This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Plug-and-Play PPO: An Adaptive Point Prompt Optimizer Making SAM Greater

Xueyu Liu[1†], Rui Wang[2†], Yexin Lai[1], Guangze Shi[1], Feixue Shao[1], Fang Hao[1],
Jianan Zhang[1], Jia Shen[1], Yongfei Wu[1*], Wen Zheng[1]
[1]Taiyuan University of Technology
[2]University of Science and Technology of China

Abstract

Powered by extensive curated training data, the Segment Anything Model (SAM) demonstrates impressive generalization capabilities in open-world scenarios, effectively guided by user-provided prompts. However, the class-agnostic characteristic of SAM renders its segmentation accuracy highly dependent on prompt quality. In this paper, we propose a novel plug-and-play dual-space Point Prompt Optimizer (PPO) designed to enhance prompt distribution through deep reinforcement learning (DRL)-based heterogeneous graph optimization. PPO optimizes initial prompts for any task without requiring additional training, thereby improving SAM's downstream segmentation performance. Specifically, PPO constructs a dual-space heterogeneous graph, leveraging the robust feature-matching capabilities of a foundational pre-trained model to create internal feature and physical distance matrices. A DRL policy network iteratively refines the distribution of prompt points, optimizing segmentation predictions. We conducted experiments on four public datasets. The ablation study explores the necessity and balance of optimizing prompts in both feature and physical spaces. The comparative study shows that PPO enables SAM to surpass recent one-shot methods. Additionally, experiments with different initial prompts demonstrate PPO's generality across prompts generated by various methods. In conclusion, PPO redefines the prompt optimization problem as a heterogeneous graph optimization task, using DRL to construct an effective, plug-and-play prompt optimizer. This approach holds potential for broader applications across diverse segmentation tasks and provides a promising solution for point prompt optimization. The source code and demo are available at https://github.com/XueyuLiu/PPO.

1. Introduction

Segmentation, as a fundamental task, plays a pivotal role in visual understanding systems. The prevailing training paradigm for segmentation relies on data-driven approaches, where large datasets are used to train models. While this method has achieved remarkable success, it faces several challenges, including the computational resources to train models for new tasks and the risk of domain-specific optimizations that can limit generalizability [18, 22].

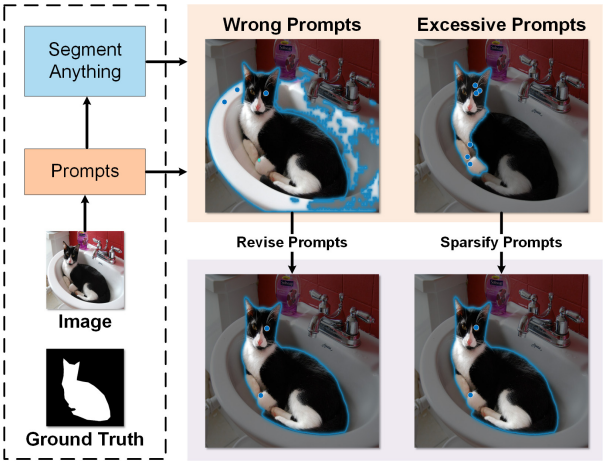


Figure 1. Ineffective prompts can diminish SAM's performance. Incorrect prompts may cause the segmentation mask to capture irrelevant regions, while excessive prompts for adjacent areas can lead to neglect of other parts of the object. Therefore, prompt optimization is essential to improve SAM's performance.

Recent breakthroughs in Pretrained Foundation Models (PFMs) in computer vision (CV) have demonstrated impressive capabilities in domain-invariant segmentation across diverse scenarios [11, 29]. Vision Foundation Models (VFMs), such as DINOv2 [2, 27], acquire generalized visual features by capturing intricate information at both the image and pixel levels using only raw image data. This approach enables these models to learn generic visual representations, ensuring robust zero-shot transferability for various downstream tasks. Meanwhile, the Segment Anything Model (SAM) [14, 30], trained on large-scale, high-quality datasets, has shown a strong ability to generalize across tasks and data distributions that are not encountered during

*Corresponding author. †Co-first author.

training. However, SAM relies heavily on manual prompts, such as bounding boxes and points, to identify objects of interest, as it is pretrained in a class-agnostic manner. Consequently, achieving task-specific adaptation of SAM is crucial to bridge the gap between its class-agnostic predictions and the specific needs of industrial applications.

Some works attempt to train prompt encoders that enables SAM to adapt to downstream segmentation tasks [8, 15, 34, 43]. However, these task-specific prompt encoders require additional training for new segmentation scenarios. Recently, other efforts have focused on automatically generating geometric visual prompts, such as point prompts, to enhance SAM's segmentation performance [21, 23, 37]. In such methods, the quality of prompt schemes remains a critical factor influencing downstream segmentation performance. As illustrated in Fig. 1, incorrect prompts can result in erroneous segmentation, while excessive prompts in specific areas may cause SAM to focus disproportionately on those regions, neglecting other parts of the target. Current automatic prompt engineering methods often lack optimization in prompt distribution, leading to lower-quality prompts and segmentation performance. Therefore, there is an urgent need for methods that can refine prompts to effectively enhance SAM's performance.

To address this, effective point prompts for SAM should exhibit certain characteristics in both feature and physical spaces. In the feature space, regions containing positive prompts should be homogeneous to ensure consistency and clarity within the segmentation target, while regions between positive and negative prompts should be heterogeneous, distinguishing the target from the background. In the physical space, positive and negative prompts should be sufficiently spaced apart to mitigate the effects of excessive prompts and prevent over-segmentation, while also being close enough to boundaries to ensure precise delineation.

