UNIVERSITY OF INFORMATION TECHNOLOGY
VIETNAM NATIONAL UNIVERSITY OF HO CHI MINH CITY

MMLab UIT

PAPER PRESENTATION

CAMO-InstSynth: Few-shot Camouflage Instance Segmentation with Multi-Conditional Background Synthesis and Generative Augmentation

Thanh-Danh Nguyen[1,2], *Vinh-Tiep Nguyen*[1,2*], *Kunpeng Li*[3], *and Tam V. Nguyen*[4]

[1]*University of Information Technology, Ho Chi Minh City, Vietnam,* [2]*Vietnam National University, Ho Chi Minh City, Vietnam*

[3]*Air Force Institute of Technology, Ohio, 45433, United States,* [4]*University of Dayton, Dayton, Ohio, 45469, United States*

*Contacts: {danhnt, tiepnv}@uit.edu.vn, kunpeng.li@us.af.mil, tamnguyen@udayton.edu,*corresponding author*

Contents

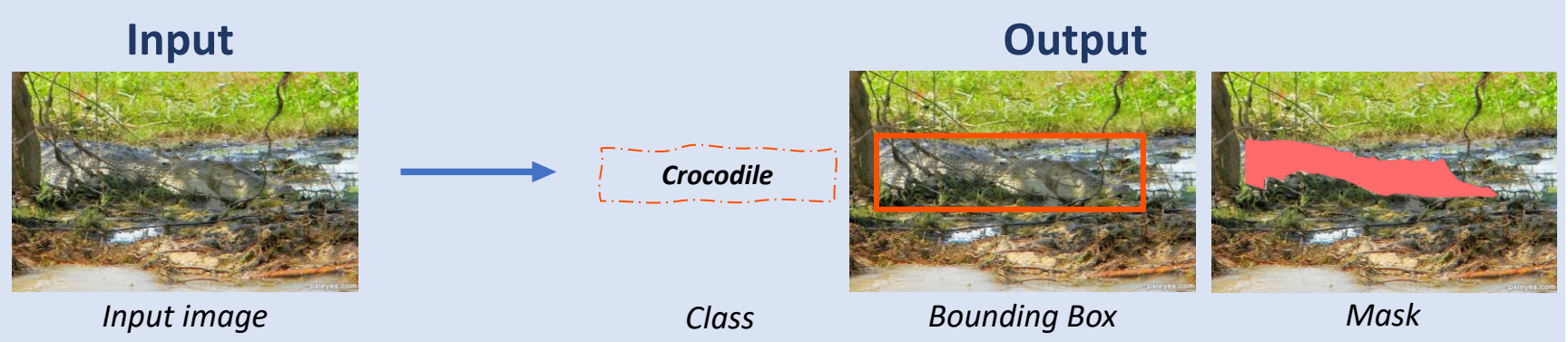
Introduction – Context

- ***Camouflage*** is **a complex biological strategy** that animals use to **conceal their appearance** by blending in with their habitat
- **Identifying camouflaged objects has various applications:** search-and-rescue work, wildlife population monitoring, media forensics, etc.

Exemplary camouflaged samples extracted from CAMO-FS dataset

Problem statement

- Camouflage object detection and instance segmentation:



- Challenges: Lacking fine-grained annotated data for camouflage domain

➔ Few-shot learning

- Remaining challenges: few-shot data lacks diversity due to limited number of samples

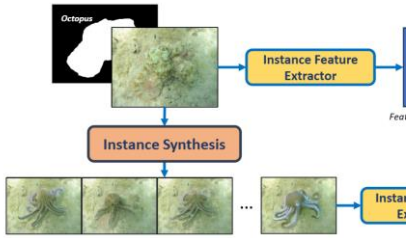
➔ Data enhancement

How to efficiently enhance data for camouflage FSOD and FSIS?

