

# DC-TTA: Divide-and-Conquer Framework for Test-Time Adaptation of Interactive Segmentation

## Supplementary Material

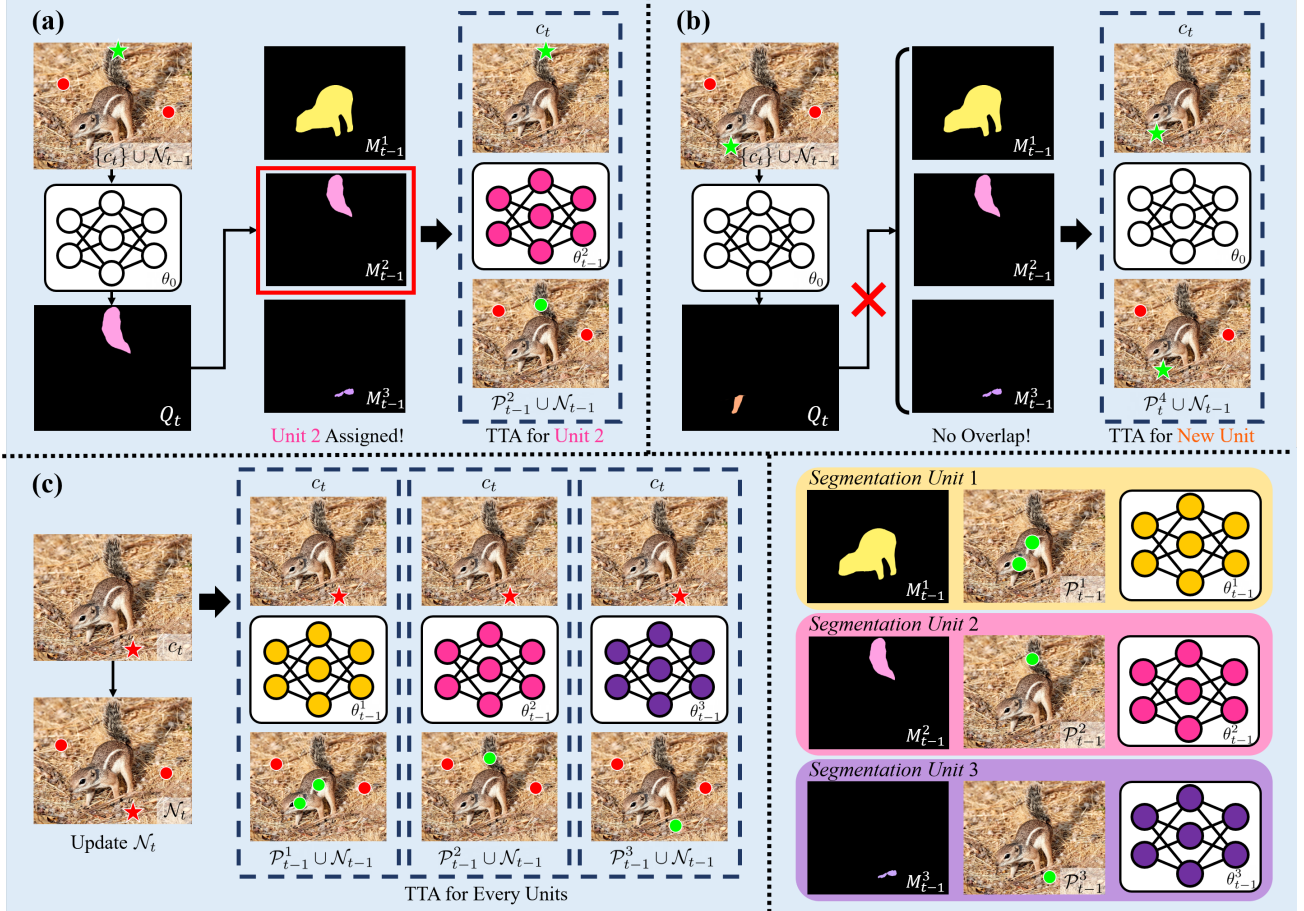


Figure A1. Detailed process of the click assignment. The possible cases are as follows: (a) The new click is positive and assigned to an existing unit; (b) The new click is positive and assigned to a newly generated unit; (c) The new click is negative. The bottom right shows the composition of each segmentation unit up to the  $t - 1$  iterations.

