ORGANIZERS
VAPR

MAPR 2023
THE 6TH INTERNATIONAL CONFERENCE ON MULTIMEDIA ANALYSIS AND PATTERN RECOGNITION

SPONSORS
IEEE FPT EARTH VENTURE foundation

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,

{danhnt, tuanld, tiepnv, thanhnd}@uit.edu.vn, †corresponding author

Quy Nhon, October 5-6th, 2023

Content

1. Introduction to Camouflage Instance Segmentation
2. Related work
3. Our proposed CE-OST framework
   - Contour Emphasis Block
   - One-stage Transformer-based CIS
4. Experiments
5. Conclusion

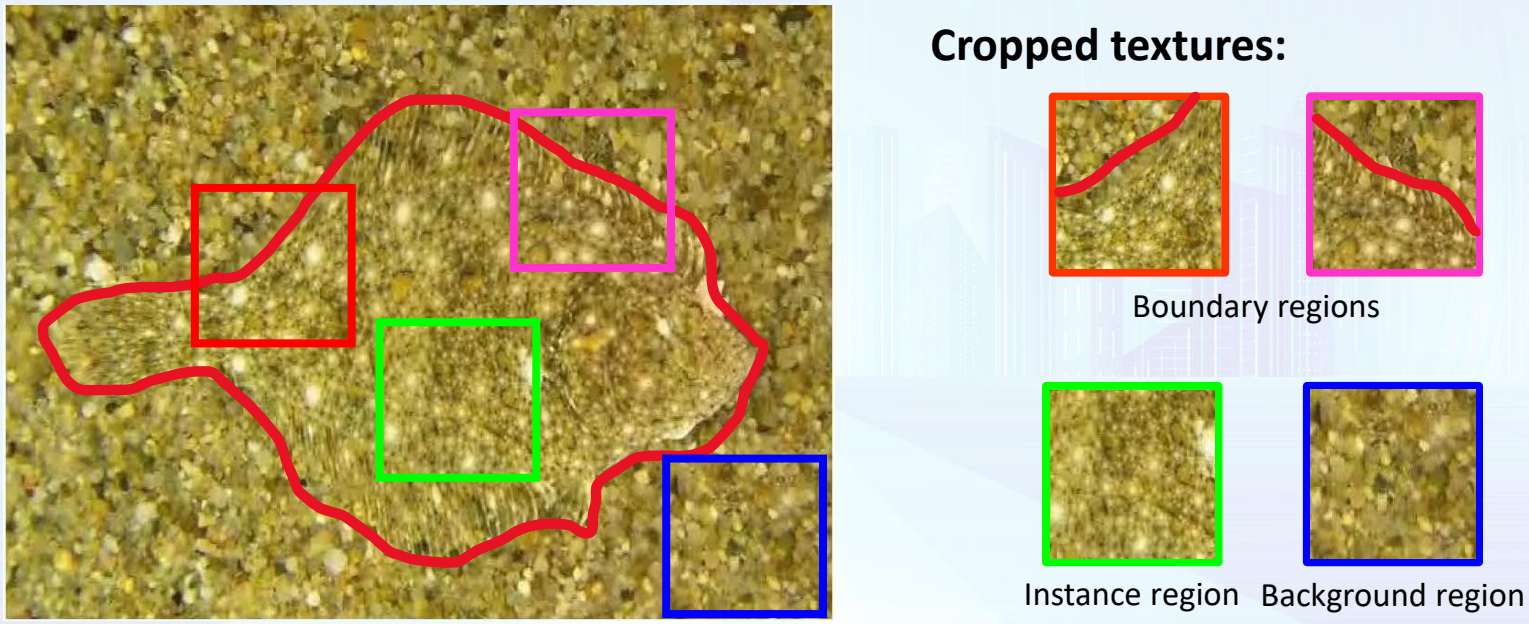
1. Introduction

- **"Camouflage"** is **a defense mechanism** that animals use to **conceal their appearance** by blending in with their environment
- **Applications:** search-and-rescue work, wild species discovery and preservation, medical diagnostic, etc.

