CE-OST: Contour Emphasis for One-Stage Transformer-based Camouflage Instance Segmentation

Thanh-Danh Nguyen[1,2], Duc-Tuan Luu[1,2], Vinh-Tiep Nguyen[†1,2], and Thanh Duc Ngo[1,2]
[1]University of Information Technology, Ho Chi Minh City, Vietnam, [2]Vietnam National University, Ho Chi Minh City, Vietnam. († corresponding author)

Introduction

Overview: Camouflage is a defense mechanism that animals use to conceal their appearance by blending in with their environment. Autonomously detecting camouflaged animals is helpful in various fields of computer vision: search-and-rescue mission; wild species discovery and preservation activities; media forensics, etc.

Motivation: Camouflaged instances have their **colors and textures similar to the background** while the **contour features can be recognized easier**.

Main contribution: CE-OST framework which employs the well-known architecture of Transformer-based models in a one-stage manner to boost the performance of camouflaged instance segmentation via contour emphasis machenism.

Method

Our **CE-OST** framework is enhanced with the two main components (Fig. 1). A camouflaged input image goes through the two blocks before meeting the Fusion Module at the end of the framework to return the segmentation mask.

1. Transformer Block:

- Inspired by OSFormer [30] whose design is specific for the task of camouflage instance segmentation.
- Features are extracted at **multi-scale** via an Encoder-Decoder model to get the **global information** along with the **precise location.**

2. Contour Emphasis Block:

- Based on Holistically-nested edge detection [22], CE-Block seeks out and enhance the contours of the camouflaged instances.
- A **Grid-Condition** is introduced to decide whether to use **Color Contrast** or **Brightness Addition** to apply the enhancement (Fig. 2).
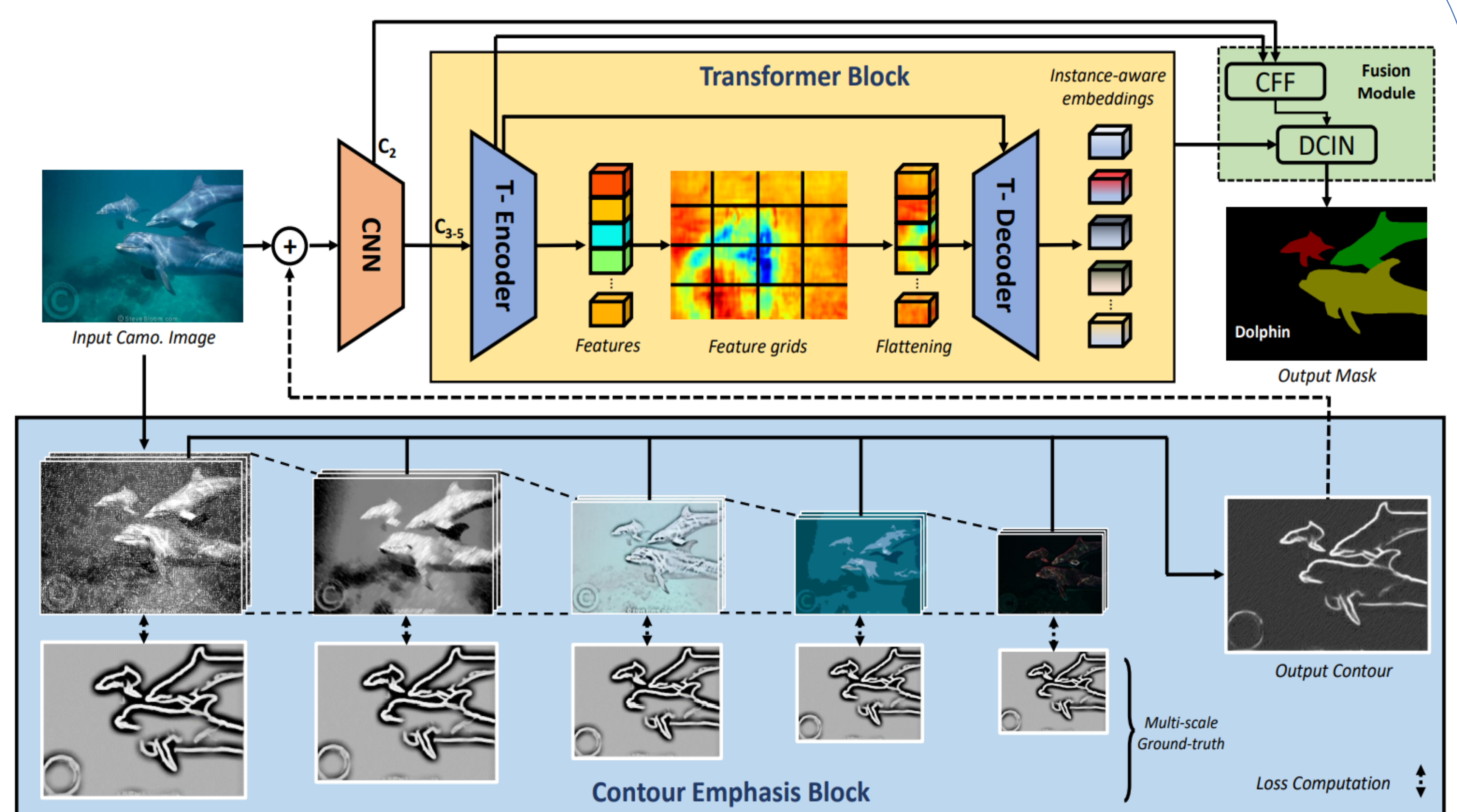

