MAPR 2025 THE 8TH INTERNATIONAL CONFERENCE ON MULTIMEDIA ANALYSIS AND PATTERN RECOGNITION

Generative One-shot Camouflage Instance Segmentation

Thanh-Danh Nguyen[1,2], Vinh-Tiep Nguyen†[1,2], and Tam V. Nguyen[3]

[1]University of Information Technology, Ho Chi Minh City, Vietnam,

[2]Vietnam National University, Ho Chi Minh City, Vietnam, [3]University of Dayton, Dayton, OH 45469, United States

{danhnt, tiepnv}@uit.edu.vn, tamnguyen@udayton.edu, †corresponding author

Project Page

INTRODUCTION

Motivation: Addressing camouflage instance segmentation (CIS) efficiently given one-shot annotated samples.

Main contribution: **CAMO-GenOS**, a CIS framework leveraging one-shot annotated samples to drive a generative process for data enrichment.

Fig. 1. Our CAMO-GenOS Concept Idea Presentation.

METHOD

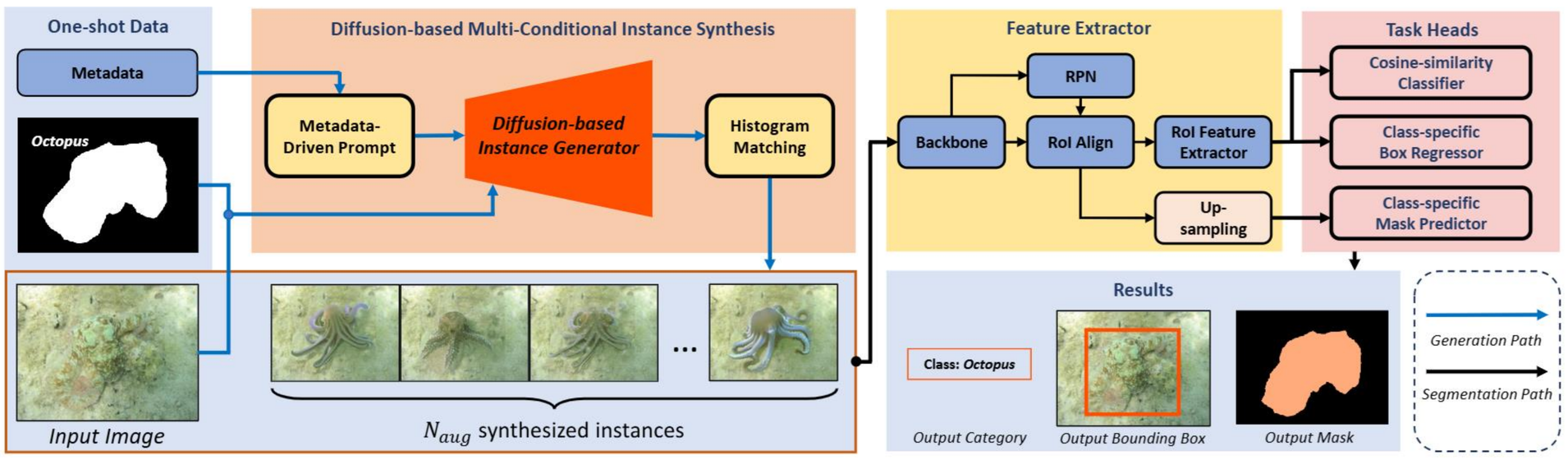

Fig. 2. Overview of our CAMO-GenOS framework leveraging generative models to enhance one-shot camouflage instance segmentation.

Our **CAMO-GenOS** is a generative approach to exploit and enrich the one-shot CIS task.

One-shot Fine-tuning CIS: The CIS model is presented in two phases:

- **Base phase**: training with abundant annotated training data C_{base}
- **Novel phase**: fine-tuning on a disjoint set of novel classes C_{novel} containing one sample per each class of novel data

In our case, the one-shot fine-tuning relies on *a single annotated sample*, enriched by our proposed generative instance synthesis method.

Diffusion-based Multi-Conditional Instance Synthesis:

The diffusion-based generative model $G(.)$ strictly takes the multiple conditions of the referenced query image $I_q \in$ novel training samples in C_{novel}, ground truth mask M_q, and guided text prompt P_q to return N_{aug} samples.

- **Metadata-Driven Conditional Text Prompt** follows the structure: *"a photo of a/an [size] [meta-class] [instance class]"*
- **Histogram Matching Post-processing** is to narrow the gap between the distribution of the original and the synthesized samples

RESULT

Tab. 1. SoTA comparison of our CAMO-GenOS evaluated on CAMO-FS benchmark. The backbones are COCO-80 FPN-ResNet-101

Method	Synthesis Base	nAP		nAP50	nAP75	nAPs	nAPm	nAPl	nAR1
Mask-RCNN [28]		2.99		5.73	3.26	20.68	3.06	2.74	12.45
iMTFA [11]		3.66		5.37	4.09	22.42	4.35	2.01	11.30
iFS-RCNN [10]		4.27		5.98	4.75	21.57	5.71	4.87	11.70
FS-CDIS [7]		4.46	-	7.34	4.84	25.50	5.60	3.48	14.77
CAMO-GenOS (ours)	BlendedDiff [19]	4.80	+0.34	7.79	5.37	28.59	5.67	3.32	17.85
	DiffInpainting [17]	4.91	+0.45	7.84	5.47	26.54	5.06	4.02	17.18
	GLIGEN [18]	4.74	+0.28	7.53	5.31	28.10	4.79	5.28	17.65

Tab. 2. Ablation study on multiple instance generation-based methods

	Instance Segmentation			Object Detection			
Method	nAP		nAP50	nAP75	nAP		nAP50	nAP75
FS-CDIS [7]	4.46		7.34	4.84	3.88		7.71	3.21
+ ITL	4.55		7.52	4.94	3.99		7.92	3.47
+ IMS	3.94		7.44	3.64	4.01		8.05	3.44
+ Both	4.10		7.40	4.15	3.99		7.82	3.40
CAMO-GenOS (ours)								
w/ BlendedDiff [19]	4.80	+0.34	7.79	5.37	4.90	+1.02	8.09	4.78
+ ITL	5.16	+0.61	8.25	5.73	4.97	+0.98	8.54	5.08
+ IMS	4.19	+0.25	7.98	4.54	4.75	+0.74	8.38	5.16
+ Both	4.25	+0.15	7.36	4.71	4.79	+0.80	7.71	4.52
w/ DiffInpainting [17]	4.91	+0.45	7.84	5.47	5.00	+1.12	8.33	5.26
+ ITL	4.80	+0.25	7.90	5.32	4.97	+0.98	8.29	4.61
+ IMS	4.04	+0.10	7.21	4.34	4.68	+0.69	7.84	4.84
+ Both	4.29	+0.19	7.30	4.60	4.70	+0.71	7.83	4.86
w/ GLIGEN [18]	4.74	+0.28	7.53	5.31	4.83	+0.95	7.94	4.85
+ ITL	5.30	+0.75	8.26	6.02	5.23	+1.24	8.63	5.61
+ IMS	4.39	+0.45	7.28	4.86	4.52	+0.51	7.87	4.52
+ Both	4.33	+0.23	7.28	4.74	4.75	+0.76	7.62	5.52

**The increased values in blue are compared to the SoTA baseline FS-CDIS [7] with the corresponding ITL, IMS, and both of them.*