Exemplary camouflaged samples extracted from CAMO++ dataset

1. Introduction

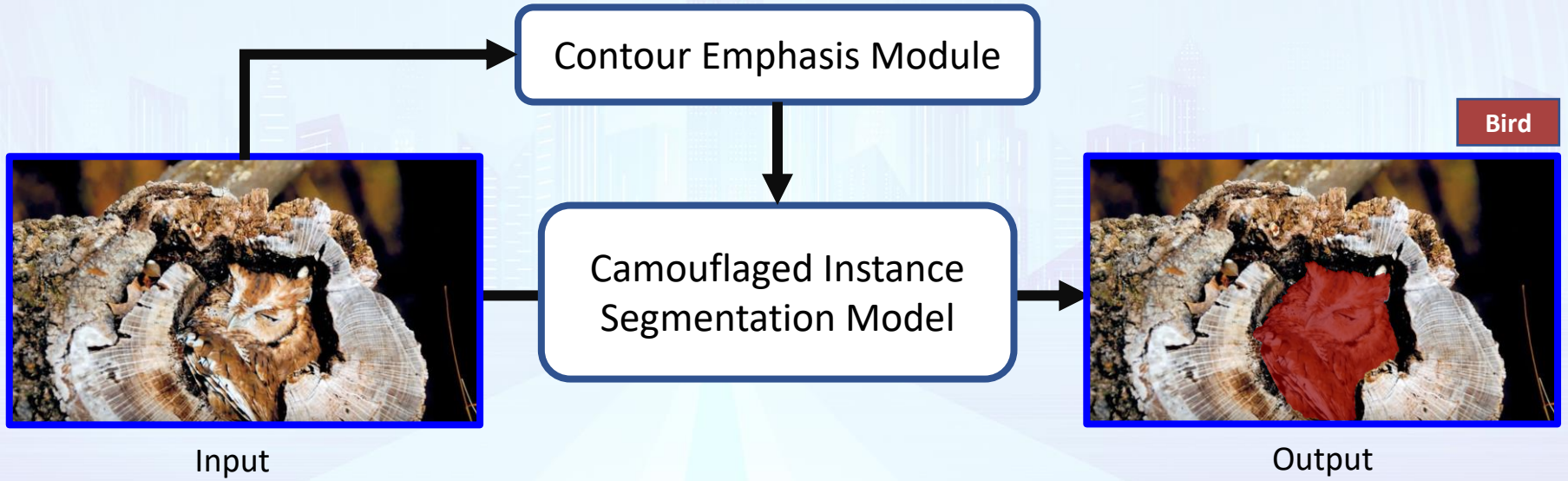
Focused challenge: Camouflaged instances are “visually transparent”, their colors and textures are similar to the background



Visual texture comparison among regions in a camouflaged image

1. Introduction

Contribution: we propose **Contour Emphasis** approach for **One-Stage Transformer-based Camouflage Instance Segmentation**, dubbed **CE-OST**