Fig. 1 Overall our CE-OST framework: Contour Emphasis for One-Stage Transformer-based Camouflage Instance Segmentation
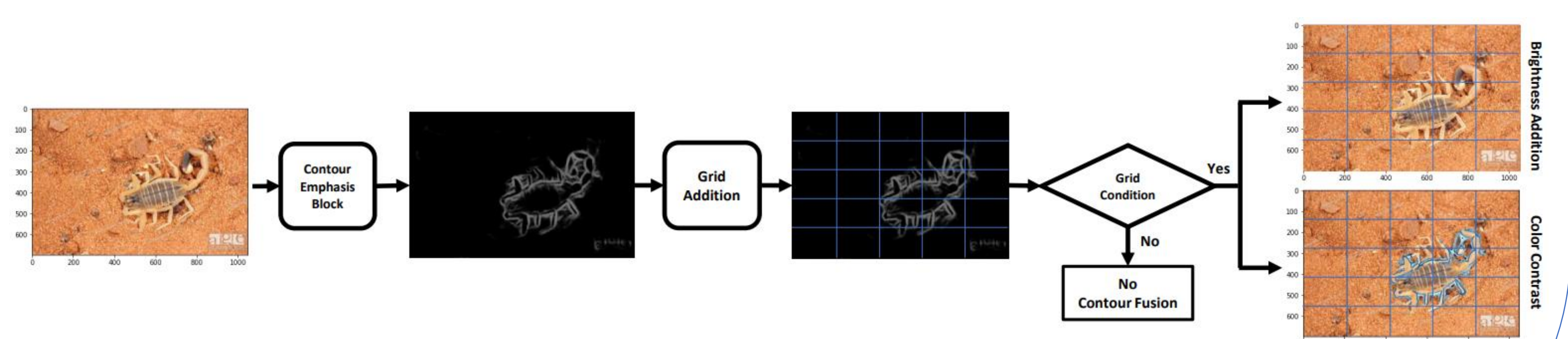

Fig. 2 Our Grid-Condition to control the Contour Emphasis. We apply a grid 5×5 to each image and decide not to apply boundary fusion if the number of eliminated cells is over half. A cell is eliminated if the number of pixels in the detected boundary area covers over a half area of the cell. This proposal is adaptive to every single image size of the camouflage dataset.