Building upon the aforementioned inspiration and committed to addressing the challenges mentioned above, we propose a novel Plug-and-Play **P**oint **P**rompt **O**ptimizer (PPO) designed to enhance SAM's segmentation performance by automatically optimizing initial point prompt schemes. Leveraging DINOv2's powerful feature association capabilities [27], we construct an internal feature and physical distance matrix within each image to create a heterogeneous graph environment. To optimize prompt distribution in dual spaces, we employ Deep Reinforcement Learning (DRL), a framework that trains agents to maximize cumulative rewards through interaction and feedback mechanisms. By designing specific actions and reward functions, PPO optimizes prompt point distribution under the guidance of DRL and inputs them into SAM to obtain refined segmentation results. We validated our approach through experiments on two natural and two medical image datasets. Ablation studies demonstrate that PPO-optimized prompt distributions significantly enhance segmentation performance. Comparative experiments show that PPO outperforms recent methods that have not been trained on in-domain datasets. Additionally, experiments with different initial prompts confirm PPO's generality across prompts generated by various methods. The main contributions are summarized as follows:

- To the best of our knowledge, we are the first to define an optimization direction for point prompts in both feature and physical spaces. Additionally, we are the first to use DRL to optimize point prompts, achieving prompt optimization without additional training for new tasks.
- PPO operates in a plug-and-play manner, allowing for the optimization of any initial geometric point prompts to improve SAM's segmentation performance.
- Extensive experiments demonstrate the effectiveness of PPO in optimizing point prompts across various initial prompt schemes.

2. Related Work

2.1. Pretrained Foundation Models

In natural language processing (NLP), PFMs, such as BERT [5], ChatGPT, and GPT-4, have become foundational tools for a wide range of downstream tasks across diverse data modalities. By leveraging extensive text data, PFMs provide effective parameter initialization for various NLP applications [44]. The success of PFMs in NLP has inspired researchers to explore their application in CV. Visual language models like CLIP [29] and GLIP [19], trained on large-scale datasets of millions of image-text pairs from the web, exhibit strong zero-shot generalization across multiple vision tasks, offering representations with robust generalization and transfer capabilities.

Building on this trend, Meta AI introduced the SAM, a notable advancement in applying VFMs. SAM allows users to specify image segmentation tasks through prompts and demonstrates impressive zero-shot learning, enabling it to generalize effectively across diverse tasks without task-specific training [14]. However, SAM's segmentation performance heavily relies on the quality of input prompts. To address this, we propose an adaptive point prompt optimizer that automatically refines prompts to enhance SAM's segmentation performance.

In another significant development, DINOv2 [27] leverages discriminative self-supervised learning at both image and patch levels, showing exceptional patch-matching capabilities that allow it to capture semantic similarities across images of the same target. By utilizing these patch-matching capabilities, we can extract feature distances between patches within an image, providing a theoretical foundation for constructing heterogeneous graphs to achieve effective prompt optimization.

2.2. Automatic Prompt Engineering

In the field of NLP, extensive researches have been conducted on designing automatic prompt engineering, with prompt tuning methods achieving significant advances in enhancing PFMs' performance through optimized language prompts [1, 12, 13, 16, 28, 41, 45, 45]. Simultaneously, efforts have been made to explore automatic prompt engineering in CV. On the one hand, some recent works have attempted to train prompt encoders that enable SAM to adapt to downstream segmentation tasks [9, 32, 34, 40]. For example, Sun et al. proposed VRP-SAM, an extension of SAM that encodes various granularities of visual references directly into prompt embeddings for mask generation [34]. These approaches encode SAM's prompts into prompt embeddings via a prompt encoder, which are then fed into SAM's mask decoder to generate target masks. However, these methods are often task-sensitive, requiring training different prompt encoders for different tasks, thus limiting rapid task transfer.

On the other hand, some works attempt to generate geometric prompts to achieve automatic prompt engineering for SAM, generally following two main paradigms. The first paradigm uses a small amount of annotated data to train a segmentation model, then samples prompts from pseudo-labels generated by this model [3, 26, 36, 42, 43]. For instance, Li et al. randomly selected prompt points from masks generated by coarse segmentation, which were subsequently input into SAM after fine-tuning to obtain segmentation results [17]. Wang et al. utilized a novel prototype learning framework to generate pseudo-labels and extract prompts, which were then incorporated into SAM for segmentation [35]. Dai et al. developed various methods to expand prompt points, exploring the potential of a single prompt point for enhanced segmentation performance [4].

However, this first paradigm still requires a small amount of data to train the prompt generator, also limiting task transferability. To improve transferability, the second paradigm generates initial prompts by leveraging feature similarity between reference and target images without additional training [21, 23, 37, 43]. For example, Zhang et al. calculated feature similarity between pixels in the target and reference images to create a point prompt scheme using only a single reference image [43]. Liu et al. introduced Matcher, an instance-level and bidirectional matching strategy for prompt point generation [23], and Liu et al. proposed a series of prompt optimization strategies for the automatic segmentation of basement membranes in electron microscopy images [21]. However, the quality of prompts generated by these methods is often limited, as they do not sufficiently consider the relationships between positive and negative prompts, resulting in redundant and erroneous prompts that negatively affect segmentation performance.

In our work, PPO addresses this limitation by constructing a heterogeneous graph based on dual-space distances between initial prompts within each image. This approach optimizes any automatically generated geometric point prompts without additional training for new tasks, enhancing SAM's segmentation performance while preserving task transferability by ensuring high prompt quality.

3. Methodology

This section introduces our plug-and-play PPO, which is based on a heterogeneous graph structure designed to enhance the segmentation performance of SAM by optimizing initial point prompts. An overview of PPO is illustrated in Fig. 2. Specifically, we first compute the feature distance matrix and the physical distance matrix for each image. Using these matrices, we construct a heterogeneous graph where the edge attributes are defined by the feature and physical distances based on the initial prompt indices. During the training phase, we use this heterogeneous graph as the environment to design an action list and reward function to train the optimization agent. In the inference phase, the trained agent activates the appropriate point prompts as a final prompt scheme, which is then input into SAM to obtain the final segmentation result. In the following subsections, we will describe the main components in detail.

3.1. Dual-Space Distance Matrix Calculation

The core objective of our method is to optimize positive and negative point prompts in both feature and distance spaces. To achieve this, we construct a heterogeneous graph comprising two homogeneous graphs that represent the feature and distance spaces. Edge values are determined by creating two patch-level distance matrices for each image X.