General idea of our contour emphasis framework contribution

2. Related work

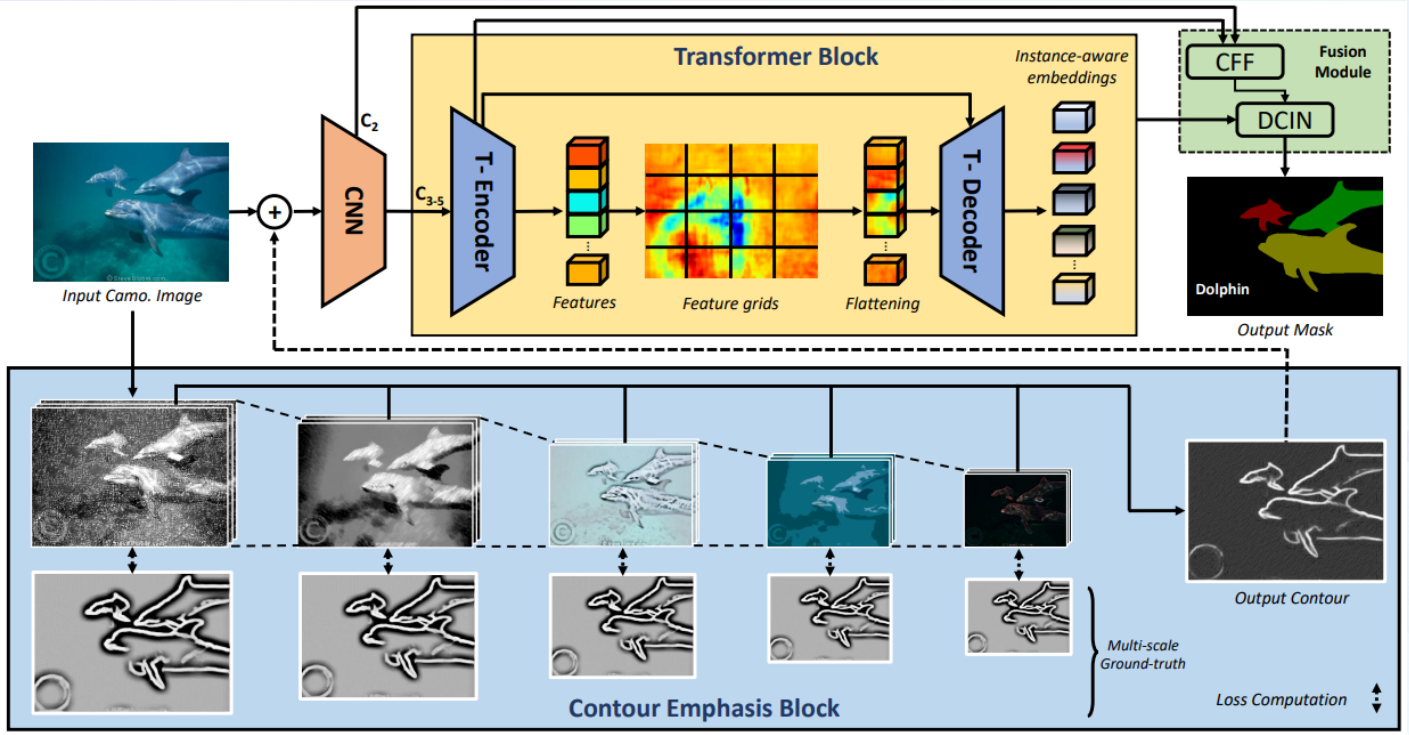
- **Camouflaged research** is attractive to the community with various tasks
- **Common approaches:** using Hand-crafted features and using Deep features
- **Instance segmentation models:** one-stage and two-stage architectures
- **Camouflaged Datasets:** COD10K, NC4K, CAMO, and CAMO++ are among the potential camouflage datasets with fine-grained annotations

Dataset	#Annot. Camo. Img.	#Meta-Cat.	#Obj. Cat.	Bbox. GT	Obj. Mask GT	Ins. Mask GT
CAMO [2]	1,250	2	8	×	✓	×
COD10K [19]	5,066	5	69	✓	✓	✓
NC4K [32]	4,121	5	69	✓	✓	✓
CAMO++ [31]	2,695	10	47	✓	✓	✓

Tab. Comparison among camouflage datasets (w/o non-camouflaged images)

3. Method

CE-OST has 2 main components:
- A Contour Emphasis Block
- A Transformer-based Instance Segmentation Block