### 1. Details on Click Assignment Process

We provide a detailed procedure for the click assignment process in Fig. A1. We illustrate three cases for the newly selected click, labeled as (a), (b), and (c), by showing possible examples. In case (a), where  $c_t$  is positive and  $\text{IoU}(Q_t, M_{t-1}^2)$  exceeds 0,  $c_t$  is assigned to segmentation unit 2 and TTA process begins. In case (b), where  $c_t$  is positive but  $\text{IoU}(Q_t, M_{t-1}^k)$  does not exceed 0 for  $k = 1, 2, 3$ ,  $c_t$  is assigned to new segmentation unit 4 with original SAM parameters and TTA process begins. If  $\mathcal{N}_{t-1}$  is  $\emptyset$ , meaning that no negative click has been entered, the TTA process will not be performed in case (b). In the last case (c), where  $c_t$  is negative, we update  $\mathcal{N}_t$  by adding  $c_t$ , and the TTA process is performed for all segmentation units. For all cases,

the first click is positive, and the first segmentation unit is assigned from this click, with no TTA process performed.

### 2. Experiments on Other SAM Checkpoints

To assess the effectiveness across different SAM checkpoints, we conduct an additional experiment using ViT-L and ViT-H backbones. We compute the NoC and FR metrics with 20 clicks and a 90 threshold IoU on the CAMO [21] and TRASHCAN [14] datasets, as shown in Table A1. Our results show that DC-TTA significantly improves performance across all checkpoints for both datasets. This indicates that our method is effective with various backbone architectures in SAM.

Table A1. Quantitative results on different SAM checkpoints. We conduct an additional experiment using ViT-L and ViT-H backbones with a 20-click setting. Our DC-TTA consistently shows improvements.

Methods	Model	CAMO [21]		TRASHCAN [14]	
		NoC90	FR90	NoC90	FR90
Baseline	ViT-B	10.45	31.20	13.30	52.29
Ours		<b>9.58</b>	<b>25.60</b>	<b>12.05</b>	<b>38.80</b>
Baseline	ViT-L	9.16	27.20	13.08	48.18
Ours		<b>8.42</b>	<b>18.40</b>	<b>11.70</b>	<b>32.81</b>
Baseline	ViT-H	8.39	25.20	12.81	47.98
Ours		<b>7.93</b>	<b>20.00</b>	<b>11.88</b>	<b>36.45</b>

Table A2. Performance comparison of NoC and FR at 95% IoU with 30 clicks. Our DC-TTA achieves lower NoC and FR values compared to the baselines and other TTA methods.

Method	CAMO [21]		COD10k [9]	
	NoC95	FR95	NoC95	FR95
Baseline [19]	24.87	74.00	25.46	80.21
AdaptSAM [41]	24.71	73.60	25.14	77.99
DC-only	24.53	72.40	25.09	77.84
Ours	<b>23.59</b>	<b>64.80</b>	<b>24.04</b>	<b>71.03</b>

Method	TRASHCAN [14]		ISTD [50]	
	NoC95	FR95	NoC95	FR95
Baseline [19]	27.13	85.58	21.48	61.07
AdaptSAM [41]	26.79	82.60	21.05	57.54
DC-only	27.05	84.93	20.71	56.20
Ours	<b>25.51</b>	<b>73.86</b>	<b>19.71</b>	<b>50.80</b>

### 3. Experiments with 30 Clicks and 95 IoU

To further validate the effectiveness of our method, we conduct an experiment with 30 clicks and a 95 threshold IoU. Table A2 presents the results on the CAMO [21], COD10k [9], TRASHCAN [14], and ISTD [50] datasets. For both the NoC95 and FR95 metrics, our DC-only method improves performance compared to the baseline SAM model. Additionally, our DC-TTA outperforms other TTA methods across all datasets. These results demonstrate that our DC strategy and DC-TTA framework provide significant and robust benefits across different settings.