First, we divide X into patches x using a sliding window approach, where $x_i = x_1, x_2, \ldots, x_n$. We then use the image encoder from DINOv2 [27] to extract patch-level features for x, represented by f, where $f_i = f_1, f_2, \ldots, f_n$. Next, we calculate the feature matrix M_f as follows:

$$M_f(i,j) = \|f_i - f_j\|, \tag{1}$$

where $M_f(i,j)$ represents the Euclidean distance D_f between the i-th patch feature f_i and the j-th patch feature f_j. Similarly, we calculate the physical distance matrix M_p using the center points of the patches x as follows:

$$M_p(i,j) = \|x_i - x_j\|, \tag{2}$$

where $M_p(i,j)$ represents the Euclidean distance D_p between the i-th patch center x_i and the j-th patch center x_j.

3.2. Graph Environment Creation

Inspired by coarse segmentation sampling [3, 26] and feature matching techniques [21, 23], we can obtain initial

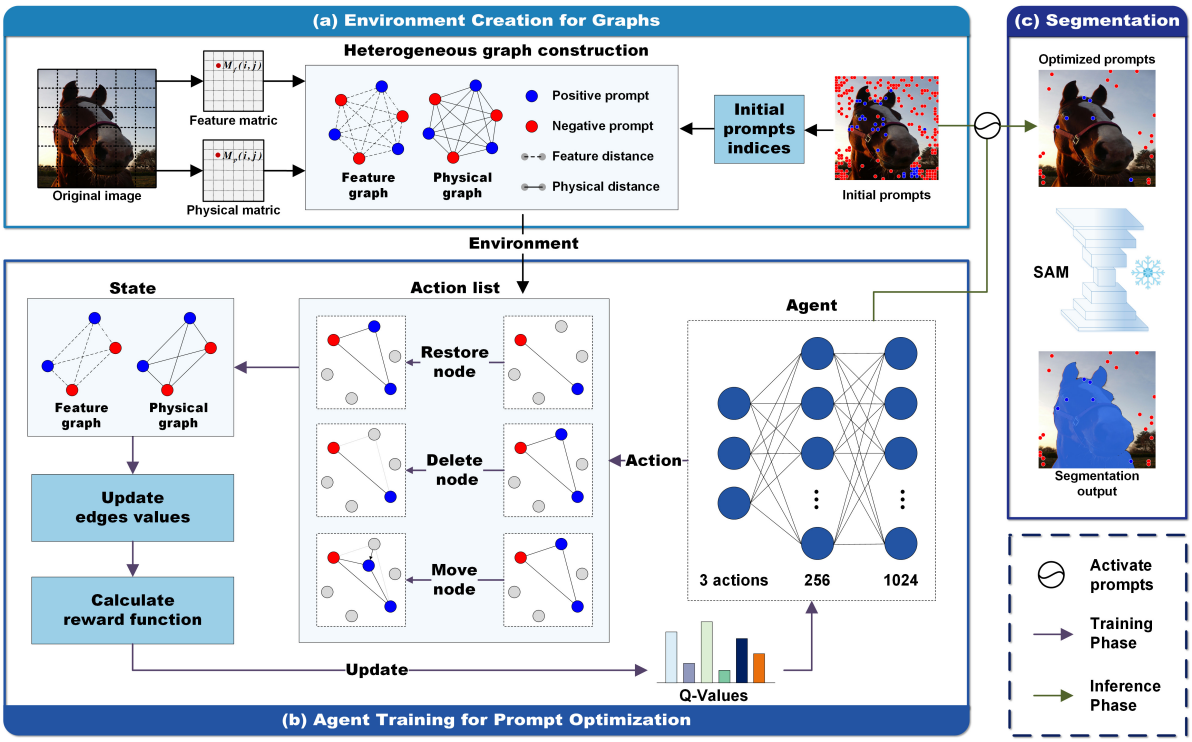


Figure 2. The PPO workflow is designed to enhance SAM's segmentation performance by optimizing initial point prompts through two key components: graph-based environment creation and agent training for prompt optimization. After optimized, the prompts are fed into SAM to obtain segmentation results.

point prompt indices for optimization. Specifically, we define a heterogeneous graph $G = (V, E, \phi, \psi)$, where V is the set of nodes and E is the set of edges. The node type set $\mathcal{A} = \{\text{positive nodes } V_p, \text{negative nodes } V_n\}$ is represented by the mapping function $\phi(v)$: $V \rightarrow \mathcal{A}$, and the edge type set $\mathcal{R} = \{\text{feature edges } E_f, \text{physical edges } E_p\}$ is represented by the mapping function $\psi(e)$: $E \rightarrow \mathcal{R}$.

We use the heterogeneous graph G as the environment for DRL. By selecting different nodes, we can obtain different edge values as feedback.

3.3. Agent Training for Prompt Optimization

On the one hand, an effective point prompt scheme should ensure correctness. In an image, the features of the segmentation target are often homogeneous, while the features of the background are heterogeneous. Therefore, we aim for the feature distance between positive point prompts to be sufficiently close, and the feature distance between positive and negative point prompts to be sufficiently far. On the other hand, an effective point prompt scheme should also ensure sparsity. Over-prompting in adjacent areas can degrade segmentation performance. Thus, the physical distance between positive point prompts (and between negative point prompts) should not be too close. Additionally, to enhance boundary segmentation accuracy, the physical distance between positive and negative point prompts should not be too far.

