InfoSAM: Fine-Tuning the Segment Anything Model from An Information-Theoretic Perspective

ICML International Conference On Machine Learning

Project Page

Yuanhong Zhang*[1], Muyao Yuan*[1], Weizhan Zhang†[1], Tieliang Gong[1], Wen Wen[1], Jiangyong Ying[2], Weijie Shi[1]

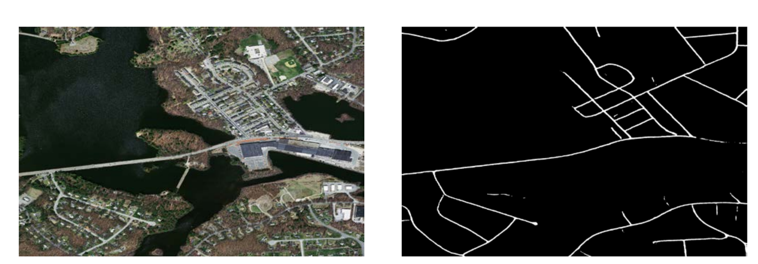
* Equal Contribution [1] Xi'an Jiaotong University [2] China Telecom E-surfing Vision Technology

Introduction

Background

Parameter-efficient fine-tuning (PEFT) is a promising approach to unleash the potential of Segment Anything (SAM) in *diverse scenarios*.

✓ Leaf disease segmentation



✓ Remote sensing road segmentation

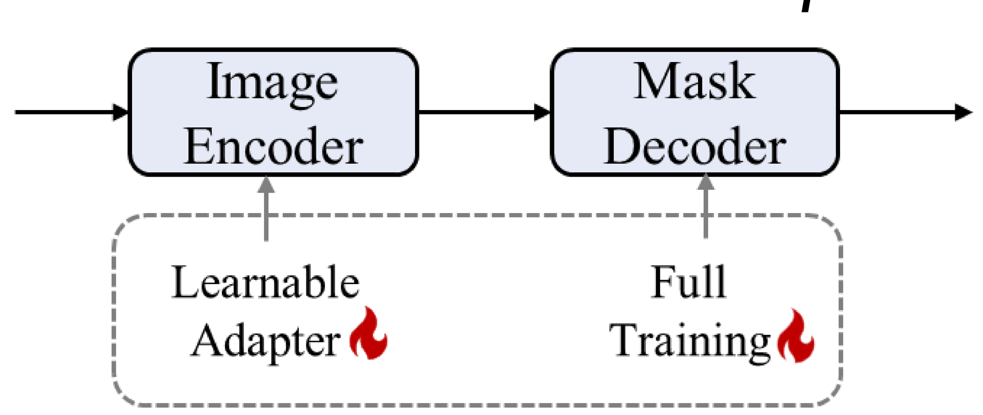
? *Can we extract the domain-invariant information preserved in the pre-trained SAM to enhance the PEFT process?*

Challenges:

- How to extract the domain-invariant information?
- How to transfer it to the finetuned process?

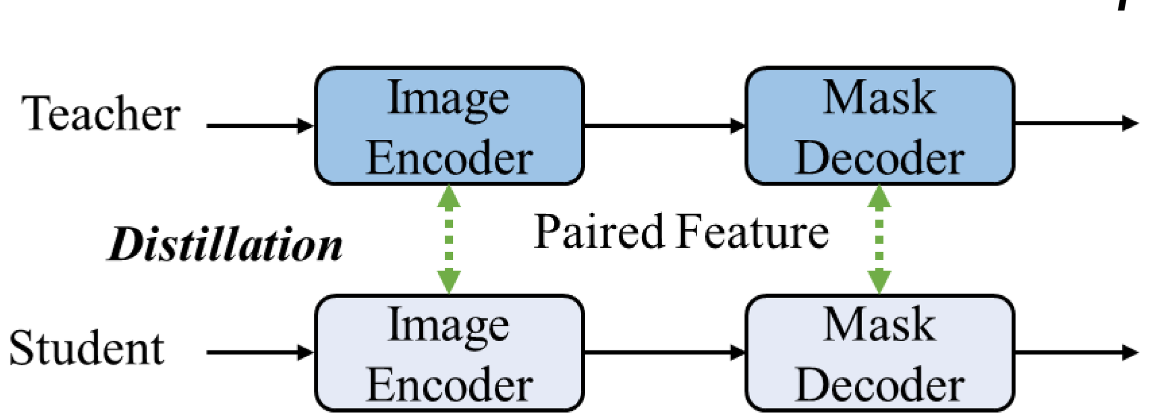
Previous solutions

- *Traditional PEFT Methods for SAM*



Not Consider Relation

- *Traditional Distillation Methods for SAM*



No Relation Guidance

Our Solutions

InfoSAM leverages matrix-based Rényi mutual information to (i) **compress domain-invariant relations** and (ii) **maximize alignment** between pre-trained and fine-tuned relational knowledge.