### 4. Hyperparameter Ablation Studies

We verify the effectiveness of the hyperparameter  $\gamma$ , as illustrated in Sec. 3.4.4. We measure the NoC90 with 20 clicks while varying the  $\gamma$  value, as shown in Fig. A2. We observe that NoC90 decreases as  $\gamma$  increases from 0.3 to 0.7, and subsequently increases as  $\gamma$  rises from 0.7 to 1.5. Based on this result, we select  $\gamma = 0.7$  for the main experiments. Note that, regardless of the value of  $\gamma$  in our

Table A3. Maximum K and Seconds Per Click (SPC) of each dataset.

Dataset	CAMO	COD10k	TRASHCAN	ISTD
max K / SPC(s)	9 / 0.275	11 / 0.307	9 / 0.243	13 / 0.338

Dataset	Berkeley	DAVIS	COCOMVal	PascalVOC
max K / SPC(s)	3 / 0.141	8 / 0.275	9 / 0.228	10 / 0.204

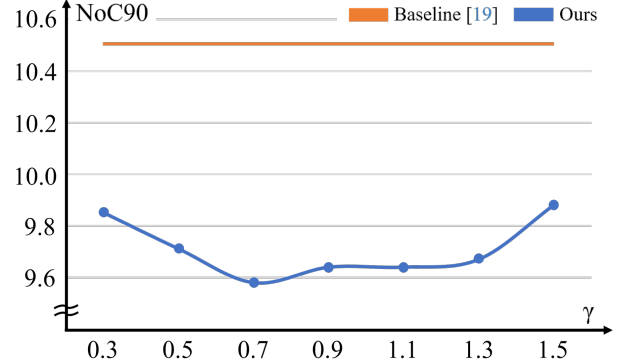


Figure A2. Verification of hyperparameter  $\gamma$ 's effectiveness on CAMO [21]. We choose  $\gamma = 0.7$  as it is the optimal point.

searching range, the proposed DC-TTA robustly outperforms baseline SAM [19].

### 5. Computational Analysis

To evaluate the efficiency of DC-TTA in interactive segmentation, we analyze its computational and memory overhead. Each adaptive unit updates only a lightweight prompt encoder and decoder, introducing approximately 4M additional parameters per unit—relatively minor compared to the frozen image encoder ( $\sim 90$ M). The average number of gradient updates per click is 3.33, and memory usage remains under 6GB throughout, enabling use even on a local GPU. We also measure the runtime responsiveness of DC-TTA using the Seconds Per Click (SPC) metric across multiple datasets. As shown in Table A3, SPC remains around 0.3 seconds in most cases, suggesting that the method maintains practical responsiveness despite the added adaptation steps. On CAMO [21], even for samples with the highest  $K$ , the SPC is only 0.378 seconds. These observations indicate that the computational overhead introduced by DC-TTA remains manageable, and the method can be applied efficiently in real-world interactive segmentation scenarios.

### 6. More Qualitative Results

We provide additional qualitative results Fig. A3. We observe that our methods successfully predict output mask with smaller amount of clicks compared to baseline and other TTA methods. In addition, we provide additional examples of how masks are divided into segmentation units in Fig. A4. These results strongly support the superiority of our method and design intention.