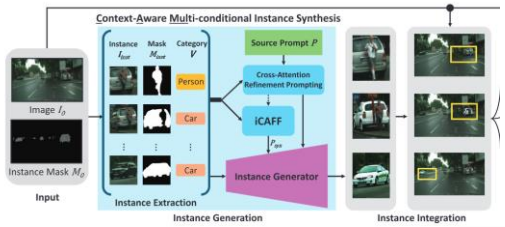
Problem statement

How to efficiently enhance data for camouflage FSOD and FSIS?

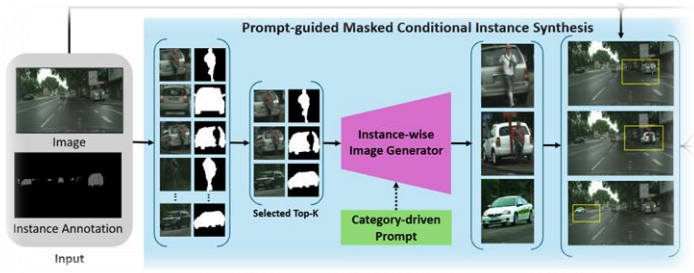
- Traditional data augmentation
- Generative data augmentation
  - Foreground-aware
  - Background-aware



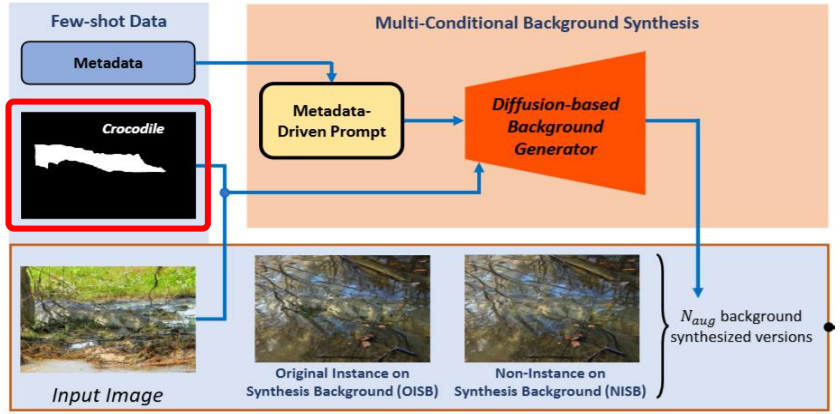
CAMO-GenOS (2025)



CAMUL (2025)



InstSynth (2024)