Formulation

- Let z_i^T / z_i^S and z_m^T / z_m^S denote the image encoder features and mask decoder tokens from the teacher (pre-trained SAM) and student (fine-tuned SAM), respectively.
- To prioritize the domain-invariant relations r^T :

$$I_\alpha(z_i^T, z_m^T; r^T) \le I_c \;\Rightarrow\; \mathcal{L}_r = -log_2\|G_r^T\|_F^2 + log_2\|G_{imr}^T\|_F^2$$

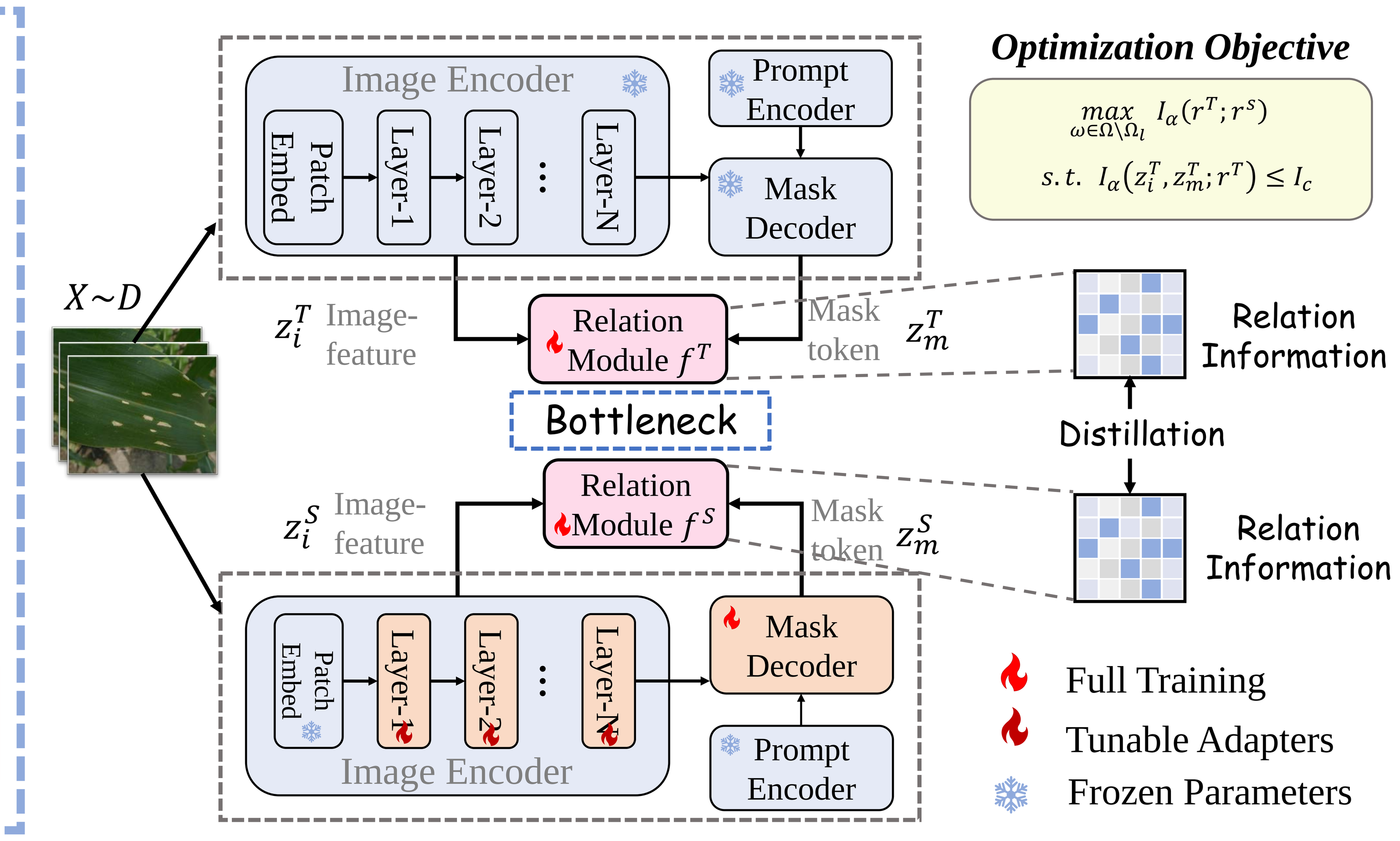
- To maximize the extracted information:

$$\max_{\omega} I_\alpha(r^T; r^S) \;\Rightarrow\; \mathcal{L}_d = log_2\|G_r^T\|_F^2 + log_2\|G_r^S\|_F^2 - log_2\|G_r^{TS}\|_F^2$$