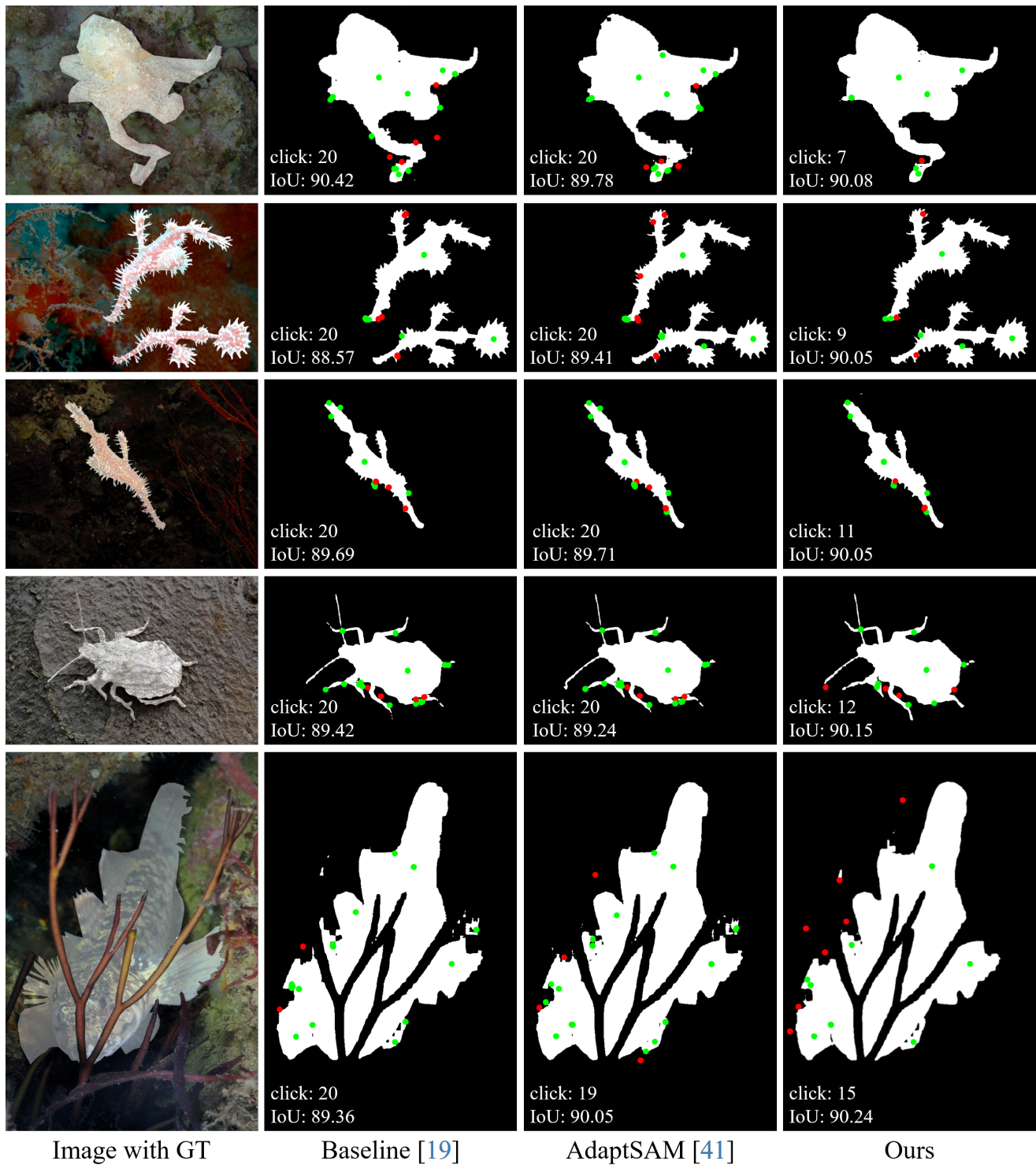


Figure A3. Additional qualitative results of challenging camouflaged objects.