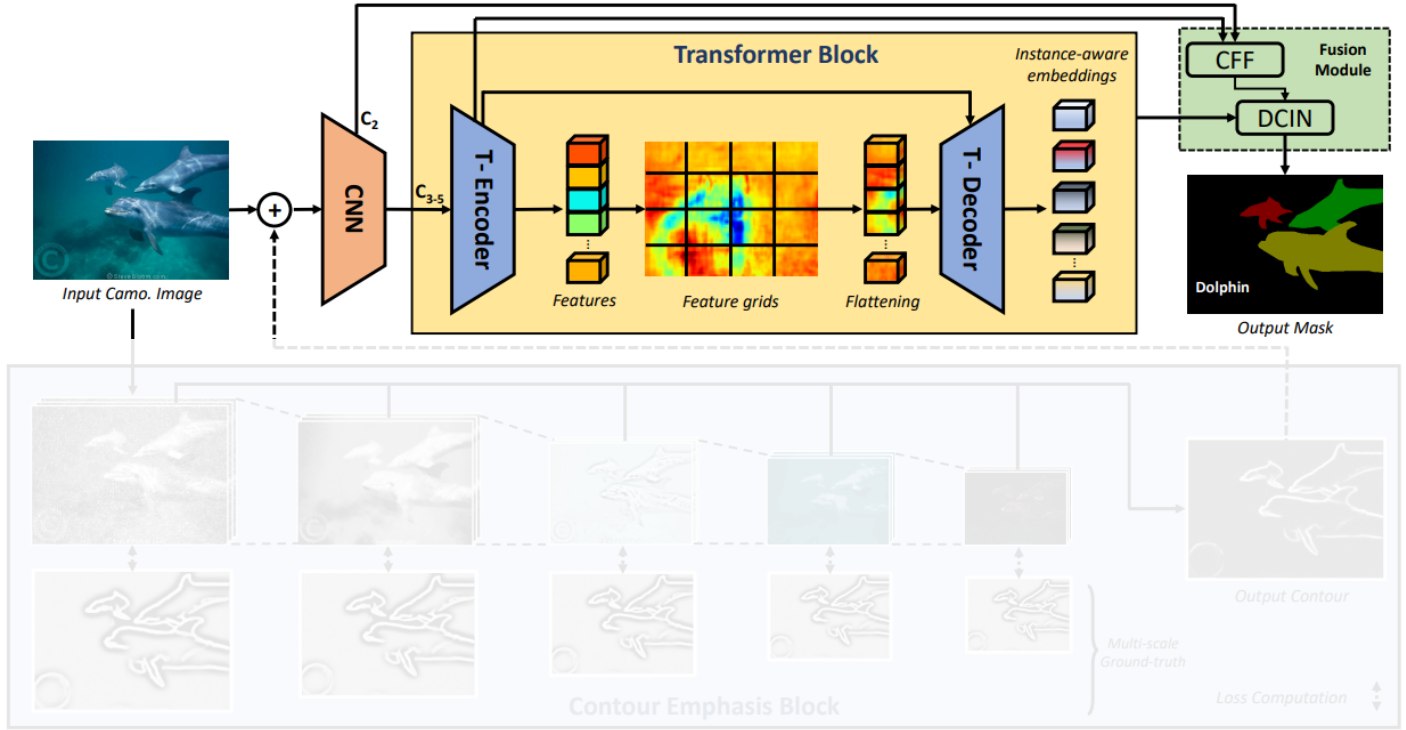


Overall our CE-OST framework: Contour Emphasis for One-Stage Transformer-based Camouflage Instance Segmentation

CE-OST enhances the contour features at image-level to boost the performance the camouflaged instance segmentation model

3. Method

A Transformer-based Instance Segmentation Block: relied on Transformer architecture with self-attention mechanism

