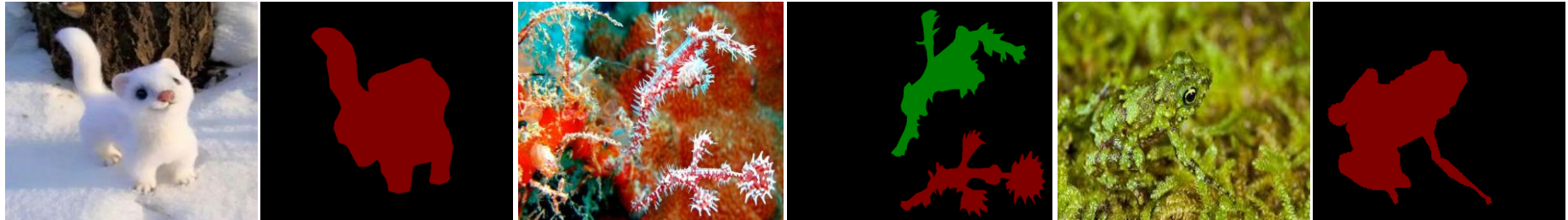




**LUẬN VĂN THẠC SĨ
NGÀNH KHOA HỌC MÁY TÍNH**

PHÂN ĐOẠN THỰC THỂ NGUY TRẠNG DỰA TRÊN ĐẶC TRƯNG CÓ TÍNH PHÂN BIỆT CAO



**NGUYỄN THÀNH DANH – 210101001
GVHD: TS. NGUYỄN VINH TIỆP**



Nội dung

1. Giới thiệu đề tài
2. Các công trình liên quan
3. Phương pháp **CE-OST** khai thác đặc trưng biên cạnh
4. Tập dữ liệu **CAMO-FS**
5. Phương pháp **FS-CDIS** học đặc trưng phân biệt với ít mẫu
6. Kết luận và hướng phát triển

Giới thiệu đề tài



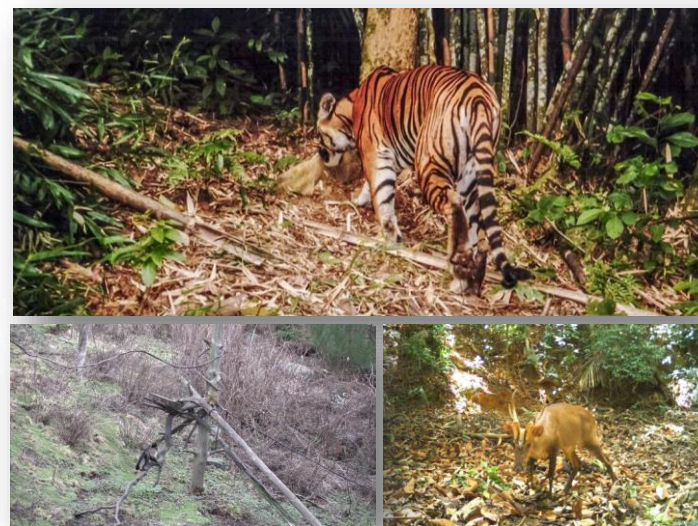
- **Ngụy trang** là **tập tính tự vệ** của các loài động vật nhằm **lẩn trốn** khả năng **quan sát** của động vật khác **bằng cách ẩn mình vào môi trường xung quanh**
- Nhận biết tự động các đối tượng **ngụy trang** với mức độ **chi tiết cao** có **nhiều ứng dụng thực tiễn**



Phát hiện mục tiêu quân sự



Phục vụ công tác tìm kiếm và giải cứu



Tìm kiếm và bảo tồn động vật quý hiếm



Định nghĩa bài toán

Bài toán: Phân đoạn thực thể nguy trang

- **Đầu vào:** ảnh chứa các thực thể nguy trang
- **Đầu ra:** nhãn ngữ nghĩa ứng với từng thực thể nguy trang



Đầu vào



Đầu ra

Định nghĩa bài toán (tt)



Đặc thù của thực thể nguy trang:

- Có **màu sắc, chất liệu, hoa văn tương đồng** với vùng nền
- Dễ nhầm lẫn với môi trường xung quanh



Minh họa một số hình ảnh nguy trang

➔ **Đặc trưng phân biệt cao**: hỗ trợ phân biệt hiệu quả **vùng thực thể** và **vùng nền**:

1. Đặc trưng **biên cạnh thực thể**
2. Đặc trưng **vùng chứa thực thể**





Các thách thức

1. Chưa có nhiều mô hình phân đoạn thực thể đối tượng ngẫu trang

- Chủ yếu phục vụ các loại đối tượng phổ biến
- Chưa khai thác các yếu tố đặc thù của thực thể ngẫu trang

