Mutual Learning for SAM Adaptation: A Dual Collaborative Network Framework for Source-Free Domain Transfer

Yabo Liu, Waikeung Wong, Chengliang Liu, Xiaoling Luo, Yong Xu, Jinghua Wang

Harbin Institute of Technology, Shenzhen
Laboratory for Artifcial Intelligence in Design, Hong Kong
School of Fashion and Textiles, Hong Kong Polytechnic University
Shenzhen University

Domain Adaptive Segmentation

CAMO

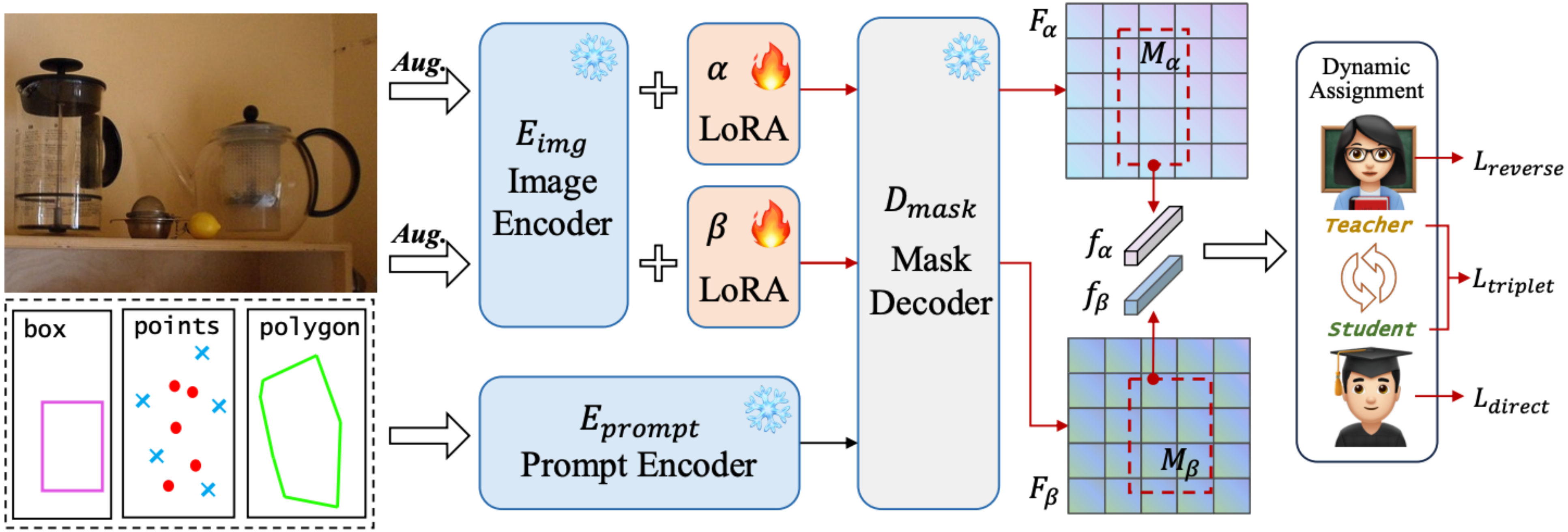
COCO

Kvasir-SEG

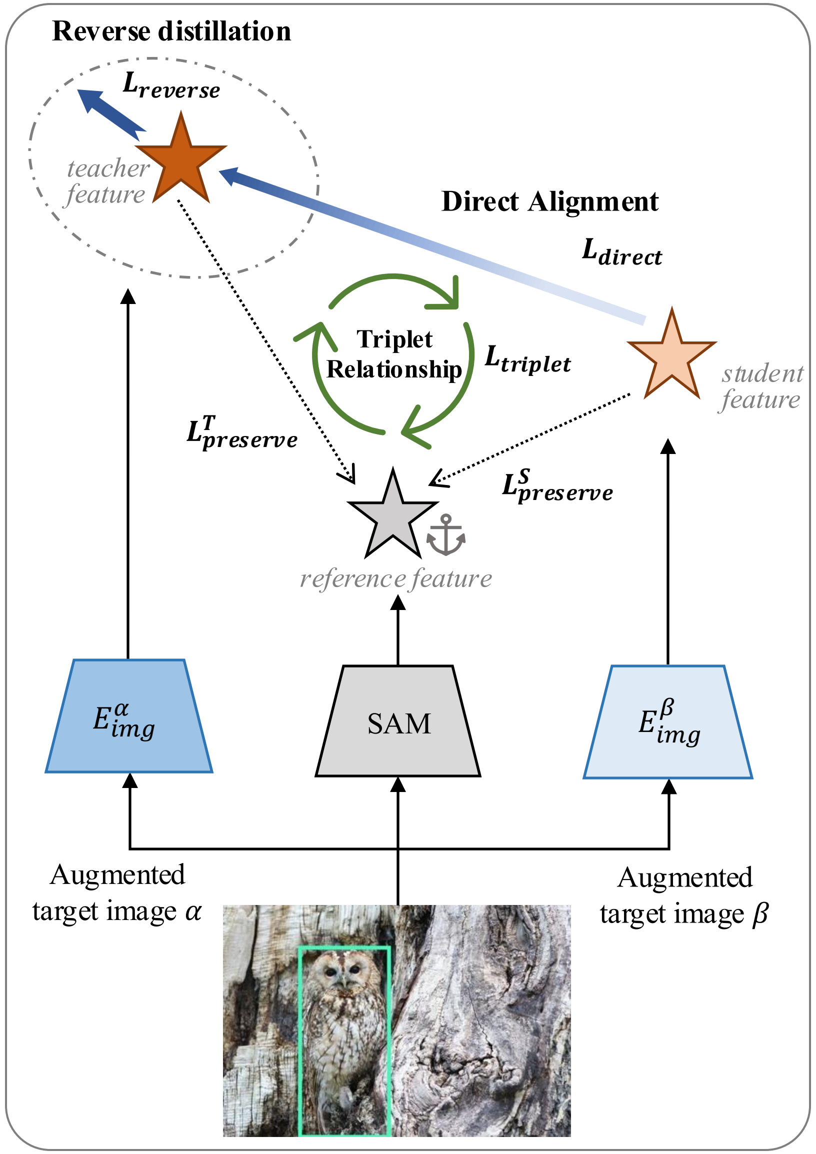
COD10K

Method

1. Add two LoRA modules after the original SAM image encoder to construct lightweight trainable collaborative image encoders α and β. $E_{img}^{\alpha/\beta} = \mathbf{perturb}(E_{img}) + \triangle E_{img}^{\alpha/\beta}$
2. Input the augmented target domain images into the collaborative image encoders to obtain predicted image features.
3. Compare the image features output by the collaborative network with those from the original SAM, and assign the roles of **"Teacher" and "Student"** networks based on the degree of knowledge retention.
4. Use **Direct Alignment Loss** to reduce the student's prediction errors and employ **Reverse Distillation Loss** to enhance the teacher's feature diversity.