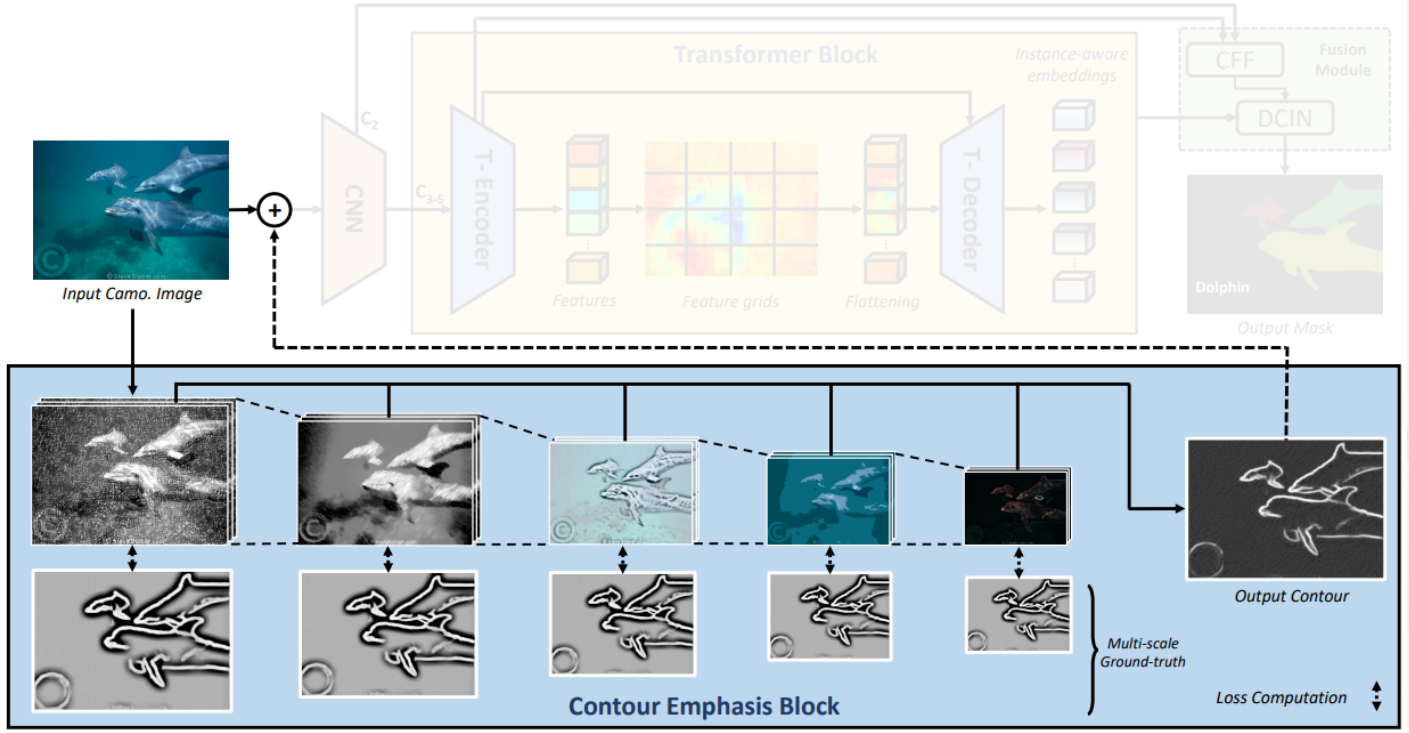


Focus: Transformer Block

- Inspired by OSFormer, a one-stage Transformer architecture
- Pre-trained backbones from OSFormer: ResNet-50, ResNet-101, SwinT, PVT
- Precise positional embeddings, necessary for instance segmentation

3. Method

Contour Emphasis Block: a multi-scale CNN (VGG-16 backbone) learning to perform edge detection*, then fuse them with the original camouflaged image



Focus: Contour Emphasis Block

- CE Block is pretrained on BSD500, a dataset for edge detection
- Total loss is the summary of Cross-entropy loss computed at each scale
- The final output contour is combined with the original image thanks to a **Grid-Condition**

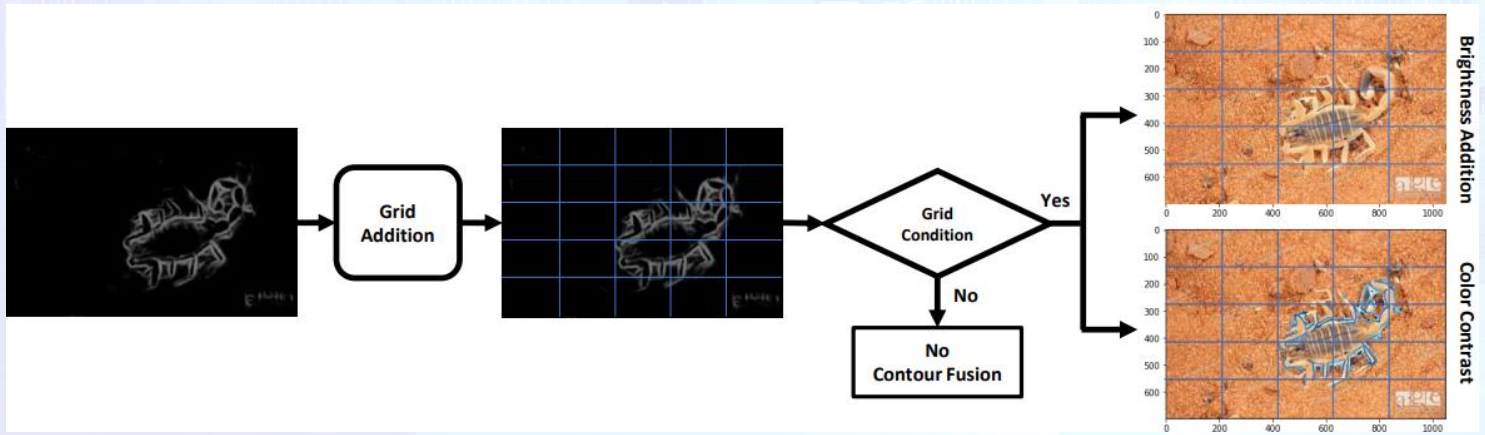
*S. Xie and Z. Tu, "Holistically-nested edge detection," in ICCV, 2015