2. Tập dữ liệu ngẫu trang còn một số hạn chế

- Số lượng mẫu dữ liệu ít
- Số tập dữ liệu cho bài toán này còn ít

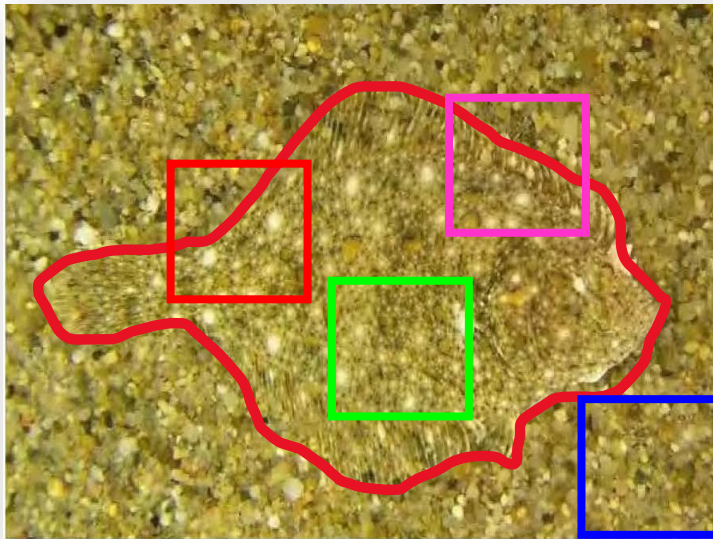
Thống kê một số tập dữ liệu về thực thể ngẫu trang (chỉ xét dữ liệu ảnh/video chứa thực thể ngẫu trang)

Tập dữ liệu	Năm	Loại	#Gán nhãn ngẫu trang	#Lớp tổng quát	#Lớp đối tượng	Nhãn khung bao	Nhãn ngữ nghĩa	Nhãn thực thể
CamouflagedAnimals [60]	2016	Video	181	-	6	×	✓	✓
MoCA [35]	2020	Video	7,617	-	67	✓	×	×
CHAMELEON [69]	2018	Ảnh	76	-	-	×	✓	×
CAMO [38]	2019	Ảnh	1,250	2	8	×	✓	×
COD [12]	2020	Ảnh	5,066	5	69	✓	✓	✓
NC4K [49]	2021	Ảnh	4,121	5	69	✓	✓	✓
CAMO++ [36]	2022	Ảnh	2,695	10	47	✓	✓	✓



Mục tiêu của luận văn

1. Khai thác đặc trưng tại vùng biên cạnh để phân đoạn thực thể nguy trang



Biên cạnh thực thể



Vùng thực thể



Vùng nền

Minh họa sự tương đồng giữa các vùng trong ảnh chứa thực thể nguy trang.

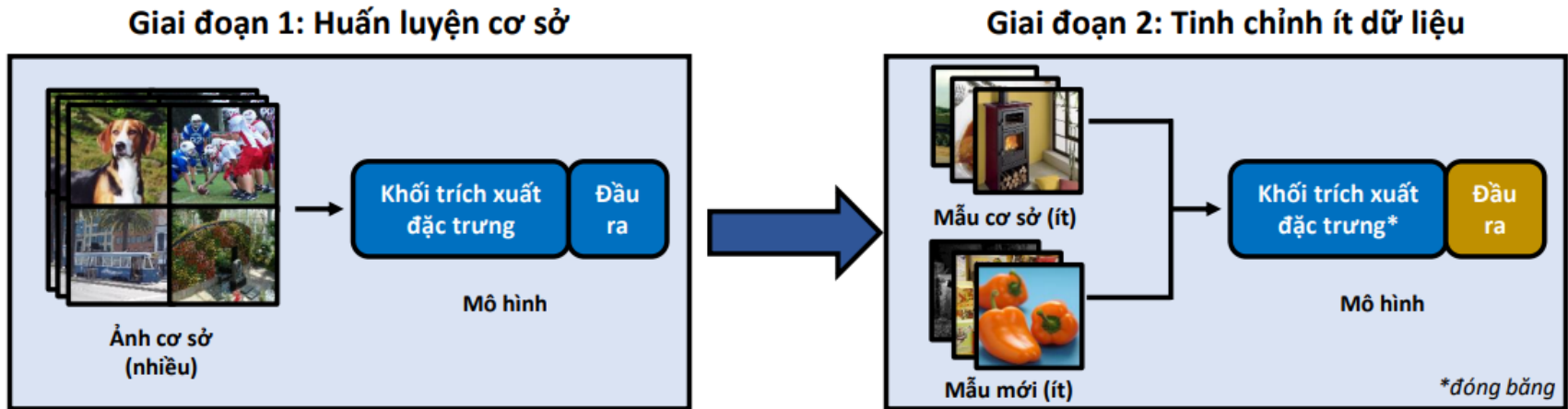
- Tính chất nguy trang **không liên tục ở vùng tiếp giáp** giữa thực thể và vùng nền

→ Đây là đặc trưng có **tính phân biệt cao**



Mục tiêu của luận văn (tt)