A DRL model can infer actions and receive rewards from the environment, reflecting the significance of these actions and updating the agent's state accordingly [38]. Using DRL, we aim to train an agent to activate appropriate point prompts to optimize the distribution. The detailed process is shown in Algorithm 1. The key components of agent training for prompt optimization will be introduced as follows:

Actions: We define three possible actions: (1) restore node A_r, which randomly activates a node and connects the corresponding edges; (2) delete node A_d, which randomly deletes a node and removes the corresponding edges; (3) move node A_m, which randomly selects a direction and moves the index position of a node. Additionally, we aim to maintain a stable number of prompts. Therefore, if the number of nodes $|V_p|$ or $|V_n|$ in the current state is less than the minimum value, the only possible action is A_r. Conversely, if the $|V_p|$ or $|V_n|$ exceeds the maximum value, the only possible action is A_d. Only when the $|V_p|$ or $|V_n|$ is between the minimum and maximum values are all three actions, A_r, A_d, and A_m, possible.

States: For each action, there will be different nodes, resulting in a different heterogeneous graph G^s. Based on

Algorithm 1 PPO training process

Input: Heterogeneous graph environment G
Parameter: Max step; Max epoch;
Max nodes N_{max}; Min nodes N_{min}
Output: Q-values

Initializing the Q-values, Epoch, Step.
while Epoch $<$ Max epoch **do**
  Initializing the Step.
  while Step $<$ Max step **do**
    Emptying all nodes and edges.
    Emptying reward.
    Calculating the number of Nodes N_{pos} and N_{neg} .
    if $N_{pos} < N_{min}$ or $N_{neg} < N_{min}$ **then**
      Action list=$\{A_r\}$.
    else if $N_{pos} > N_{max}$ or $N_{neg} > N_{max}$ **then**
      Action list=$\{A_d\}$.
    else
      Action list=$\{A_r, A_d, A_m\}$.
    end if
    Perform a random action and update the status s
    Adding the reward r_s based on the reward function.
    Updating Q-values .
    Step+=1.
  end while
  Epoch+=1.
end while
return solution

G^s, we can calculate the following mean edge values:

$$D_f^s(p,p) = \frac{1}{|V_p||V_p|} \sum_{i \in V_p} \sum_{j \in V_p} M_f(i,j), \tag{3}$$

$$D_f^s(p,n) = \frac{1}{|V_p||V_n|} \sum_{i \in V_p} \sum_{j \in V_n} M_f(i,j), \tag{4}$$

$$D_p^s(p,p) = \frac{1}{|V_p||V_p|} \sum_{i \in V_p} \sum_{j \in V_p} M_p(i,j), \tag{5}$$

$$D_p^s(n,n) = \frac{1}{|V_n||V_n|} \sum_{i \in V_n} \sum_{j \in V_n} M_p(i,j), \tag{6}$$

$$D_p^s(p,n) = \frac{1}{|V_p||V_n|} \sum_{i \in V_p} \sum_{j \in V_n} M_p(i,j). \tag{7}$$

In state s, $D_f^s(p,p)$ and $D_f^s(p,n)$ represents the mean feature distance between positive prompts, and between positive and negative prompts. Similarly, $D_p^s(p,p)$, $D_p^s(n,n)$, and $D_p^s(p,n)$ represent the mean physical distances in state s between positive prompts, negative prompts, and between positive and negative prompts, respectively.

Rewards: Based on the various distances obtained from G^s, we can set a reward function according to the optimization objectives. The reward function for the current state, r_s, is defined based on the changes in distance from the previous state:

$$\begin{aligned} r_s = &-\alpha(D_f^s(p,p) - D_f^{s-1}(p,p)) \\ &+ \alpha(D_f^s(p,n) - D_f^{s-1}(p,n)) \\ &+ \beta(D_p^s(p,p) - D_p^{s-1}(p,p)) \\ &+ \beta(D_p^s(n,n) - D_p^{s-1}(n,n)) \\ &- \beta(D_p^s(p,n) - D_p^{s-1}(p,n)), \end{aligned} \tag{8}$$

where coefficients α and β balance the feature and physical distance, respectively. Using this reward function, if the current state's $D_f^s(p,p)$ and $D_p^s(p,n)$ increase, a penalty is applied. Conversely, if the current state's $D_f^s(p,n)$, $D_p^s(p,p)$, and $D_p^s(n,n)$ increase, a reward is given.

Policy Network: We adopt the Q-network [25] as the policy network, which consists of two fully connected layers with 1024 and 64 neurons, respectively, and a single output layer with 3 neurons. This output layer predicts the Q-value of 3 actions based on the current state. The policy network is regarded as the approximate action-value function Q_s and takes the state s as its input to approximate the underlying action-value function. This approximation is achieved by repeatedly performing a gradient descent step on $(y_s - Q_s)^2$ with respect to the policy network parameters at each step. The Q-value y_s is defined as:

$$y_s = \begin{cases} r_s + \gamma \max\limits_{s+1} Q_{s+1}, & \text{if not stop} \\ r_s, & \text{if stop} \end{cases}, \tag{9}$$

where the discount factor γ represents the balance between the current and future rewards.

In the training process of PPO, we initially delete all nodes and treat them as potential activatable nodes in the action environment. As the agent attempts to activate different nodes, the edges of the heterogeneous graph acquire different values. Based on the reward function, the agent receives rewards or penalties and updates the Q-values accordingly. The policy network repeatedly interacts with the segmentation network, updates the network weights according to Eq. (11), and ultimately learns effective policies.

4. Experimental Results

4.1. Implementation Details

Our experiments are carried out on a Linux server platform equipped with an NVIDIA Tesla V100. During the PPO training phase, we set the step size for each iteration to 100 and the number of training epochs to 1000. The discount factor γ of Q-learning is set to 0.1. We use the ϵ-greedy strategy to balance exploration and exploitation [25]. During ϵ-greedy training, the agent can select a random action

with probability ϵ and a predicted action generated by the Q-network with probability $1 - \epsilon$. The ϵ parameters are set as follows: $\epsilon_{\text{start}} = 1$, $\epsilon_{\text{end}} = 0.1$, and $\epsilon_{\text{decay}} = 0.995$. Additionally, to ensure stability during PPO training, the probability of action A_m is always set to be 0.5 times the probability of the other actions.