5. Select the collaborative network with better performance as the target domain model.



Method



Foreground Representations

Extract foreground representations from the image feature map based on the predicted pseudo-labels.

$$f_{\alpha/\beta/\gamma} = \frac{\sum_{i,j}^{h,w} F^{i,j}_{\alpha/\beta/\gamma} \cdot \mathbf{1}(S^{i,j}_{\alpha/\beta/\gamma} > \mathcal{T})}{\sum_{i,j}^{h,w} \mathbf{1}(S^{i,j}_{\alpha/\beta/\gamma} > \mathcal{T})}$$

Role Assignment

Assign roles based on the similarity between the outputs of the collaborative network and the original network.

$$sim_{\alpha} = \frac{f_{\alpha} \cdot f_{\gamma}}{\|f_{\alpha}\| \|f_{\gamma}\|}, \quad sim_{\beta} = \frac{f_{\beta} \cdot f_{\gamma}}{\|f_{\beta}\| \|f_{\gamma}\|}. \qquad (\mathcal{T}, \mathcal{S}) = \begin{cases} (\alpha, \beta) & \text{if } sim_{\alpha} > sim_{\beta} \\ (\beta, \alpha) & \text{otherwise.} \end{cases}$$

Direct alignment loss

Align the student network's predicted feature map with the teacher network to reduce the prediction gap between them.

Reverse distillation loss

Push the teacher network's predicted feature map away from the student network's results to increase the prediction gap between them.