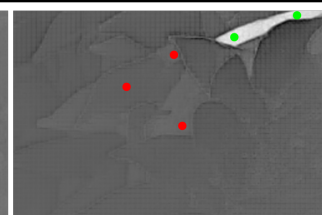
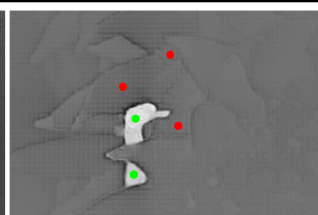
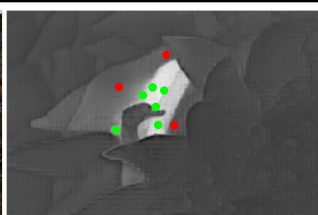
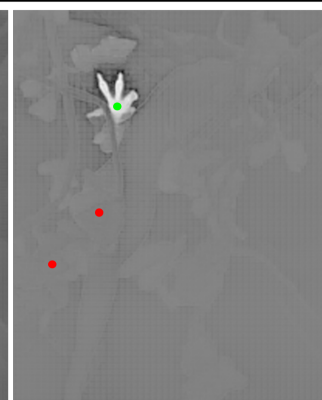
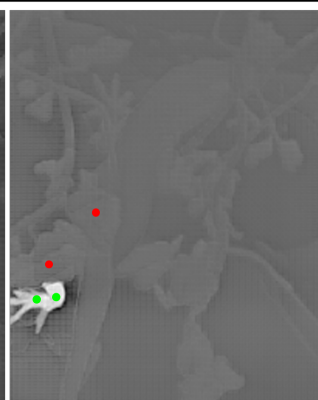
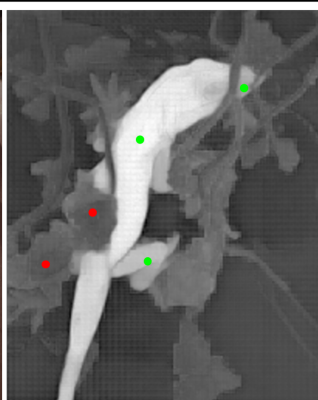
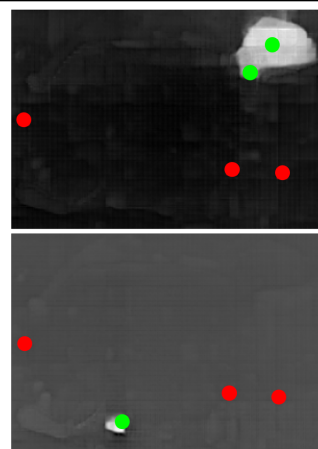
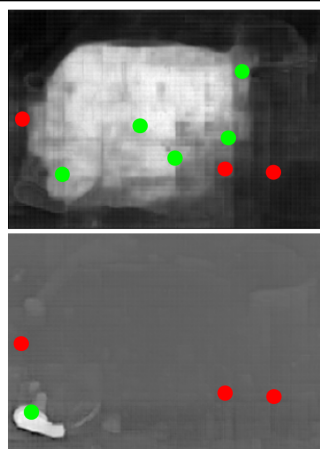
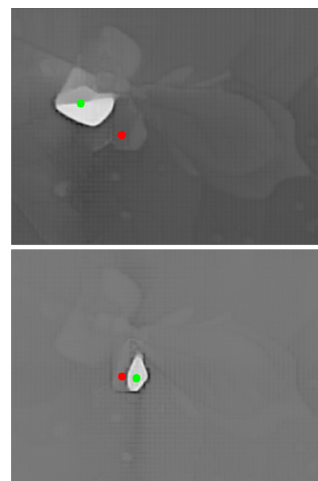
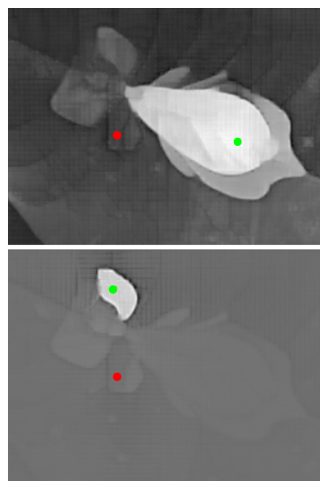


Image with GT

Unit Partitions

Figure A4. Additional examples of segmentation units identified in our DC-TTA.