Results

The visualization and the state-of-the-art comparison is presented in Fig. 3 and Tab. 1 below. The comparison in Grid-Condition is also presented in Fig. 4. Please refer to the manuscript for more explanation.
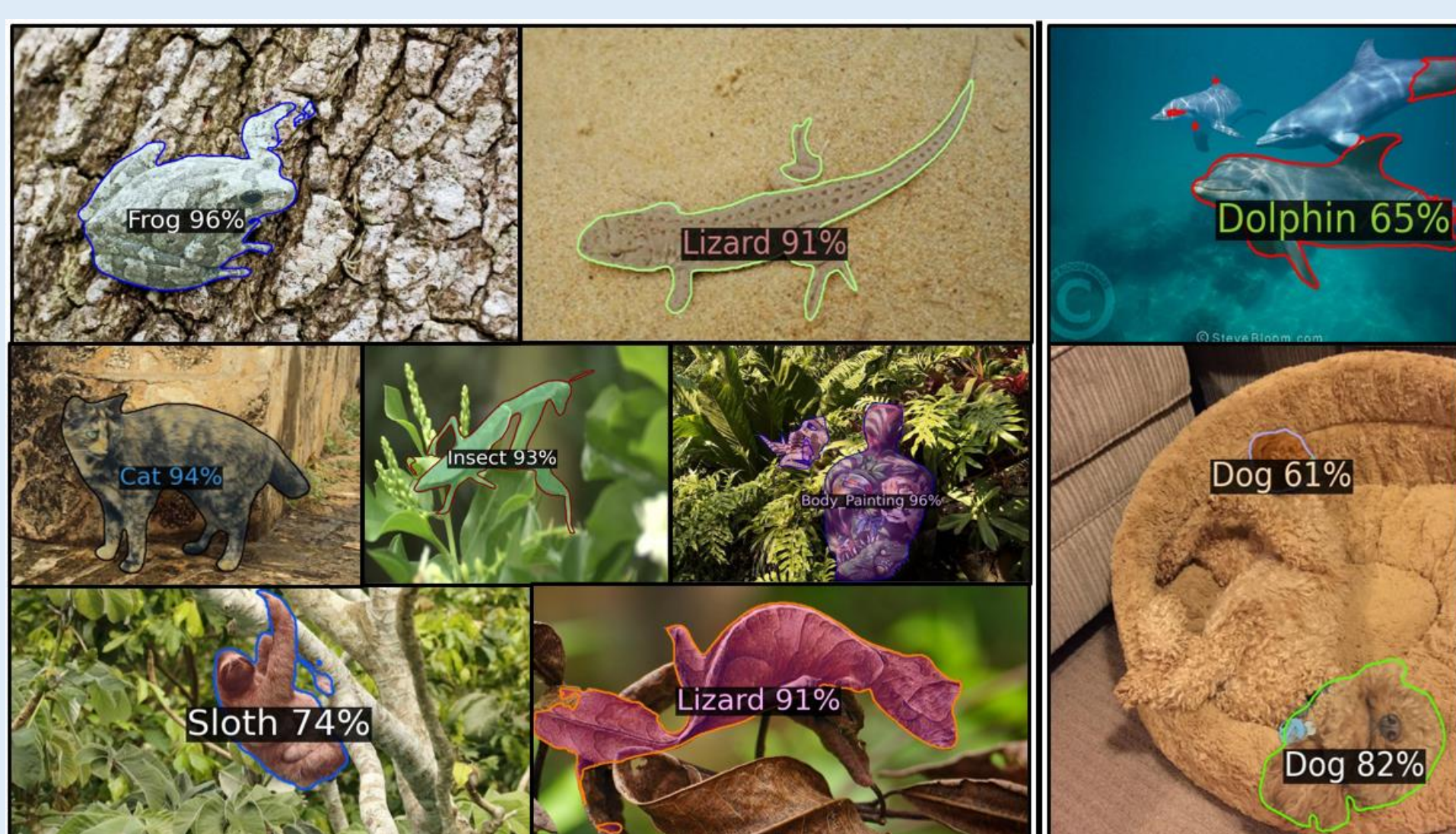

Fig. 3 Visualization results on the CAMO++ [31] testing set on our **CE-OST-PVT**. The confidence threshold is 0.5. Good and bad cases are left-right separated.

	Method	COD10K			NC4K		
		AP	AP50	AP75	AP	AP50	AP75
Two-Stage	Mask R-CNN [41]	28.7	60.1	25.7	36.1	68.9	33.5
	MS R-CNN [43]	33.3	61.0	32.9	35.7	63.4	34.7
	Cascade R-CNN [44]	29.5	61.0	25.9	34.6	66.3	31.5
	HTC [47]	30.9	61.0	28.7	34.2	64.5	31.6
	BlendMask [46]	31.2	60.0	28.9	31.4	61.2	28.8
	Mask Transfiner [64]	31.2	60.7	29.8	34.0	63.1	32.6
One-Stage	YOLACT [49]	29.0	60.1	25.3	37.8	70.6	35.6
	CondInst [65]	34.3	67.9	31.6	38.0	71.1	35.6
	QueryInst [66]	32.5	65.1	28.6	38.7	72.1	37.6
	SOTR [67]	32.0	63.6	29.2	34.3	65.7	32.4
	SOLOv2 [51]	35.2	65.7	33.4	37.8	69.2	36.1
	OSFormer [30]	42.0	71.3	42.8	44.4	73.7	45.1
	CE-OST (Ours)	**43.2**	**72.2**	**44.1**	**45.1**	**74.0**	**46.4**