2. Học đặc trưng phân biệt với ít dữ liệu huấn luyện



- **Dữ liệu gán nhãn** cho phân đoạn thực thể nguy trang **không nhiều**
- ➔ Tiếp cận **sử dụng ít mẫu huấn luyện** với 2 giai đoạn: học trên dữ liệu tổng quát (GĐ1) và tinh chỉnh trên dữ liệu nguy trang (GĐ2)

Đóng góp chính của luận văn



1

Mô hình **CE-OST**

Tận dụng hiệu quả đặc trưng biên cạnh thực thể [CT1]

2

Tập dữ liệu **CAMO-FS**

Xây dựng bộ dữ liệu cho hướng cận học ít mẫu [CT2, CT3]

3

Mô hình **FS-CDIS**

Kỹ thuật học tương phản với ít mẫu huấn luyện [CT2, CT3]



- [CT1]** **Thanh-Danh Nguyen**, Duc-Tuan Luu, Vinh-Tiep Nguyen, and Thanh Duc Ngo. “*CE-OST: Contour Emphasis for One-Stage Transformer-based Camouflage Instance Segmentation*”. In Proceedings of the International Conference on Multimedia Analysis and Pattern Recognition (MAPR), 2023. (**Best Student Paper Award**)

- [CT2]** **Thanh-Danh Nguyen**, Anh-Khoa Nguyen Vu, Nhat-Duy Nguyen, Vinh-Tiep Nguyen, Thanh Duc Ngo, Thanh-Toan Do, Minh-Triet Tran, and Tam V. Nguyen. “*Few-shot Camouflaged Animal Detection and Segmentation*”. In CV4Animals: Computer Vision for Animal Behavior Tracking and Modeling, in conjunction with CVPR2022 (**Invited Poster**).

- [CT3]** **Thanh-Danh Nguyen**, Anh-Khoa Nguyen Vu, Nhat-Duy Nguyen, Vinh-Tiep Nguyen, Thanh Duc Ngo, Thanh-Toan Do, Minh-Triet Tran, and Tam V. Nguyen. “*The Art of Camouflage: Few-shot Learning for Animal Detection and Segmentation*”. Under review at Journal of Signal, Image and Video Processing, 2023.



Các công trình liên quan

- **OSFormer** (ECCV22)^[1]: 1 giai đoạn, học đặc trưng vị trí thực thể
- **De-Camouflaging** (CVPR23)^[2]: 2 giai đoạn, học loại bỏ đặc trưng ngụy trang
- ➔ Chưa khai thác đặc trưng ở vùng biên cạnh trên thực thể ngụy trang

- **iMTFA** (CVPR21)^[3], **iFS-RCNN** (CVPR22)^[4]: học ít mẫu dữ liệu có hiệu quả trên MS-COCO
- ➔ Chưa có mô hình học ít mẫu cho thực thể ngụy trang

- Dữ liệu phục vụ nghiên cứu thực thể ngụy trang: **CAMO++**, **COD10K**, **NC4K**
- ➔ Chưa có dữ liệu chuẩn cho bài toán học ít mẫu ngụy trang

➔ Khai thác đặc trưng vùng biên cạnh trên thực thể ngụy trang và học hiệu quả với ít dữ liệu huấn luyện

[1] J. Pei et al., "OSFormer: One-Stage Camouflaged Instance Segmentation with Transformers", ECCV2022

[2] N. Luo et al., "Camouflaged Instance Segmentation via Explicit De-Camouflaging", CVPR2023

[3] DA. Ganea et al., "Incremental few-shot instance segmentation", CVPR2021

[4] K. Ng et al., "iFS-RCNN: An Incremental Few-shot Instance Segmenter", CVPR2022