- **The overall objectives:**

$$\max_{\omega} I_\alpha(r^T; r^S) - \beta I_\alpha(z_i^T, z_m^T; r^T) \;\Rightarrow\; \mathcal{L}_{info} = \lambda_1\mathcal{L}_r + \lambda_2\mathcal{L}_d$$

Overview of InfoSAM



Results

- **Comparison of PEFT methods on various segmentation tasks.**

Method	Natural Images			Medical				Agriculture		Remote Sensing	
	CAMO			ISIC 2017		Kvasir		Leaf		Road	
	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$	Jac ↑	Dice ↑	$S_\alpha \uparrow$	$E_\phi \uparrow$	IoU ↑	Dice ↑	IoU ↑	Dice ↑
SAM	$79.7_{\pm 0.02}$	$88.8_{\pm 0.09}$	$79.6_{\pm 0.01}$	$61.0_{\pm 0.12}$	$71.7_{\pm 0.14}$	$71.4_{\pm 0.16}$	$77.9_{\pm 0.17}$	$37.6_{\pm 0.11}$	$47.0_{\pm 0.16}$	$7.2_{\pm 0.24}$	$12.9_{\pm 0.29}$
decoder-only	$84.9_{\pm 0.38}$	$92.7_{\pm 0.34}$	$81.8_{\pm 0.33}$	$85.9_{\pm 0.34}$	$92.2_{\pm 0.20}$	$90.9_{\pm 0.05}$	$95.2_{\pm 0.18}$	$55.6_{\pm 1.12}$	$68.8_{\pm 1.17}$	$47.6_{\pm 0.47}$	$64.1_{\pm 0.47}$
BitFit	$87.5_{\pm 0.13}$	$94.5_{\pm 0.08}$	$85.3_{\pm 0.48}$	$87.7_{\pm 0.14}$	$93.2_{\pm 0.08}$	$92.5_{\pm 0.12}$	$96.3_{\pm 0.20}$	$69.2_{\pm 0.67}$	$80.3_{\pm 0.68}$	$58.1_{\pm 0.06}$	$73.1_{\pm 0.06}$
AdaptFormer	$87.9_{\pm 0.10}$	$94.8_{\pm 0.21}$	$86.2_{\pm 0.19}$	$87.6_{\pm 0.24}$	$93.2_{\pm 0.15}$	$93.3_{\pm 0.68}$	$97.0_{\pm 0.81}$	$75.0_{\pm 0.11}$	$84.8_{\pm 0.08}$	$61.1_{\pm 0.15}$	$75.5_{\pm 0.12}$
LoRA	$87.7_{\pm 0.59}$	$94.6_{\pm 0.50}$	$85.1_{\pm 0.64}$	$87.8_{\pm 0.24}$	$93.3_{\pm 0.13}$	$93.0_{\pm 0.14}$	$96.6_{\pm 0.11}$	$71.4_{\pm 0.54}$	$82.1_{\pm 0.62}$	$59.0_{\pm 0.19}$	$74.0_{\pm 0.17}$
Adapter	$88.2_{\pm 0.44}$	$94.8_{\pm 0.34}$	$86.7_{\pm 0.92}$	$87.7_{\pm 0.23}$	$93.2_{\pm 0.16}$	$93.4_{\pm 0.12}$	$97.1_{\pm 0.15}$	$74.4_{\pm 0.16}$	$84.3_{\pm 0.28}$	$60.5_{\pm 0.10}$	$75.1_{\pm 0.08}$
HQ-SAM	$85.1_{\pm 0.10}$	$92.6_{\pm 0.10}$	$81.0_{\pm 0.61}$	$86.3_{\pm 0.32}$	$92.4_{\pm 0.19}$	$91.1_{\pm 0.50}$	$95.5_{\pm 0.57}$	$66.2_{\pm 0.44}$	$77.8_{\pm 0.43}$	$54.9_{\pm 0.16}$	$70.6_{\pm 0.13}$
SU-SAM	$88.3_{\pm 0.21}$	$95.0_{\pm 0.22}$	$86.2_{\pm 0.59}$	$87.8_{\pm 0.18}$	$93.2_{\pm 0.09}$	$93.8_{\pm 0.02}$	$97.5_{\pm 0.06}$	$74.7_{\pm 0.53}$	$84.5_{\pm 0.56}$	$60.2_{\pm 0.26}$	$74.8_{\pm 0.22}$
ConvLoRA-SAM	$87.5_{\pm 0.39}$	$94.5_{\pm 0.17}$	$85.4_{\pm 0.41}$	$87.7_{\pm 0.22}$	$93.2_{\pm 0.11}$	$92.9_{\pm 0.13}$	$96.6_{\pm 0.28}$	$71.4_{\pm 0.44}$	$82.2_{\pm 0.37}$	$59.6_{\pm 0.22}$	$74.4_{\pm 0.20}$
LoRA+Ours	$\mathbf{88.3}_{\pm 0.05}$	$\mathbf{95.2}_{\pm 0.00}$	$\mathbf{85.8}_{\pm 0.59}$	$\mathbf{88.1}_{\pm 0.08}$	$\mathbf{93.5}_{\pm 0.05}$	$\mathbf{93.4}_{\pm 0.11}$	$\mathbf{96.8}_{\pm 0.09}$	$\mathbf{72.2}_{\pm 0.06}$	$\mathbf{82.8}_{\pm 0.04}$	$\mathbf{59.9}_{\pm 0.20}$	$\mathbf{74.6}_{\pm 0.17}$
Adapter+Ours	$\mathbf{88.6}_{\pm 0.09}$	$\mathbf{95.1}_{\pm 0.05}$	$\mathbf{87.1}_{\pm 0.37}$	$\mathbf{88.0}_{\pm 0.05}$	$\mathbf{93.4}_{\pm 0.00}$	$\mathbf{94.4}_{\pm 0.12}$	$\mathbf{97.9}_{\pm 0.09}$	$\mathbf{75.6}_{\pm 0.27}$	$\mathbf{85.2}_{\pm 0.23}$	$\mathbf{61.4}_{\pm 0.30}$	$\mathbf{75.8}_{\pm 0.27}$