Tab. 1 State-of-the-art comparison on COD10K [19] and NC4K [32] dataset. The chosen backbone is the common ResNet-101 [59].
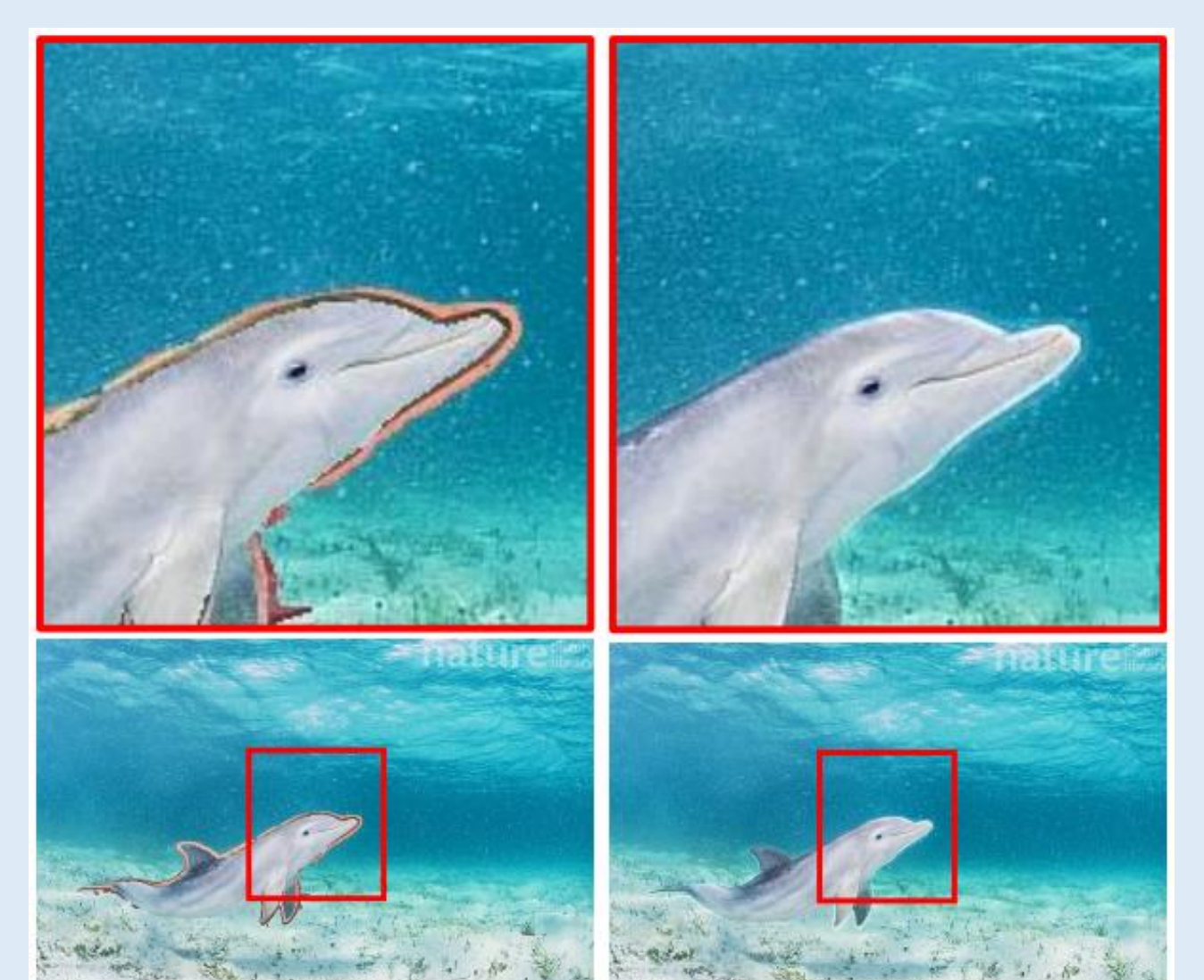

Fig. 4 Color Contrast (left) and Brightness Addition (right) comparison.