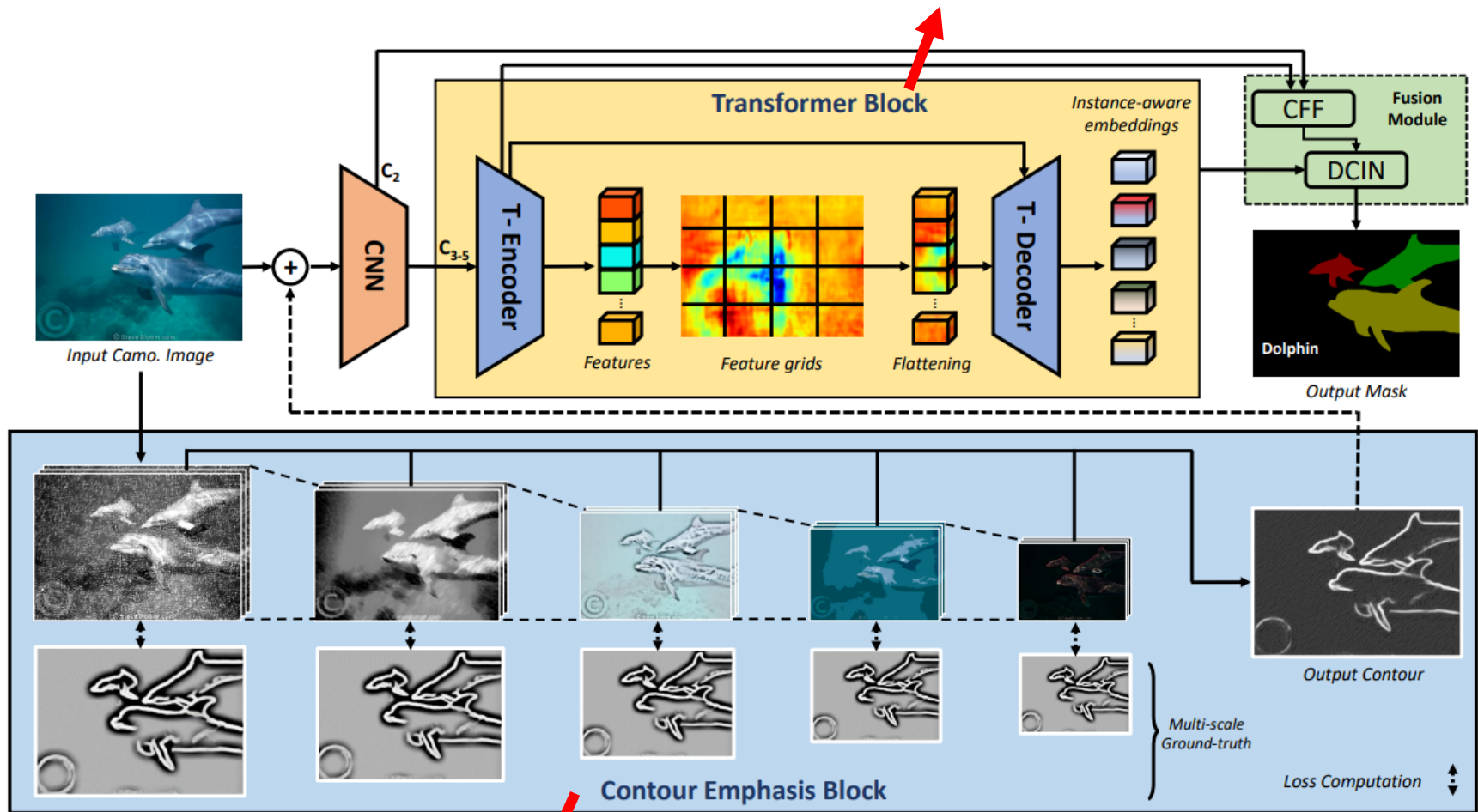
Nội dung

1. Giới thiệu đề tài
2. Các công trình liên quan
- 3. Phương pháp CE-OST khai thác đặc trưng biên cạnh**
4. Tập dữ liệu CAMO-FS
5. Phương pháp FS-CDIS học đặc trưng phân biệt với ít mẫu
6. Kết luận và hướng phát triển

Phương pháp CE-OST



Khối phân đoạn thực thể

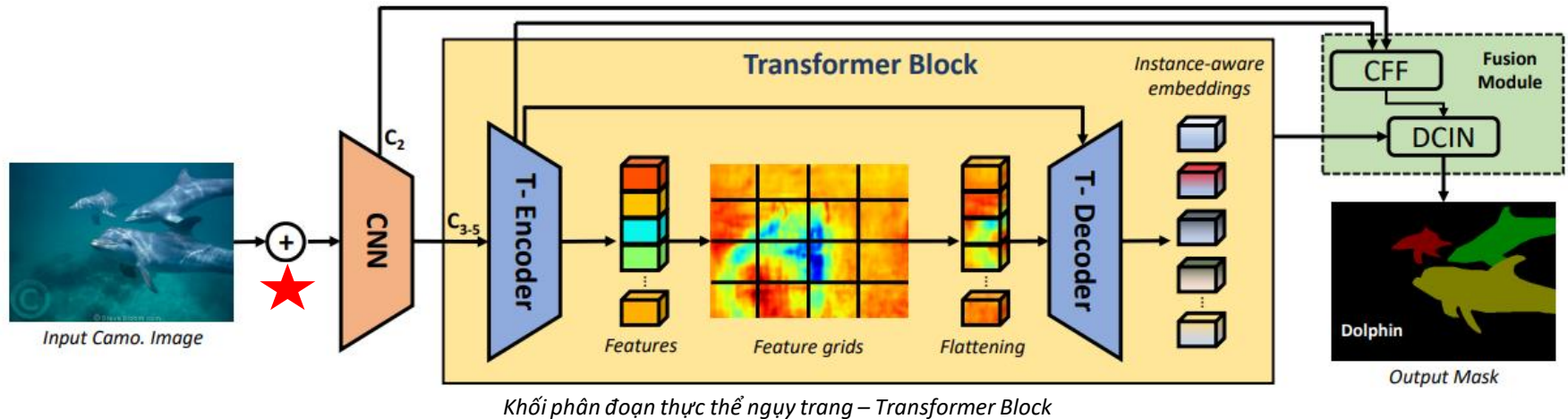


Khối tăng cường đặc trưng biên cạnh

CE-OST – Khối Transformer



Khối Transformer phân đoạn thực thể ngụy trang:



Kế thừa OSFormer*:

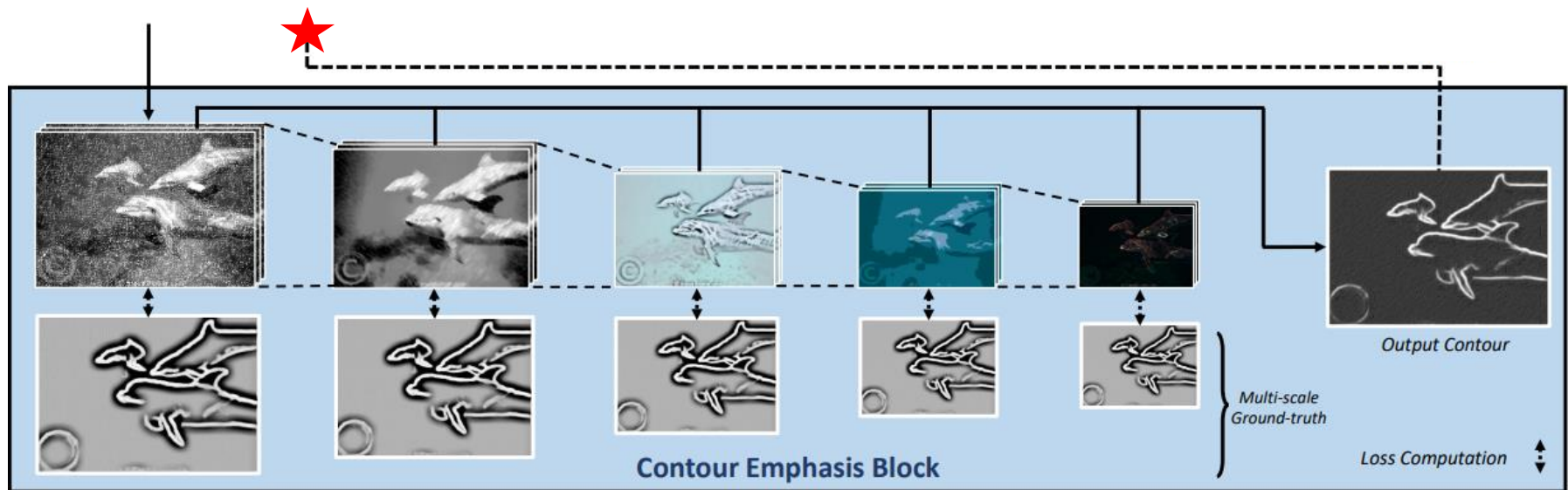
1. **Cơ chế chia lưới** → Khai thác thông vị trí chính xác
2. Các mô-đun hậu xử lý (CFF và DCIN) → **tổng hợp đặc trưng phân tầng ở nhiều mức độ ngữ nghĩa** → Tăng mức độ chi tiết của kết quả phân đoạn

*J. Pei et al., "OSFormer: One-Stage Camouflaged Instance Segmentation with Transformers", ECCV2022

CE-OST – Khối tăng cường biên cạnh



Khối tăng cường đặc trưng biên cạnh:



Khối tăng cường đặc trưng biên cạnh – Contour Emphasis Block

Mục tiêu: làm nổi bật đặc trưng vùng biên cạnh

- Sử dụng mô hình **CNN đa kích thước** để học phát hiện biên cạnh*
- **Biên cạnh được tăng cường** cho ảnh đầu vào thông qua cơ chế **Grid-Condition**

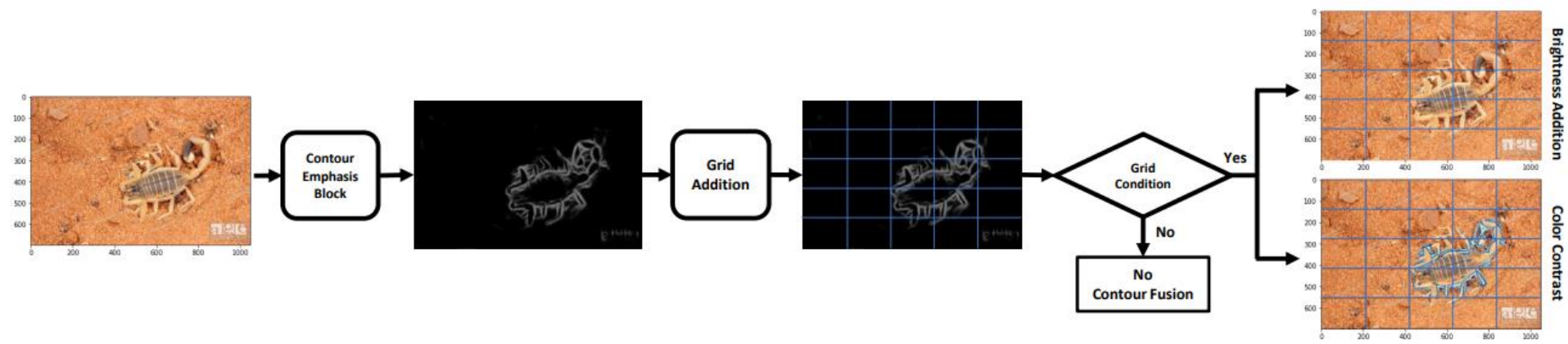
*S. Xie and Z. Tu, "Holistically-nested edge detection", ICCV2015

Cơ chế Grid-Condition



Grid-Condition: đặt điều kiện để xác định có/không tăng cường biên cạnh cho ảnh đầu vào.