In the testing phase, PPO requires no additional training for new tasks; it simply inputs initial prompts and images to output optimized prompt schemes that serve SAM. For evaluation, we use the mean Dice Similarity Coefficient (mDSC) and the mean Hausdorff Distance (mHD) as metrics to assess the segmentation performance.

4.2. Dataset Description

To ensure that PPO remains unseen to in-domain datasets for downstream tasks, we trained PPO on the GBMSD dataset [21], generating initial prompts through feature matching and constructing a heterogeneous graph as the environment for DRL learning. During testing, we evaluated the model on natural image domains using the FSS-1000 [20] and VOC2012 [6] datasets. The FSS-1000 dataset comprises mask-annotated images from 1,000 classes, with 10 images per class, while the VOC2012 dataset is a widely-used benchmark for semantic segmentation tasks, containing 20 object classes and a total of 4,396 images. For medical image domains, we used the ISIC2018 [33] and Kvasir [10]. The ISIC2018 is a large-scale collection of dermatoscopic images focused on lesion segmentation, consisting of 2,594 images. The Kvasir, on the other hand, is an endoscopic dataset aimed at colorectal polyp segmentation, containing 1,000 gastroenterology images of colorectal polyps along with their corresponding segmentation masks.

We generated initial prompts based on two mainstream point prompt-generation paradigms. The first paradigm is the feature matching (FM) approach [21, 23], which requires no additional training for new tasks. In this approach, only one annotated reference image is selected for each segmentation task, and this image is used to generate the initial prompt for testing images. The second paradigm is the coarse segmentation sampling (CS) approach, which involves few-shot training for new tasks. For each task, we selected 20 annotated images, trained a U-Net [31] with 200 epochs, and used the resulting coarse segmentation masks to sample initial prompts for the segmentation tasks.

4.3. Ablation Study

We conduct an ablation study on the initial prompts generated by FM to evaluate the segmentation performance of SAM with PPO under varying parameter settings. Quantitative results are presented in Table 1. Initially, when SAM applies unoptimized prompts ($\alpha = 0$, $\beta = 0$), segmentation outcomes are suboptimal across all datasets. In contrast, applying PPO with optimized settings significantly enhances prompt quality, leading to notable performance gains for SAM across various tasks.

When we examine optimization within the feature space ($\alpha = 1$, $\beta = 0$) and physical space ($\alpha = 0$, $\beta = 1$) separately, both settings improve SAM's performance; however, feature space optimization proves to be more effective, yielding a marked improvement over the initial prompts. The better results are achieved when both feature and physical spaces are optimized concurrently ($\alpha = 1$, $\beta = 1$).

Interestingly, further increasing the weight on the physical space ($\beta = 5$) reduces segmentation performance, likely due to an overemphasis on physical distributions at the expense of essential feature space information. In contrast, placing a higher weight on the feature space ($\alpha = 5$, $\beta = 1$) yields the best segmentation results. These experiments demonstrate that the accuracy of point prompts is more crucial than sparsity, though sparsity remains beneficial.

In Fig. 3, we compare SAM's qualitative segmentation results using initial prompts versus PPO-optimized prompts under the best parameter settings. This figure includes images from both natural and medical datasets. The initial prompts contain inaccuracies and excessive redundancy, impairing SAM's segmentation accuracy. After PPO refinement, the prompts become more precise and sparse, allowing the segmentation results to closely align with the ground truth, demonstrating PPO's effectiveness in enhancing SAM's adaptability across diverse segmentation tasks.

4.4. Comparison with One-Shot Methods

Our goal is to enable SAM to adapt to segmentation tasks across diverse scenarios with minimal cost. To this end, we compared recently proposed methods with our PPO-optimized FM prompt generation method, both following the one-shot paradigm, across multiple datasets. As shown in Table 2, FM-PPO achieves competitive performance in various natural image segmentation tasks without training on in-domain datasets, performing comparably to models trained on in-domain data and surpassing other SAM-based prompt generation methods. Additionally, FM-PPO significantly outperforms other segmentation methods on medical datasets without requiring additional training. This demonstrates that PPO-optimized visual prompt generation retains its training-free task transfer capability while ensuring robust segmentation performance for downstream SAM through optimized prompts.

Moreover, Fig. 4 illustrates the qualitative results of training-free methods on natural and medical datasets, providing visual evidence of FM-PPO's effectiveness. The experimental results show that FM-PPO performs well even in small target segmentation by reducing erroneous prompt points. This success is attributed to the homogeneous feature expectation within the segmentation target and the heterogeneous feature expectation between the target and background during agent training. For medical image segmentation, incorrect and redundant prompts due to domain

Table 1. Ablation study on different spatial weights during PPO training. The results with α and β set to 0 represent the initial prompts.

Feature Space	Physical Space	FSS-1000		VOC2012		ISIC		Kvasir	
		mDSC (%↑)	mHD (↓)	mDSC (%↑)	mHD (↓)	mDSC (%↑)	mHD (↓)	mDSC (%↑)	mHD (↓)
α=0	β=0	72.1	9.9	34.7	12.1	48.5	9.9	27.2	13.2
α=1	β=0	81.3	8.5	59.4	10.5	61.7	9.2	56.6	10.3
α=0	β=1	73.5	9.9	38.4	11.7	51.9	9.9	30.5	13.1
α=1	β=1	86.6	7.8	63.7	10.3	69.2	8.5	61.3	9.6
α=1	β=5	82.4	8.9	56.9	10.2	64.9	9.0	58.8	9.7
α=5	β=1	**91.9**	**7.4**	**73.7**	**9.9**	**75.4**	**8.1**	**70.3**	**9.0**