- **Ablation study**

➢ Ablation study results of two losses

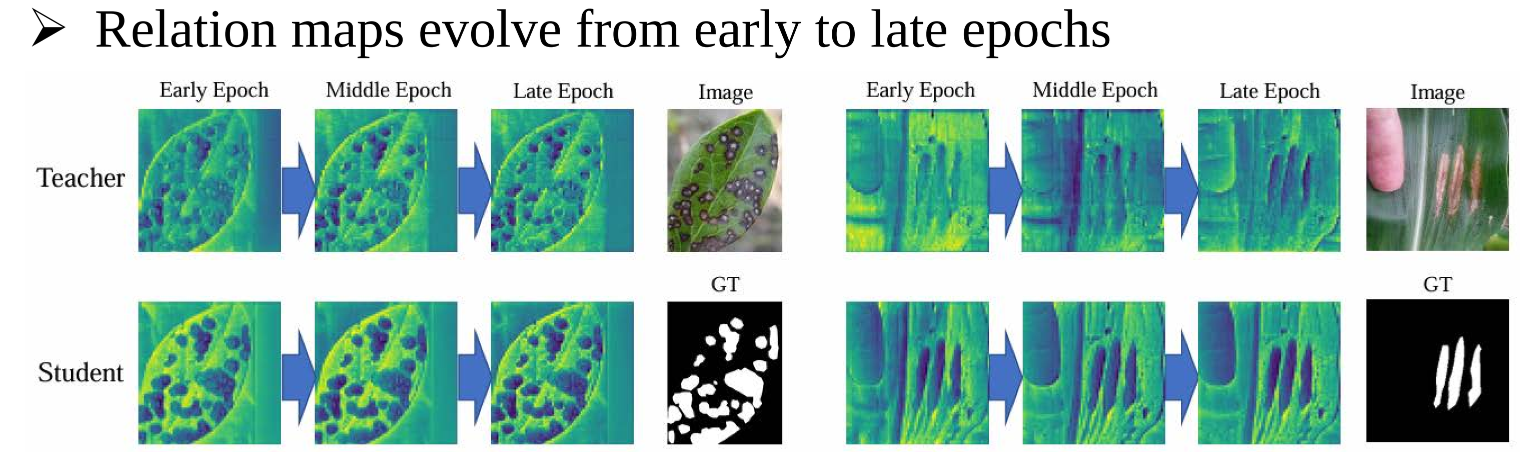
L_r	L_d	Medical	Agriculture	Remote Sensing
		S_α (Kvasir)	IoU (Leaf)	IoU (Road)
		93.4	74.4	60.5
	✓	93.6 (+0.2)	75.2 (+0.8)	61.0 (+0.5)
✓	✓	94.4 (+1.0)	75.6 (+1.2)	61.4 (+0.9)

➢ Transferability of the Relation Module

Method	Frozen RM from	Medical S_α (Kvasir)	Agriculture IoU (Leaf)
InfoSAM	-	$94.4_{\pm 0.12}$	-
InfoSAM-T	Leaf	$93.7_{\pm 0.24}$	-
InfoSAM	-	-	$75.6_{\pm 0.30}$
InfoSAM-T	Kvasir	-	$75.4_{\pm 0.45}$

- **Visualization results**

➢ Relation maps evolve from early to late epochs



➢ Visualization results on camouflaged object segmentation