- **Mục tiêu:** loại bỏ các trường hợp phát hiện biên cạnh sai, gây nhiễu cho ảnh ngụy trang
- Phương pháp tăng cường biên cạnh:
 1. **Brightness Addition:** cộng giá trị biên cạnh tương ứng
 2. **Color Contrast:** cộng giá trị màu tương phản



Kết quả thực nghiệm CE-OST



So sánh hiệu năng trên tập dữ liệu COD10K và NC4K

Phương pháp		COD10K			NC4K		
		AP	AP50	AP75	AP	AP50	AP75
Two-Stage	Mask R-CNN [28]	28.7	60.1	25.7	36.1	68.9	33.5
	MS R-CNN [30]	33.3	61.0	32.9	35.7	63.4	34.7
	Cascade R-CNN [2]	29.5	61.0	25.9	34.6	66.3	31.5
	HTC [6]	30.9	61.0	28.7	34.2	64.5	31.6
	BlendMask [5]	31.2	60.0	28.9	31.4	61.2	28.8
	Mask Transfuser [33]	31.2	60.7	29.8	34.0	63.1	32.6
One-Stage	YOLACT [1]	29.0	60.1	25.3	37.8	70.6	35.6
	CondInst [71]	34.3	67.9	31.6	38.0	71.1	35.6
	QueryInst [17]	32.5	65.1	28.6	38.7	72.1	37.6
	SOTR [23]	32.0	63.6	29.2	34.3	65.7	32.4
	SOLO [78]	35.2	65.7	33.4	37.8	69.2	36.1
	OSFormer [59]	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

- CE-OST đạt hiệu năng cao nhất với **43.2% AP trên COD10K** và **45.1% AP trên NC4K** nhờ phương pháp tăng cường đặc trưng biên cạnh

Kết quả thực nghiệm CE-OST (tt)



Phương pháp	Mô hình cơ sở	COD10K			NC4K			CAMO++		
		AP	AP50	AP75	AP	AP50	AP75	AP	AP50	AP75
OSFormer	ResNet-50 [26]	41.0	71.1	40.8	42.5	72.5	42.3	19.0	33.8	18.3
	ResNet-50-550 [26]	-	-	-	-	-	-	20.1	36.3	19.3
	ResNet-101 [26]	42.0	71.3	42.8	44.4	73.7	45.1	20.6	34.4	20.2
	PVTv2-B2-Li [76]	47.2	74.9	49.8	-	-	-	27.7	44.7	27.9
	Swin-T [46]	47.7	78.6	49.3	-	-	-	22.3	36.6	21.8
CE-OST (Color Contrast)	ResNet-50 [26]	41.6	70.7	42.3	42.4	71.4	42.6	20.1	34.2	19.6
	ResNet-50-550 [26]	35.9	65.2	34.3	41.1	70.9	41.1	20.6	35.7	20.0
	ResNet-101 [26]	43.2	72.2	44.1	45.1	74.0	46.4	21.7	36.6	21.3
	PVTv2-B2-Li [76]	48.4	75.7	51.3	51.4	77.9	55.0	28.5	45.3	29.9
	Swin-T [46]	49.1	78.0	52.1	50.5	78.9	53.1	22.7	37.6	22.4
CE-OST (Brightness Addition)	ResNet-50 [26]	41.2	69.0	41.6	42.4	71.1	42.9	20.2	34.8	19.5
	ResNet-50-550 [26]	35.9	65.2	34.6	40.8	71.1	40.3	21.0	37.1	20.3
	ResNet-101 [26]	42.4	70.8	43.7	44.2	73.1	45.0	21.1	34.4	20.9
	PVTv2-B2-Li [76]	47.9	74.6	50.5	51.1	77.3	54.9	27.9	45.1	29.2
	Swin-T [46]	49.0	78.5	51.4	50.8	79.3	53.9	22.7	38.4	23.1

Thực nghiệm loại suy với các tập dữ liệu COD10K, NC4K, và CAMO++.



Trực quan hóa kết quả trên **CAMO++**

- **CE-OST** cải thiện so với baseline OSFormer trên nhiều mô hình cơ sở khác nhau
- **PVT** là mô hình cơ sở **có kết quả tốt ổn định nhất** trên 3 tập dữ liệu nhờ vào bộ trích xuất đặc trưng đa kích thước (thể hiện rõ trên tập CAMO++)

*Kết quả tốt thứ nhất, thứ hai và thứ ba lần lượt được kí hiệu với màu đỏ, xanh dương, và xanh lá.