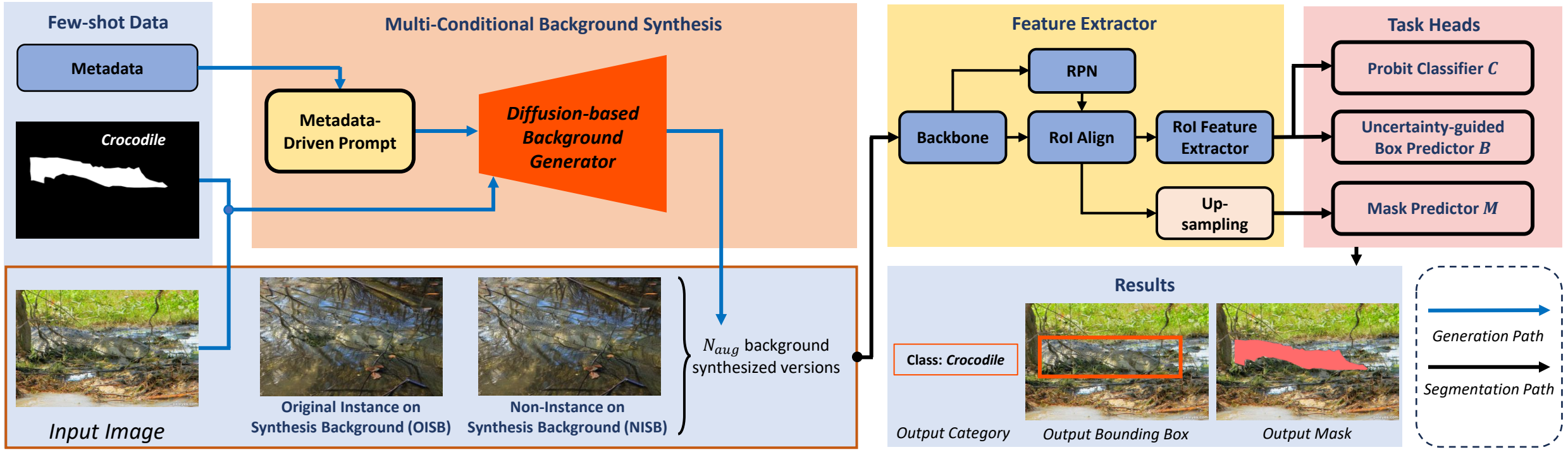


CAMO-InstSynth (ours)

Key contributions

- **CAMO-InstSynth** framework with a **generative multi-conditional background synthesis** method to enhance existing **few-shot** camouflage samples on FSOD and FSIS
- Extensive experiments on CAMO-FS benchmark to demonstrate the robustness of CAMO-InstSynth over existing SoTAs

Proposed method – CAMO-InstSynth



Multi-Conditional Background Synthesis

- We consider image I_{real} consisting of a foreground I_{fg} and a background I_{bg}

$$I_{\text{real}} = I_{\text{fg}} \odot M \; + \; I_{\text{bg}} \odot (1 - M)$$

- Then, we synthesize two background with two versions:

$$I_{\text{OISB}} = I_{\text{fg}} \odot M \; + \; \hat{I}_{\text{bg}} \odot (1 - M)$$

Input Image *Mask M*

$$I_{\text{NISB}} = \hat{I}_{\text{bg}} \odot (1 - M)$$

Input Image *Mask 1 - M*

- These inputs are fed into the Diffusion-based Background Generator, i.e., Blended Diffusion[1]

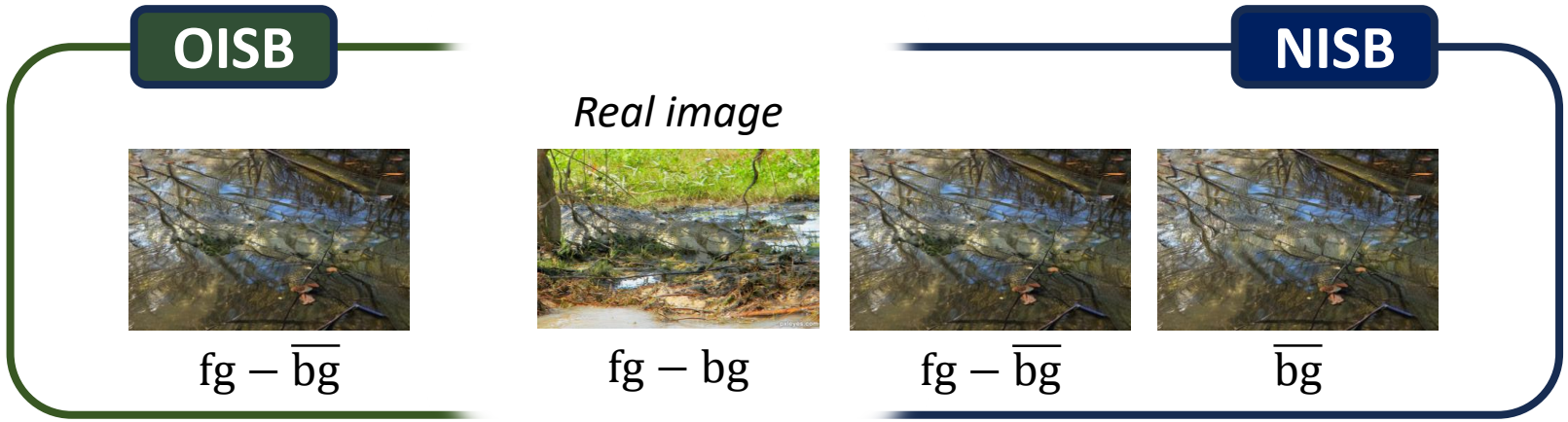
[1] Avrahami, O., Lischinski, D., Fried, O.: Blended diffusion for text-driven editing of natural images. In: CVPR 2022