Quantitative Comparison

Table 1. Comparison Results on COCO and Pascal VOC. Source and Target denote the models trained with source domain data and target domain data, respectively. WeSAM* denotes reproduced results of WeSAM

Method	COCO					Pascal VOC				
	box	point	poly	Average	GAIN	box	point	poly	Average	GAIN
Source	74.29	55.06	65.64	65.00	-	69.21	69.21	60.79	66.40	-
Target	81.50	69.77	73.39	74.89	9.90	81.23	76.98	71.32	76.51	10.11
TENT (Wang et al., 2020)	78.21	52.99	71.51	67.57	2.58	80.24	74.97	65.03	73.41	7.01
SHOT (Liang et al., 2021)	75.18	58.46	69.26	67.63	2.64	79.80	74.26	63.38	72.48	6.08
soft Teacher (Xu et al., 2021)	75.94	43.36	68.27	62.52	-2.47	72.93	56.09	62.20	63.74	-2.66
TRIBE (Su et al., 2024)	77.56	49.56	70.99	66.04	1.05	78.87	69.21	65.39	71.16	4.76
DePT (Gao et al., 2022)	71.00	37.35	63.27	57.21	-7.78	74.09	42.99	59.94	59.01	-7.39
WDASS (Das et al., 2023)	77.29	60.55	70.19	69.34	4.35	80.12	76.15	66.98	74.42	8.02
WeSAM* (Zhang et al., 2024)	77.32	60.50	70.77	69.53	4.54	80.27	74.15	66.72	73.71	7.31
ours	**78.97**	**63.00**	**72.54**	**71.50**	**6.51**	**82.90**	**76.24**	**70.20**	**76.45**	**10.05**

Table 2. Comparison Results on CAMO and COD10K. Source and Target denote the models trained with source domain data and target domain data, respectively.

Method	CAMO					COD10K				
	box	point	poly	Average	GAIN	box	point	poly	Average	GAIN
Source	62.72	57.43	50.85	57.00	-	66.32	63.61	40.04	56.66	-
Target	79.17	77.01	67.12	74.43	17.43	78.06	78.44	64.90	73.80	17.15
TENT (Wang et al., 2020)	71.24	59.59	60.29	63.71	6.71	69.36	61.94	43.36	58.22	1.57
SHOT (Liang et al., 2021)	71.61	62.78	58.72	64.37	7.37	69.09	65.25	42.38	58.91	2.26
soft Teacher (Xu et al., 2021)	62.30	48.64	51.26	54.07	-2.93	66.32	50.04	32.27	49.54	-7.11
TRIBE (Su et al., 2024)	66.00	61.97	60.54	62.84	5.84	67.84	63.62	42.75	58.07	1.42
DePT (Gao et al., 2022)	55.44	33.07	48.63	45.71	-11.29	59.32	34.06	35.51	42.96	-13.69
WDASS (Das et al., 2023)	71.25	63.39	62.29	65.64	8.64	71.42	65.61	43.93	60.32	3.67
WeSAM (Zhang et al., 2024)	73.42	65.55	62.90	67.29	10.29	71.93	70.55	45.87	62.78	6.13
ours	**74.46**	**70.21**	**67.54**	**70.74**	**13.74**	**73.89**	**72.83**	**47.27**	**64.66**	**8.01**

Quantitative Comparison

Table 3. Comparison Results on kvasir-SEG. Source and Target denote the models trained with source domain data and target domain data, respectively.