3. Method

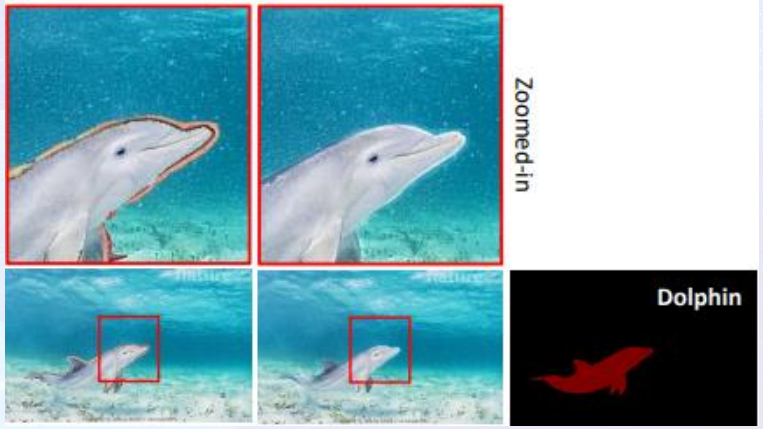
Grid-Condition: A grid 5×5 is applied to each image. Not boundary fusion if the number of eliminated cells is over half. A cell is eliminated if the number of pixels in the detected boundary area covers over a half area of the cell.

Two ways of enhancement:

- **Brightness Addition:** add **pixel-wise value** to the original image
- **Color Contrast:** add **pixel wise contrast value** to the original image



Our Grid-Condition to control the Contour Emphasis



Exemplary contour emphasized images on CAMO++

4. Experiments

- State-of-the-art comparison among **one-stage** and **two-stage methods**
- Our CE-OST achieved the highest AP score at 43.2% on COD10K and 45.1% on NC4K thanks to the contour emphasizing

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. State-of-the-art comparison on COD10K and NC4K dataset. The chosen backbone is the common ResNet-101.

4. Experiments – Ablation Study

- **PVT backbone stably holds the best performance** as its multi-scale feature extractor can well handle the various scales of CAMO++
- **CAMO++** is the most intensive dataset, following by NC4K and COD10K

Method	Base-Model	COD10K			NC4K			CAMO++		
		AP	AP50	AP75	AP	AP50	AP75	AP	AP50	AP75
OSFormer	ResNet-50 [59]	41.0	71.1	40.8	42.5	72.5	42.3	19.0	33.8	18.3
	ResNet-50-550 [59]	-	-	-	-	-	-	20.1	36.3	19.3
	ResNet-101 [59]	42.0	71.3	42.8	44.4	73.7	45.1	20.6	34.4	20.2
	PVTv2-B2-Li [60]	47.2	74.9	49.8	-	-	-	27.7	44.7	27.9
	Swin-T [61]	47.7	78.6	49.3	-	-	-	22.3	36.6	21.8
CE-OST (Color Contrast)	ResNet-50 [59]	41.6	70.7	42.3	42.4	71.4	42.6	20.1	34.2	19.6
	ResNet-50-550 [59]	35.9	65.2	34.3	41.1	70.9	41.1	20.6	35.7	20.0
	ResNet-101 [59]	43.2	72.2	44.1	45.1	74.0	46.4	21.7	36.6	21.3
	PVTv2-B2-Li [60]	48.4	75.7	51.3	51.4	77.9	55.0	28.5	45.3	29.9
	Swin-T [61]	49.1	78.0	52.1	50.5	78.9	53.1	22.7	37.6	22.4
CE-OST (Brightness Addition)	ResNet-50 [59]	41.2	69.0	41.6	42.4	71.1	42.9	20.2	34.8	19.5
	ResNet-50-550 [59]	35.9	65.2	34.6	40.8	71.1	40.3	21.0	37.1	20.3
	ResNet-101 [59]	42.4	70.8	43.7	44.2	73.1	45.0	21.1	34.4	20.9
	PVTv2-B2-Li [60]	47.9	74.6	50.5	51.1	77.3	54.9	27.9	45.1	29.2
	Swin-T [61]	49.0	78.5	51.4	50.8	79.3	53.9	22.7	38.4	23.1