Sampling strategy

Notation:

- Real image: $\mathrm{fg} - \mathrm{bg} + \mathrm{fg} - \overline{\mathrm{bg}} + \overline{\mathrm{bg}}$
- Real foreground: fg
- Synthetic background: $\overline{\mathrm{bg}}$

➔ Sampling ratio: $1\ \mathrm{fg} - \mathrm{bg} : 1\ \mathrm{fg} - \overline{\mathrm{bg}} : 1\ \overline{\mathrm{bg}}$



Experiments – SoTA comparison

Model			Novel AP									
			Instance Segmentation					Object Detection				
Method	Year	Baseline	1	2	3	5	Avg.	1	2	3	5	Avg.
MTFA [4]	2021		2.48	6.67	5.81	6.40	5.34	1.98	6.47	5.82	6.17	5.11
M-RCNN† [9]	2017	ResNet-50	4.08	6.79	6.90	8.29	6.52	2.82	5.09	5.46	6.18	4.89
iFS-RCNN [22]	2022		4.17	6.26	5.73	6.38	5.64	3.92	6.06	5.47	6.60	5.51
MTFA [4]	2021		3.66	6.21	6.16	5.95	5.50	2.93	5.90	5.84	5.84	5.13
M-RCNN† [9]	2017		4.39	7.69	7.94	10.09	7.53	3.03	5.80	6.20	7.79	5.71
iFS-RCNN [22]	2022		4.27	6.55	6.07	7.80	6.17	3.79	6.28	6.01	8.08	6.04
FS-CDIS-ITL [28]	2024	ResNet-101	4.46	5.57	6.41	8.48	6.23	4.04	7.28	7.49	9.76	7.14
FS-CDIS-IMS [28]	2024		5.46	6.95	7.36	9.61	7.35	4.50	6.95	7.55	10.36	7.34
CAMO-Freq [24]	2025		5.71	-	-	8.31	7.01	5.56	-	-	8.89	7.23
CAMO-GenOS [26]	2025		4.91	-	-	-	4.91	5.00	-	-	-	5.00
			Our performance									
CAMO-InstSynth†	2026	ResNet-101	6.17	8.51	8.81	9.52	**8.25**	5.91	8.32	8.68	8.10	**7.75**
CAMO-InstSynth††	2026		6.12	8.62	9.16	10.48	**8.60**	5.79	8.55	8.63	8.33	**7.83**

M-RCNN† is Mask R-CNN [9] with sigmoid classifier.
CAMO-InstSynth† utilizes OISB; CAMO-InstSynth†† utilizes NISB; Both are built on top of iFS-RCNN [22].

- CAMO-InstSynth achieves the highest average accuracy over SoTAs on CAMO-FS benchmark