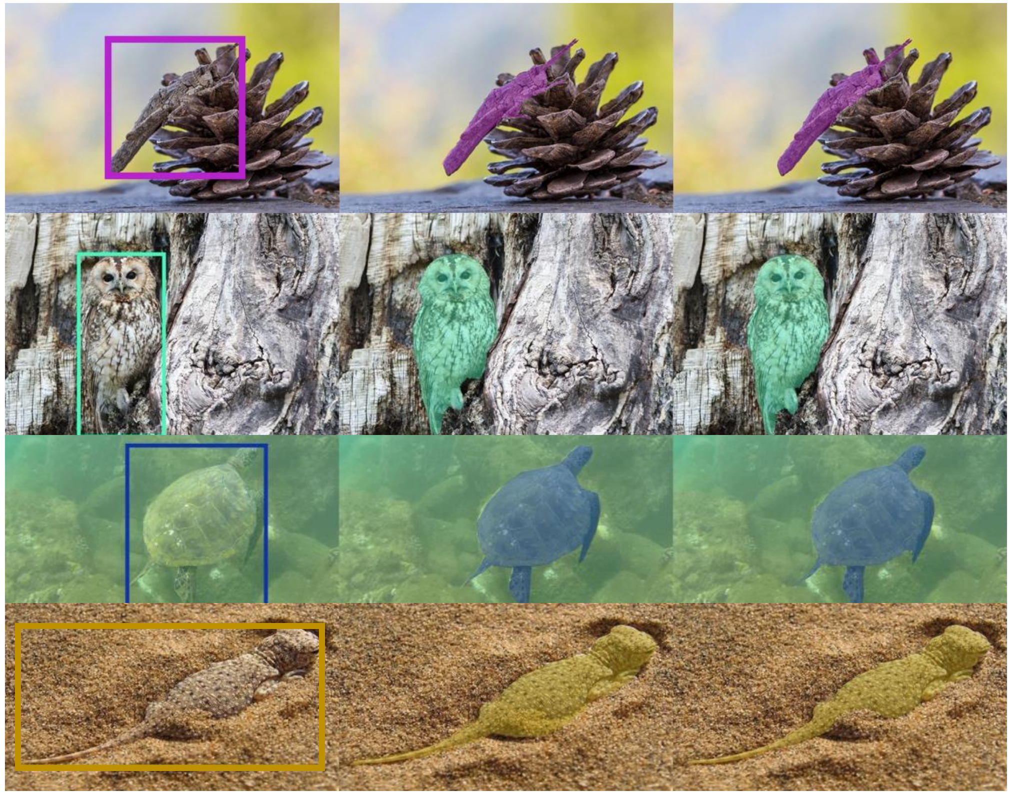
Method	kvasir-SEG				
	box	point	poly	Average	GAIN
Source	81.59	62.30	54.03	65.97	-
Target	85.89	77.54	81.64	81.69	15.72
TENT	82.47	61.84	62.97	69.09	3.12
SHOT	82.30	63.76	61.34	69.13	3.16
soft Teacher	84.12	73.53	58.15	71.93	5.96
TRIBE	85.05	73.03	64.61	74.23	8.26
DePT	81.91	52.06	61.55	65.17	-0.80
WDASS	84.01	63.78	64.78	70.86	4.89
WeSAM	85.47	75.23	67.40	76.03	10.06
ours	**86.92**	**76.18**	**86.78**	**83.29**	**17.32**

Table 4. Comparison Results on OCID. Source and Target denote the models trained with source domain data and target domain data, respectively.

Method	OCID				
	box	point	poly	Average	GAIN
Source	86.35	71.41	72.81	76.86	-
Target	91.24	89.22	79.23	86.56	9.71
TENT	87.77	66.61	77.53	77.30	0.45
SHOT	88.06	74.39	76.25	79.57	2.71
soft Teacher	84.98	68.46	73.75	75.73	-1.13
TRIBE	86.77	67.86	76.50	77.04	0.19
DePT	82.00	56.52	70.92	69.81	-7.04
WDASS	87.68	77.13	76.70	80.50	3.65
WeSAM	**88.09**	**80.14**	77.41	81.88	5.02
ours	88.07	77.33	**86.66**	**84.02**	**7.16**