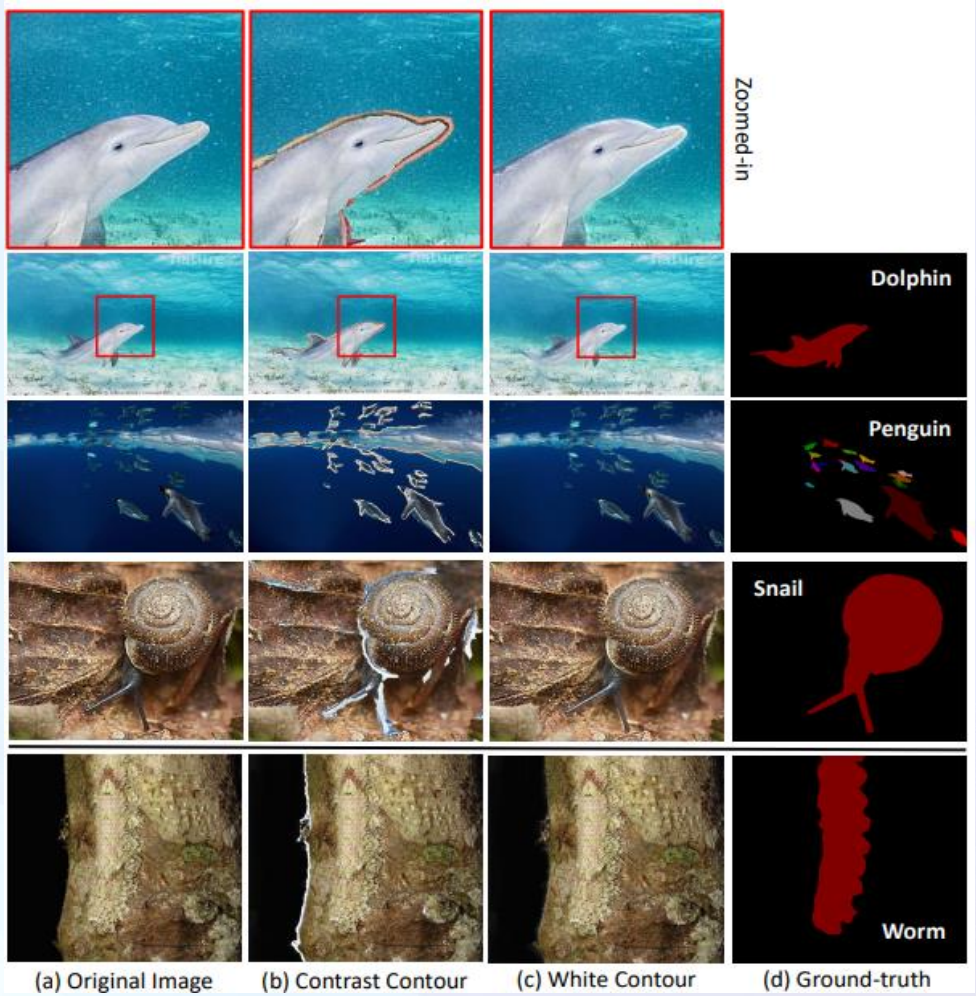
**The first, second, and third best results are marked in red, blue, and green, respectively.*

Tab. Ablation study on different base models on COD10K, NC4K, and CAMO++.

Qualitity visualization result on the CAMO++ testing set on our CE-OST-PVT. The confidence threshold is 0.5.

4. Experiments – Ablation Study

- **Color Contrast shifts the color values to another value in the contrast range**, bringing better distinguished contour views compared to Brightness Addition
- The contour emphasis is **highly affected** by the contour detection results



Exemplary contour emphasized images on CAMO++. The first rows are the zoomed-in regions.

5. Conclusion

In this work:

- We proposed **CE-OST framework –** a **Contour Emphasis** approach for **One-Stage Transformer-based model** to address the instance segmentation task on camouflaged images
- Experimental results proves our SOTA results among the surveyed methods on CAMO++, NC4K, and COD10K datasets

In the future:

- Train the camouflaged specific Edge Detection model
- Extend our idea to other specific domains of medical imaging where the instances carry camouflaged features

ORGANIZERS

MAPR 2023
THE 6TH INTERNATIONAL CONFERENCE ON MULTIMEDIA ANALYSIS AND PATTERN RECOGNITION

SPONSORS

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,

{danhnt, tuanld, tiepnv, thanhnd}@uit.edu.vn, †corresponding author

Acknowledgements

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
VNUHCM - UIT

Quy Nhon, October 5-6th, 2023