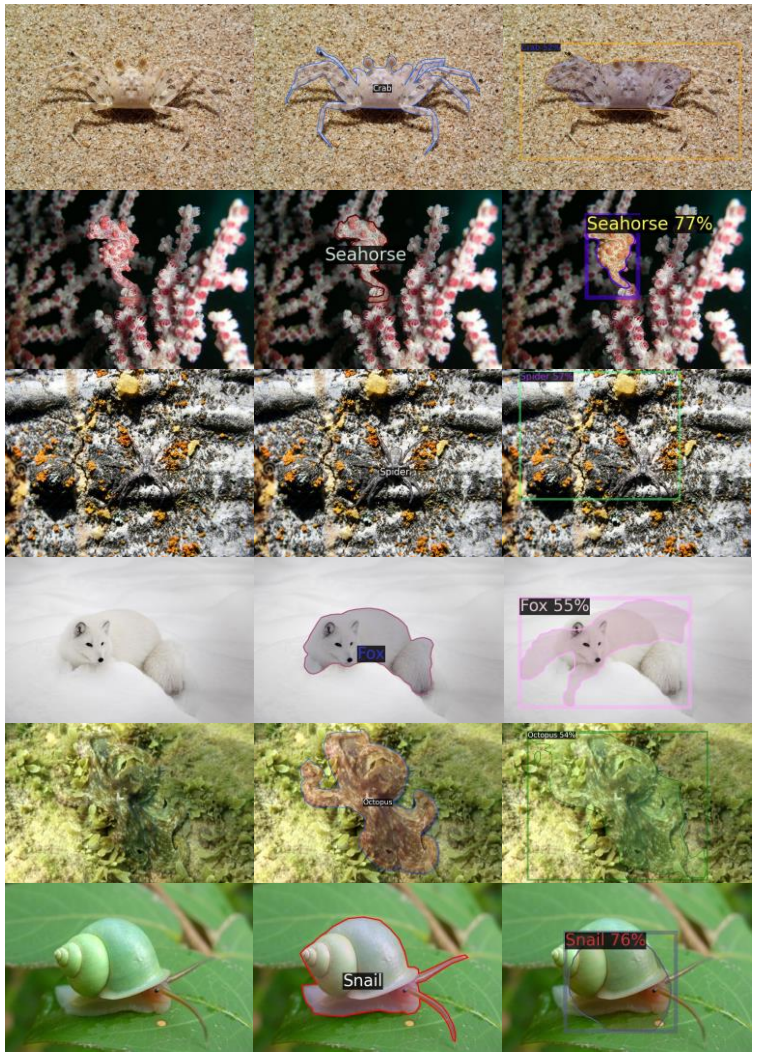


Image Ground-truth Prediction

Experiments – Ablation study

#	Method	AP	AP50	AP75	APs	APm	APl	AR1	AR10	AR100	ARs	ARm	ARl
	Instance Segmentation												
1	**Baseline**	5.49	8.14	6.04	25.86	5.48	4.40	23.49	28.26	28.39	36.05	20.01	28.69
	iFS-RCNN + OISB	6.17	8.96	7.00	27.63	6.34	6.55	24.41	29.76	29.80	38.71	19.93	29.71
	iFS-RCNN + NISB	6.12	8.87	6.81	26.64	5.57	5.31	24.28	30.33	30.35	37.85	21.51	30.07
2	**Baseline**	8.07	12.08	9.27	32.59	6.19	6.40	27.47	36.62	36.86	41.49	23.89	36.86
	iFS-RCNN + OISB	8.51	12.62	10.10	34.45	6.88	6.51	28.79	36.97	37.19	42.87	24.83	36.40
	iFS-RCNN + NISB	8.62	12.53	9.98	34.29	7.99	5.56	29.48	37.17	37.41	41.63	26.16	36.75
3	**Baseline**	8.26	12.23	9.37	36.68	6.98	6.92	29.83	39.92	40.43	44.37	25.00	39.55
	iFS-RCNN + OISB	8.81	12.78	9.97	38.29	7.12	6.08	30.41	39.76	40.06	47.48	27.19	38.64
	iFS-RCNN + NISB	9.16	13.22	10.24	36.78	7.24	8.69	30.15	39.37	39.61	44.43	26.81	38.30
5	**Baseline**	8.76	13.41	10.04	38.73	6.02	8.36	28.80	40.51	41.40	43.38	27.06	41.03
	iFS-RCNN + OISB	9.52	14.39	10.77	39.02	6.73	7.48	29.71	41.21	41.69	43.67	27.75	40.90