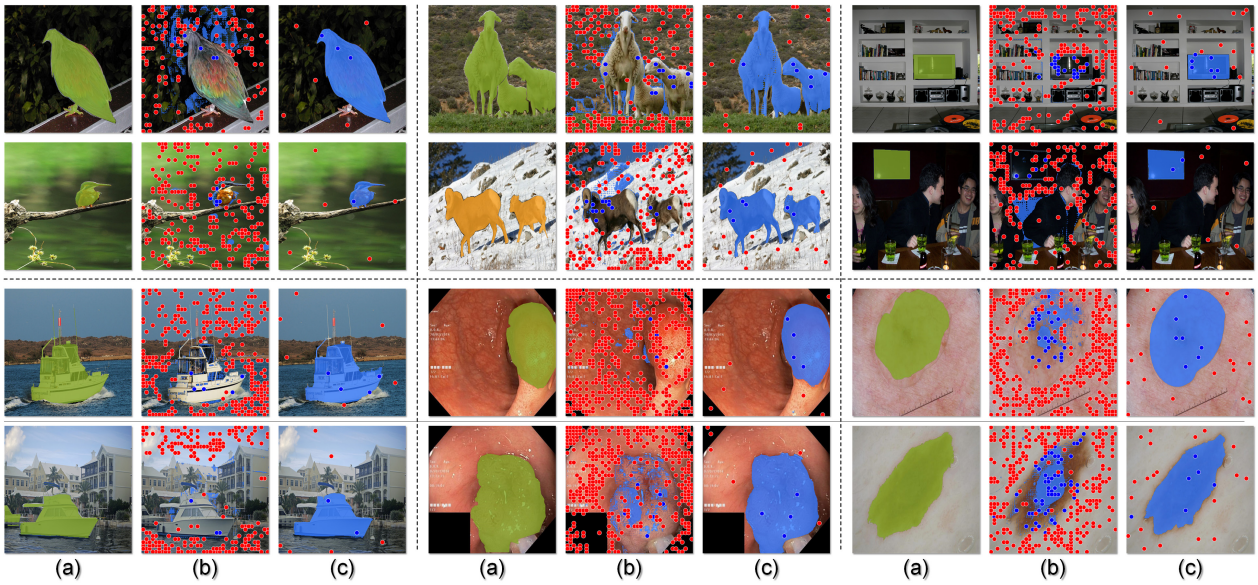


Figure 3. Qualitative segmentation results of initial prompts and optimized prompts. (a) ground truth, (b) segmentation results with initial prompts, (c) segmentation results with optimized prompts.

differences can lead to poorer segmentation outcomes for SAM. Through prompt optimization, FM-PPO effectively bridges the domain gap between medical and natural images, enhancing segmentation performance across both domains. Additionally, FM-PPO promotes a relatively uniform physical distribution of point prompts and close proximity between positive and negative prompts. In contrast, Matcher lacks negative prompts, which leads to excessive background segmentation on medical images with domain shifts, particularly when segmenting smaller targets, resulting in suboptimal performance. We further elaborate on the necessity of negative prompts in the supplementary materials. As a result, our method excels at boundary segmentation, especially in complex medical images where other methods struggle with background-foreground adhesion. FM-PPO can optimize the initial prompts generated by FM without additional training, achieving efficient segmentation across various scenarios.

4.5. Different Initial Prompts Scheme Optimization

To further validate PPO's effectiveness in optimizing prompts, we applied it to various initial prompt generation schemes and evaluated the downstream SAM segmentation. We selected medical datasets with domain gap, as prompts have a greater impact on SAM's segmentation. Table 3 presents the experimental results, with corresponding qualitative results in the supplementary materials. FM and CS serve as two baseline prompt generation methods. As shown, PPO significantly enhances segmentation performance on both datasets by optimizing the prompts. FM generates a richer prompt set based on patch matching, allowing PPO to achieve stronger optimization. In contrast, CS samples within coarse segmentation regions, leading to a concentrated prompt distribution that may overlook some target areas, thus limiting PPO's optimization potential.

Additionally, Matcher [23], a recent approach in visual prompt engineering, utilizes the FM paradigm to achieve task-agnostic segmentation. To demonstrate PPO's

Table 2. Comparison results with one-shot methods in natural and medical image datasets. Gray metrics indicates the model is already trained by in-domain datasets. † indicates the training-free method for segmentation tasks.

Methods	FSS-1000		VOC2012		ISIC		Kvasir	
	mDSC (%↑)	mHD (↓)	mDSC (%↑)	mHD (↓)	mDSC (%↑)	mHD (↓)	mDSC (%↑)	mHD (↓)
HSNet [24]	91.1	8.2	44.5	11.4	57.4	11.3	29.7	16.5
VAT [7]	93.6	7.5	48.2	11.4	49.8	12.9	26.4	16.8
SegGPT† [39]	93.7	7.1	78.2	7.8	61.2	9.5	44.1	14.4
PerSAM† [43]	81.1	8.6	53.6	10.9	37.6	16.1	45.8	16.3
Matcher† [23]	91.6	7.6	69.0	11.1	73.7	9.7	46.2	16.0
FM-PPO† (Ours)	**91.9**	**7.4**	**73.7**	**9.9**	**75.4**	**8.1**	**70.3**	**9.0**

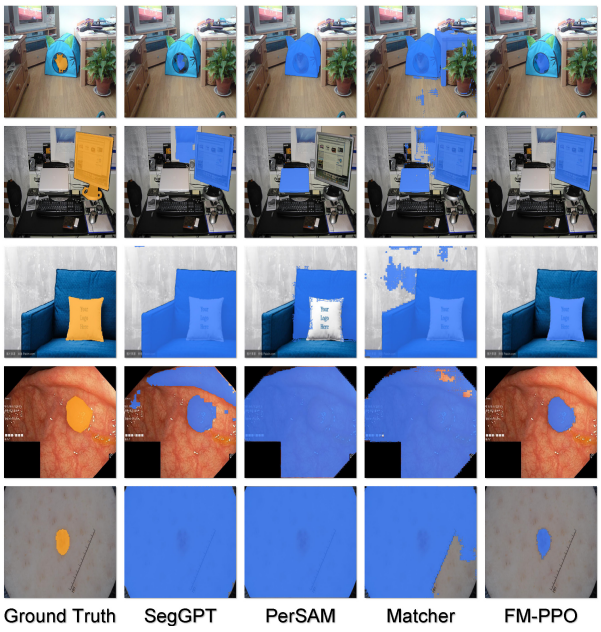


Figure 4. Qualitative segmentation results of different one-shot methods in natural and medical images.

plug-and-play capability, we applied PPO to optimize the prompts generated by Matcher and then fed them back into Matcher for segmentation. Table 3 shows that PPO-optimized prompts significantly improve segmentation performance, particularly on the Kvasir, where the DSC increased by 32.6% for smaller targets. This demonstrates that our method effectively boosts SAM's segmentation performance by optimizing various types of initial prompts.