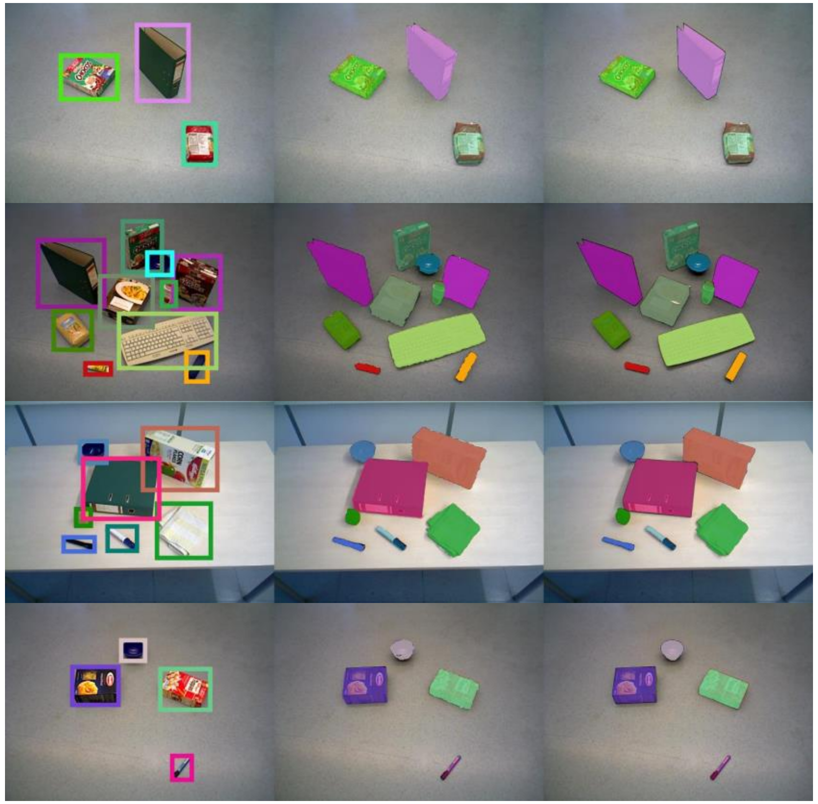
Qualitative Results

Image and box prompt Ground truth Ours



Segmentation results on the CAMO target domain

Image and box prompt Ground truth Ours

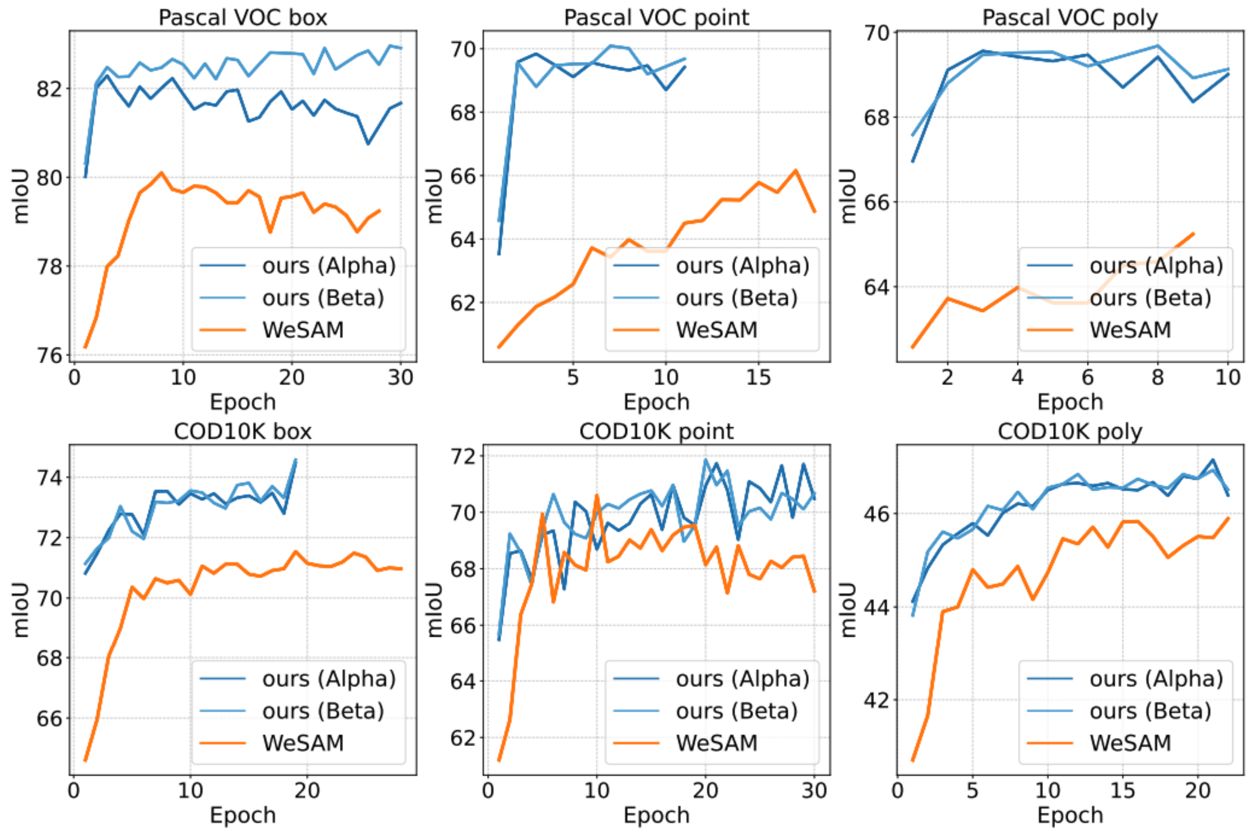


Segmentation results on the OCID target domain

Performance comparison

Performance comparison of our proposed method with different loss components.

Direct	Reverse	Triplet	CAMO	COD10K
✓			68.83	63.21
✓	✓		70.22	64.13
✓		✓	69.75	63.75
✓	✓	✓	70.74	64.66



Segmentation performance curves of our method and WeSAM. The experiments were conducted in different prompt forms on two target domain datasets. The orange, blue, and light blue curves represent the performances of WeSAM, SAMα, and SAMα in our method, respectively

Thanks