	iFS-RCNN + NISB	10.48	15.84	11.54	38.18	7.19	9.03	29.06	41.14	41.76	42.83	29.41	41.55
	Object Detection												
1	**Baseline**	5.25	8.16	5.90	29.74	7.06	4.17	20.60	25.53	25.72	34.33	19.34	26.95
	iFS-RCNN + OISB	5.91	9.11	6.75	30.05	7.91	6.47	22.56	27.65	27.72	34.91	19.50	27.89
	iFS-RCNN + NISB	5.79	8.93	6.63	29.44	7.59	5.03	22.73	28.13	28.23	35.04	20.86	29.08
2	**Baseline**	7.76	12.13	9.15	32.56	8.44	6.34	24.93	32.83	32.98	38.27	22.65	33.82
	iFS-RCNN + OISB	8.32	12.55	9.60	34.56	10.17	6.26	27.70	34.93	35.08	40.53	23.69	34.50
	iFS-RCNN + NISB	8.55	12.40	9.68	35.00	10.48	5.18	28.13	35.38	35.65	39.86	24.65	35.35
3	**Baseline**	7.77	12.32	8.47	37.18	8.52	6.33	26.36	35.28	35.70	41.37	22.84	36.22
	iFS-RCNN + OISB	8.68	12.64	9.95	39.36	12.29	6.36	29.36	38.32	38.68	45.55	25.99	38.36
	iFS-RCNN + NISB	8.63	13.13	9.29	38.08	11.50	8.08	28.34	36.62	36.88	42.02	25.66	36.36
5	**Baseline**	7.20	13.87	7.14	33.45	7.27	6.81	24.32	33.67	34.52	36.09	21.76	34.69
	iFS-RCNN + OISB	8.10	14.52	7.99	33.10	10.58	6.89	25.58	34.00	34.98	34.97	22.81	34.95
	iFS-RCNN + NISB	8.33	16.21	8.14	32.70	10.43	7.66	24.18	35.14	35.87	35.21	23.17	36.10