5. Conclusion

In this paper, we introduce PPO, designed to enhance the distribution of point prompts, thereby improving the downstream automatic segmentation performance of SAM. We first formalize the distribution of point prompts as a dual-space heterogeneous graph and define the optimization direction. Next, we train an agent within this graph environment to automatically activate effective prompt schemes, which are then input into SAM to obtain segmentation results. Our experiments on natural and medical image datasets demonstrate that PPO significantly improves segmentation performance compared to various initial point prompt schemes. In conclusion, PPO offers a promising solution for optimizing any initial point prompts, enhancing the performance of diverse downstream segmentation tasks.

Limitations. Although PPO demonstrates strong performance in optimizing incorrect and redundant point prompts, its effectiveness is limited when the initial prompts lack coverage of the target. For example, in the CS approach, if the pseudo-labels do not include the target, prompts are not sampled over the target region. Future research will focus on designing more comprehensive point prompt generation methods to build a more robust prompt optimizer.

Table 3. Optimization performance of different initial prompts.

Prompt Scheme	ISIC		Kvasir	
	mDSC (%↑)	mHD (↓)	mDSC (%↑)	mHD (↓)
FM	48.5	9.9	27.2	13.2
FM-PPO	**75.4 (↑21.9)**	**8.1 (↓1.8)**	**70.3 (↑43.1)**	**9.0 (↓4.2)**
CS	57.7	11.9	69.1	10.2
CS-PPO	**68.6 (↑10.9)**	**10.8 (↓1.1)**	**79.8 (↑10.7)**	**8.9 (↓1.3)**
Matcher [23]	73.7	9.7	46.2	16.0
Matcher-PPO	**85.9 (↑12.2)**	**9.0 (↓0.7)**	**77.6 (↑32.6)**	**9.3 (↓6.7)**

Acknowledgment

This research was supported by the National Natural Science Foundation of China (Grant No. 61901292 and Grant No. 62201331), the Fundamental Research Program of Shanxi Province (Grant No. 201901D211080, Grant No. 202303021211082, Grant No. 202403021222276 and Grant No. 202403021212262), the Shanxi Province Graduate Education Innovation Project (Grant No. 2022Y196). We also express our gratitude to the all reviewers and area chairs for their constructive feedback.

References

[1] Hyojin Bahng, Ali Jahanian, Swami Sankaranarayanan, and Phillip Isola. Exploring visual prompts for adapting large-scale models. *arXiv preprint arXiv:2203.17274*, 2022. 3

[2] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9650–9660, 2021. 1

[3] Zhen Chen, Qing Xu, Xinyu Liu, and Yixuan Yuan. Un-sam: Universal prompt-free segmentation for generalized nuclei images. *arXiv preprint arXiv:2402.16663*, 2024. 3

[4] Haixing Dai, Chong Ma, Zhengliang Liu, Yiwei Li, Peng Shu, Xiaozheng Wei, Lin Zhao, Zihao Wu, Dajiang Zhu, and Wei Liu. Samaug: Point prompt augmentation for segment anything model. *arXiv preprint arXiv:2307.01187*, 2023. 3

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018. 2

[6] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88:303–338, 2010. 6

[7] Sunghwan Hong, Seokju Cho, Jisu Nam, Stephen Lin, and Seungryong Kim. Cost aggregation with 4d convolutional swin transformer for few-shot segmentation. In *European Conference on Computer Vision*, pages 108–126. Springer, 2022. 8

[8] Duojun Huang, Xinyu Xiong, Jie Ma, Jichang Li, Zequn Jie, Lin Ma, and Guanbin Li. Alignsam: Aligning segment anything model to open context via reinforcement learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3205–3215, 2024. 2

[9] Jiaxing Huang, Kai Jiang, Jingyi Zhang, Han Qiu, Lewei Lu, Shijian Lu, and Eric Xing. Learning to prompt segment anything models. *arXiv preprint arXiv:2401.04651*, 2024. 3

[10] 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, MMM 2020, Daejeon, South Korea, January 5–8, 2020, Proceedings, Part II 26*, pages 451–462. Springer, 2020. 6

[11] 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*, pages 4904–4916. PMLR, 2021. 1

[12] Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. Visual prompt tuning. In *European Conference on Computer Vision*, pages 709–727. Springer, 2022. 3

[13] Muhammad Uzair Khattak, Hanoona Rasheed, Muhammad Maaz, Salman Khan, and Fahad Shahbaz Khan. Maple: Multi-modal prompt learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19113–19122, 2023. 3

[14] 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. 1, 2

[15] Hyeokjun Kweon and Kuk-Jin Yoon. From sam to cams: Exploring segment anything model for weakly supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19499–19509, 2024. 2

[16] Yiming Lei, Jingqi Li, Zilong Li, Yuan Cao, and Hongming Shan. Prompt learning in computer vision: a survey. *Frontiers of Information Technology & Electronic Engineering*, 25(1):42–63, 2024. 3

[17] Chengyin Li, Prashant Khanduri, Yao Qiang, Rafi Ibn Sultan, Indrin Chetty, and Dongxiao Zhu. Autoprompting sam for mobile friendly 3d medical image segmentation. *arXiv preprint arXiv:2308.14936*, 2023. 3

[18] Jingjing Li, Zhiqi Yu, Zhekai Du, Lei Zhu, and Heng Tao Shen. A comprehensive survey on source-free domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 1

