXiv:1902.03368*, 2019.
- [9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *CVPR*, pages 248–255, 2009.
- [10] Deng-Ping Fan, Ge-Peng Ji, Tao Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. PraNet: Parallel reverse attention network for polyp segmentation. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, pages 263–273. Springer, 2020.
- [11] 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.
- [12] Debesh Jha, Pia H Smedsrød, 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.
- [13] 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.
- [14] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on Challenges in Representation Learning, ICML*, volume 3, page 896. Atlanta, 2013.
- [15] Xiangde Luo, Jieneng Chen, Tao Song, and Guotai Wang. Semi-supervised medical image segmentation through dual-task consistency. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 8801–8809, 2021.
- [16] Jun Ma and Bo Wang. Segment anything in medical images. *arXiv preprint arXiv:2304.12306*, 2023.
- [17] Himashi Peiris, Zhaolin Chen, Gary Egan, and Mehrtash Harandi. Duo-SegNet: Adversarial dual-views for semi-supervised medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Proceedings, Part II*, pages 428–438. Springer, 2021.
- [18] 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.
- [19] Mamshad Nayeem Rizve, Kevin Duarte, Yogesh S Rawat, and Mubarak Shah. In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning. In *International Conference on Learning Representations*, 2021.
- [20] Constantin Marc Seibold, Simon Reiß, Jens Kleesiek, and Rainer Stiefelhagen. Reference-guided pseudo-label generation for medical semantic segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 2171–2179, 2022.
- [21] 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.
- [22] Nima Tajbakhsh, Suryakanth R Gurudu, and Jianming Liang. Automated polyp detection in colonoscopy videos using shape and context information. *IEEE Transactions on Medi-*

- cal Imaging*, 35(2):630–644, 2015.
- [23] Bethany H Thompson, Gaetano Di Caterina, and Jeremy P Voisey. Pseudo-label refinement using superpixels for semi-supervised brain tumour segmentation. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2022.
- [24] David Vázquez, Jorge Bernal, F Javier Sánchez, Gloria Fernández-Esparrach, Antonio M López, Adriana Romero, Michal Drozdzał, and Aaron Courville. A benchmark for endoluminal scene segmentation of colonoscopy images. *Journal of Healthcare Engineering*, 2017, 2017.
- [25] Ping Wang, Jizong Peng, Marco Pedersoli, Yuanfeng Zhou, Caiming Zhang, and Christian Desrosiers. CAT: Constrained adversarial training for anatomically-plausible semi-supervised segmentation. *IEEE Transactions on Medical Imaging*, 2023.
- [26] 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.
- [27] Yicheng Wu, Zongyuan Ge, Donghao Zhang, Minfeng Xu, Lei Zhang, Yong Xia, and Jianfei Cai. Mutual consistency learning for semi-supervised medical image segmentation. *Medical Image Analysis*, 81:102530, 2022.
- [28] 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.
- [29] 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.
- [30] Ying Zhang, Tao Xiang, Timothy M Hospedales, and Huchuan Lu. Deep mutual learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4320–4328, 2018.
- [31] Yizhe Zhang, Lin Yang, Jianxu Chen, Maridel Fredericksen, David P Hughes, and Danny Z Chen. Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Proceedings, Part III 20*, pages 408–416. Springer, 2017.
- [32] Tao Zhou, Yizhe Zhang, Yi Zhou, Ye Wu, and Chen Gong. Can SAM segment polyps? *arXiv preprint arXiv:2304.07583*, 2023.