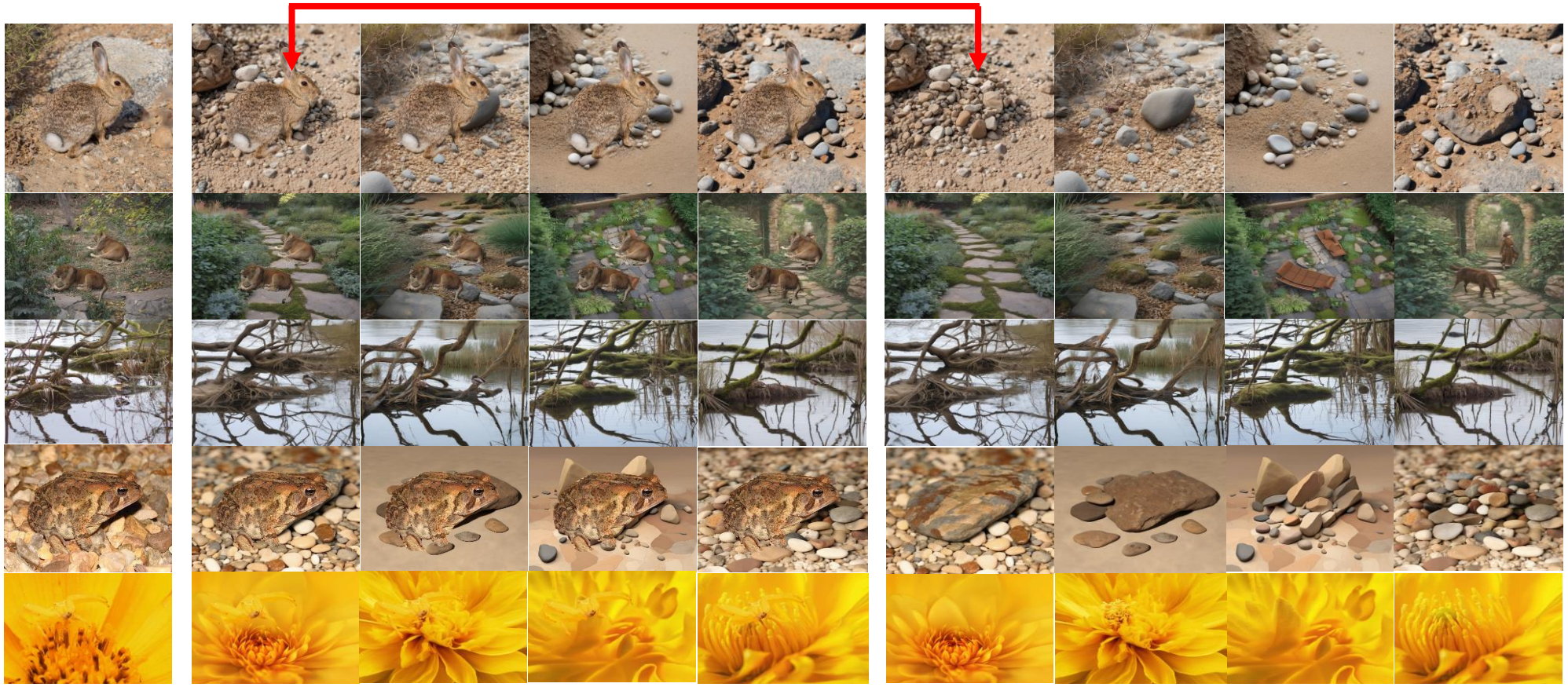
The breakdown improvement of our proposed CAMO-InstSynth over two different approaches on top of baseline iFS-RCNN

- Both OISB and NISB variants outperform the vanilla iFS-RCNN
- iFS-RCNN + NISB consistently achieves the best performance

➔ Background-aware synthesis better captures camouflage characteristics where foreground and background are strongly entangled

Experiments – Ablation study

Exemplary background synthesis results between the two variants of OISB and NISB



Original Image | **Original Instance on Synthesis Background (OISB)** | **Non-Instance on Synthesis Background (NISB)**

Conclusion

In this work, we contribute:

- **CAMO-InstSynth** framework with a **generative multi-conditional background synthesis** method to enhance existing **few-shot** camouflage samples on FSOD and FSIS
- Extensive experiments on CAMO-FS benchmark to demonstrate the robustness of CAMO-InstSynth over existing SoTAs

In the future,

- Handle more complex camouflage scenarios, i.e., multi-object, multi-scale, cluttered environments
- Investigate the framework generalization and consider an end-to-end manner

UNIVERSITY OF INFORMATION TECHNOLOGY
VIETNAM NATIONAL UNIVERSITY OF HO CHI MINH CITY

MMLab UIT

PAPER PRESENTATION

CAMO-InstSynth: Few-shot Camouflage Instance Segmentation with Multi-Conditional Background Synthesis and Generative Augmentation

Thanh-Danh Nguyen[1,2], *Vinh-Tiep Nguyen*[1,2*], *Kunpeng Li*[3], *and Tam V. Nguyen*[4]

[1]*University of Information Technology, Ho Chi Minh City, Vietnam,* [2]*Vietnam National University, Ho Chi Minh City, Vietnam*

[3]*Air Force Institute of Technology, Ohio, 45433, United States,* [4]*University of Dayton, Dayton, Ohio, 45469, United States*

*Contacts: {danhnt, tiepnv}@uit.edu.vn, kunpeng.li@us.af.mil, tamnguyen@udayton.edu,*corresponding author*

This research is funded by Vietnam National University Ho Chi Minh City (VNU-HCM) under grant number DS.C2025-26-08