[19] Liunian Harold Li, Pengchuan Zhang, Haotian Zhang, Jianwei Yang, Chunyuan Li, Yiwu Zhong, Lijuan Wang, Lu Yuan, Lei Zhang, Jenq-Neng Hwang, et al. Grounded language-image pre-training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10965–10975, 2022. 2

[20] Xiang Li, Tianhan Wei, Yau Pun Chen, Yu-Wing Tai, and Chi-Keung Tang. Fss-1000: A 1000-class dataset for few-shot segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2869–2878, 2020. 6

[21] Xueyu Liu, Guangze Shi, Rui Wang, Yexin Lai, Jianan Zhang, Lele Sun, Quan Yang, Yongfei Wu, Ming Li, Weixia Han, et al. Feature-prompting gbmseg:

One-shot reference guided training-free prompt engineering for glomerular basement membrane segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 276–285. Springer, 2024. 2, 3, 6

[22] Xueyu Liu, Rui Wang, Yexin Lai, Yongfei Wu, Hangbei Cheng, Yuanyue Lu, Jianan Zhang, Ning Hao, Chenglong Ban, Yanru Wang, Shuqin Tang, Yuxuan Yang, Ming Li, Xiaoshuang Zhou, and Wen Zheng. Msmtseg: Multi-stained multi-tissue segmentation of kidney histology images via generative self-supervised meta-learning framework. *IEEE Journal of Biomedical and Health Informatics*, pages 1–12, 2024. 1

[23] Yang Liu, Muzhi Zhu, Hengtao Li, Hao Chen, Xinlong Wang, and Chunhua Shen. Matcher: Segment anything with one shot using all-purpose feature matching. *arXiv preprint arXiv:2305.13310*, 2023. 2, 3, 6, 7, 8

[24] Juhong Min, Dahyun Kang, and Minsu Cho. Hypercorrelation squeeze for few-shot segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6941–6952, 2021. 8

[25] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015. 5

[26] Saiyang Na, Yuzhi Guo, Feng Jiang, Hehuan Ma, and Junzhou Huang. Segment any cell: A sam-based auto-prompting fine-tuning framework for nuclei segmentation. *arXiv preprint arXiv:2401.13220*, 2024. 3

[27] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 1, 2, 3

[28] Xiaoyu Qiu, Hao Feng, Yuechen Wang, Wengang Zhou, and Houqiang Li. Progressive multi-modal conditional prompt tuning. In *Proceedings of the 2024 International Conference on Multimedia Retrieval*, pages 46–54, 2024. 3

[29] 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*, pages 8748–8763. PMLR, 2021. 1, 2

[30] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, et al. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*, 2024. 1

[31] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 6

[32] Tal Shaharabany, Aviad Dahan, Raja Giryes, and Lior Wolf. Autosam: Adapting sam to medical images by overloading the prompt encoder. *arXiv preprint arXiv:2306.06370*, 2023. 3

[33] Junji Shiraishi, Shigehiko Katsuragawa, Junpei Ikezoe, Tsuneo Matsumoto, Takeshi Kobayashi, Kenichi Komatsu, Mitate Matsui, Hiroshi Fujita, Yoshie Kodera, and Kunio Doi. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *American journal of roentgenology*, 174(1):71–74, 2000. 6

[34] Yanpeng Sun, Jiahui Chen, Shan Zhang, Xinyu Zhang, Qiang Chen, Gang Zhang, Errui Ding, Jingdong Wang, and Zechao Li. Vrp-sam: Sam with visual reference prompt. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23565–23574, 2024. 2, 3

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

[36] Jiaqi Wang, Zhengliang Liu, Lin Zhao, Zihao Wu, Chong Ma, Sigang Yu, Haixing Dai, Qiushi Yang, Yiheng Liu, Songyao Zhang, et al. Review of large vision models and visual prompt engineering. *Meta-Radiology*, page 100047, 2023. 3

[37] Rui Wang, Zhouwang Yang, and Yanzhi Song. Osam-fundus: A training-free, one-shot segmentation framework for optic disc and cup in fundus images. *Biomedical Signal Processing and Control*, 100:107069, 2025. 2, 3

[38] Xu Wang, Sen Wang, Xingxing Liang, Dawei Zhao, Jincai Huang, Xin Xu, Bin Dai, and Qiguang Miao. Deep reinforcement learning: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 35 (4):5064–5078, 2022. 4

[39] Xinlong Wang, Xiaosong Zhang, Yue Cao, Wen Wang, Chunhua Shen, and Tiejun Huang. Seggpt: Towards segmenting everything in context. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1130–1140, 2023. 8

[40] Bin Xie, Hao Tang, Bin Duan, Dawen Cai, and Yan Yan. Masksam: Towards auto-prompt sam with mask classification for medical image segmentation. *arXiv preprint arXiv:2403.14103*, 2024. 3

[41] Hantao Yao, Rui Zhang, and Changsheng Xu. Visual-language prompt tuning with knowledge-guided context optimization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6757–6767, 2023. 3

[42] Chaoning Zhang, Fachrina Dewi Puspitasari, Sheng Zheng, Chenghao Li, Yu Qiao, Taegoo Kang, Xinru Shan, Chenshuang Zhang, Caiyan Qin, Francois Rameau, et al. A survey on segment anything model (sam): Vision foundation model meets prompt engineering. *arXiv preprint arXiv:2306.06211*, 2023. 3

[43] Renrui Zhang, Zhengkai Jiang, Ziyu Guo, Shilin Yan, Junting Pan, Hao Dong, Peng Gao, and Hongsheng Li. Personalize segment anything model with one shot. *arXiv preprint arXiv:2305.03048*, 2023. 2, 3, 8

[44] Ce Zhou, Qian Li, Chen Li, Jun Yu, Yixin Liu, Guangjing Wang, Kai Zhang, Cheng Ji, Qiben Yan, Lifang He, et al. A comprehensive survey on pretrained foundation models: A history from bert to chatgpt. *arXiv preprint arXiv:2302.09419*, 2023. 2

[45] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Conditional prompt learning for vision-language models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16816–16825, 2022. 3