Nội dung

1. Giới thiệu đề tài
2. Các công trình liên quan
3. Phương pháp CE-OST khai thác đặc trưng biên cạnh
- 4. Tập dữ liệu CAMO-FS**
5. Phương pháp FS-CDIS học đặc trưng phân biệt với ít mẫu
6. Kết luận và hướng phát triển



Thực trạng dữ liệu nguy trang

Động lực:

- Số tập dữ liệu cũng như số mẫu dữ liệu ảnh nguy trang còn ít
- Ngữ cảnh huấn luyện ít mẫu dữ liệu nguy trang **cần tập dữ liệu đặc thù**

Tập dữ liệu	Năm	Hội nghị	Loại	#Gán nhãn nguy trang	#Lớp tổng quát	#Lớp đối tượng	#Thực thể hoặc #Đối tượng trên ảnh	Nhãn khung bao	Nhãn ngữ nghĩa	Nhãn thực thể	Few-shot
CamouflagedAnimals [60]	2016	ECCV	Video	181	-	6	1.238	×	✓	✓	×
MoCA [35]	2020	ACCV	Video	7,617	-	67	1.000	✓	×	×	×
CHAMELEON [69]	2018	-	Ảnh	76	-	-	1.000	×	✓	×	×
CAMO [38]	2019	CVIU	Ảnh	1,250	2	8	1.000	×	✓	×	×
COD [12]	2020	CVPR	Ảnh	5,066	5	69	1.171	✓	✓	✓	×
NC4K [49]	2021	CVPR	Ảnh	4,121	5	69	1.171	✓	✓	✓	×
CAMO++ [36]	2022	TIP	Ảnh	2,695	10	47	1.171	✓	✓	✓	×
CAMO-FS	2023	-	Ảnh	2,858	10	47	1.172	✓	✓	✓	✓

Thống kê một số tập dữ liệu về thực thể nguy trang

→ Đề xuất tập dữ liệu CAMO-FS

- Tập dữ liệu đầu tiên** phục vụ phân đoạn thực thể nguy trang với ít mẫu huấn luyện
- CAMO-FS** thừa kế tập dữ liệu CAMO++* với sự đa dạng về số lớp đối tượng

*Le, T.-N. et al., "Camouflaged instance segmentation in-the-wild", IEEE TIP 2022

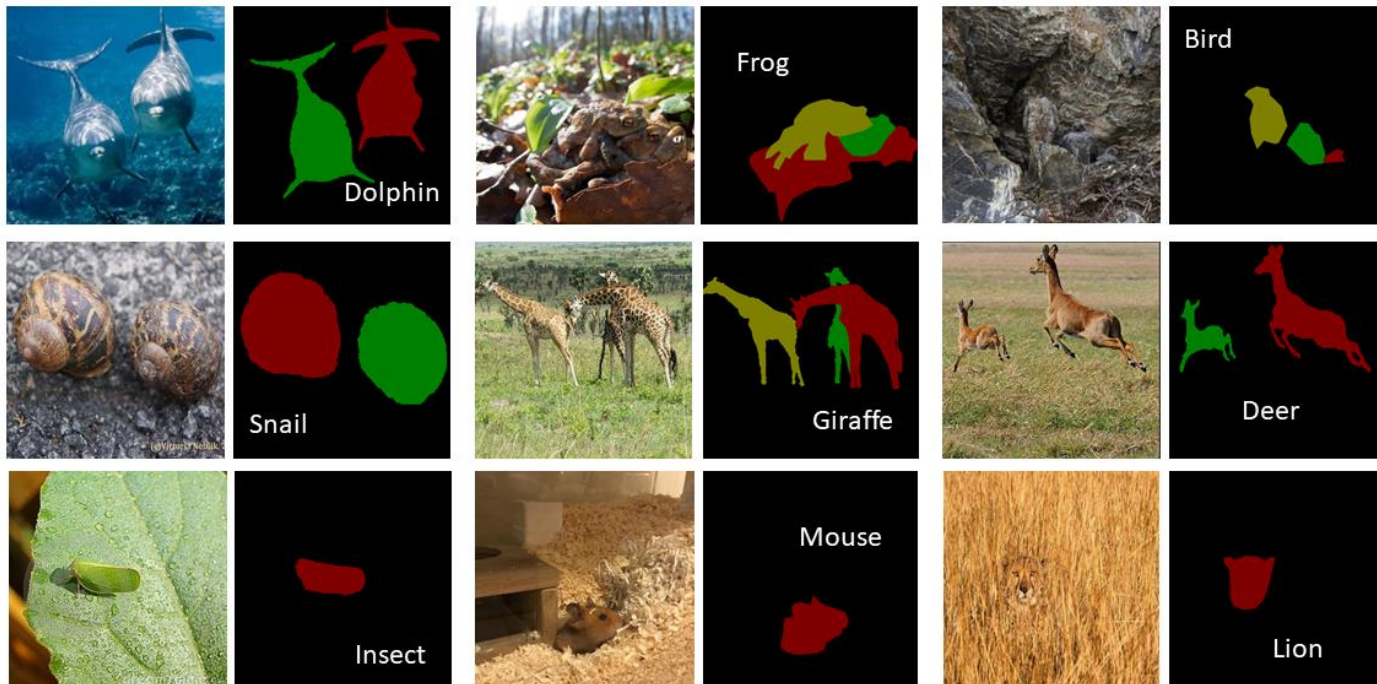


Tập dữ liệu CAMO-FS

CAMO-FS có:

- **2,858 ảnh/ 3,342 thực thể** ngẫu trang, **tăng 163 ảnh so với CAMO++***
- **47 lớp ngữ nghĩa** tạo lập cấu trúc phục vụ bài toán học với ít dữ liệu huấn luyện ($K=5$)

CAMO-FS tiên phong giải quyết bài toán trên thực thể ngẫu trang với **ít dữ liệu huấn luyện**



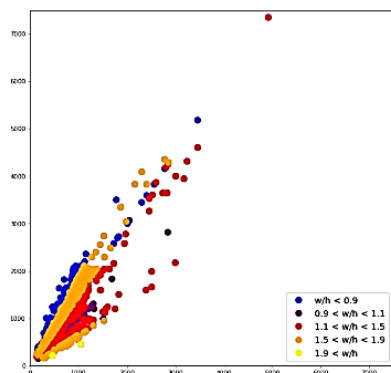
Mình họa dữ liệu của tập CAMO-FS

*Le, T.-N. et al., "Camouflaged instance segmentation in-the-wild", IEEE TIP 2022

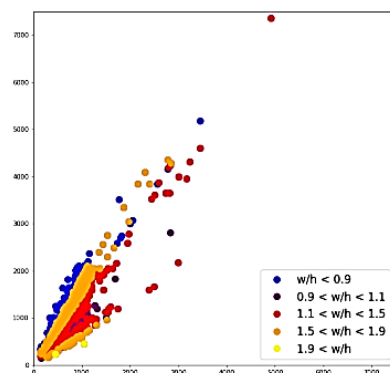
Một số phân tích về dữ liệu*



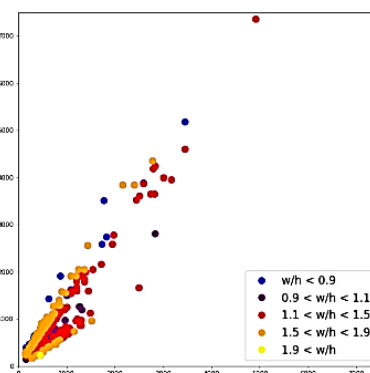
Phân phối độ phân giải ảnh:



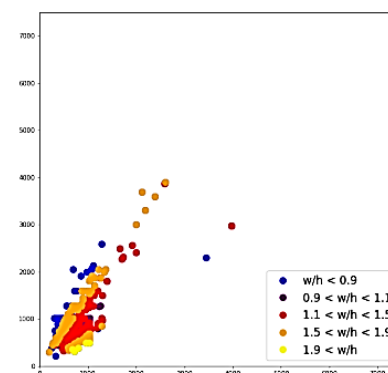
(a) CAMO-FS đề xuất



(b) CAMO++ [36]

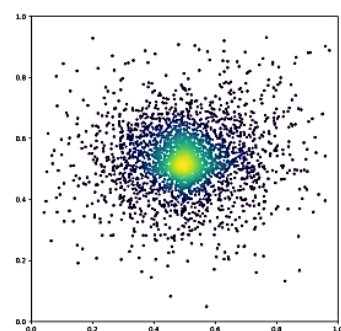


(c) CAMO [38]

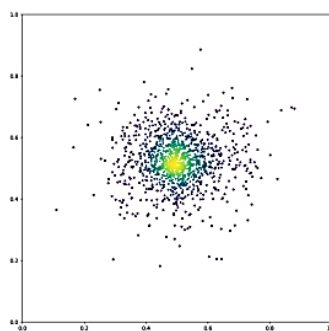


(d) COD10K [12]

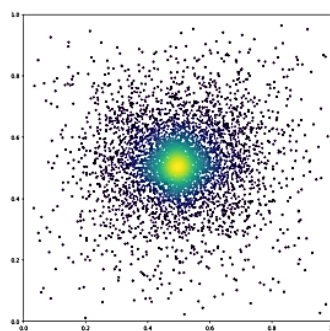
Phân phối độ lệch tâm:



(a) CAMO-FS



(b) CAMO [38]



(c) COD10K [12]

Phân phối số lượng thực thể/ảnh

Số lượng thực thể	Tỉ lệ (%)	Số lượng ảnh
1	90.5	2581
2	1.05	190
3	1.79	51
3+	6.66	30

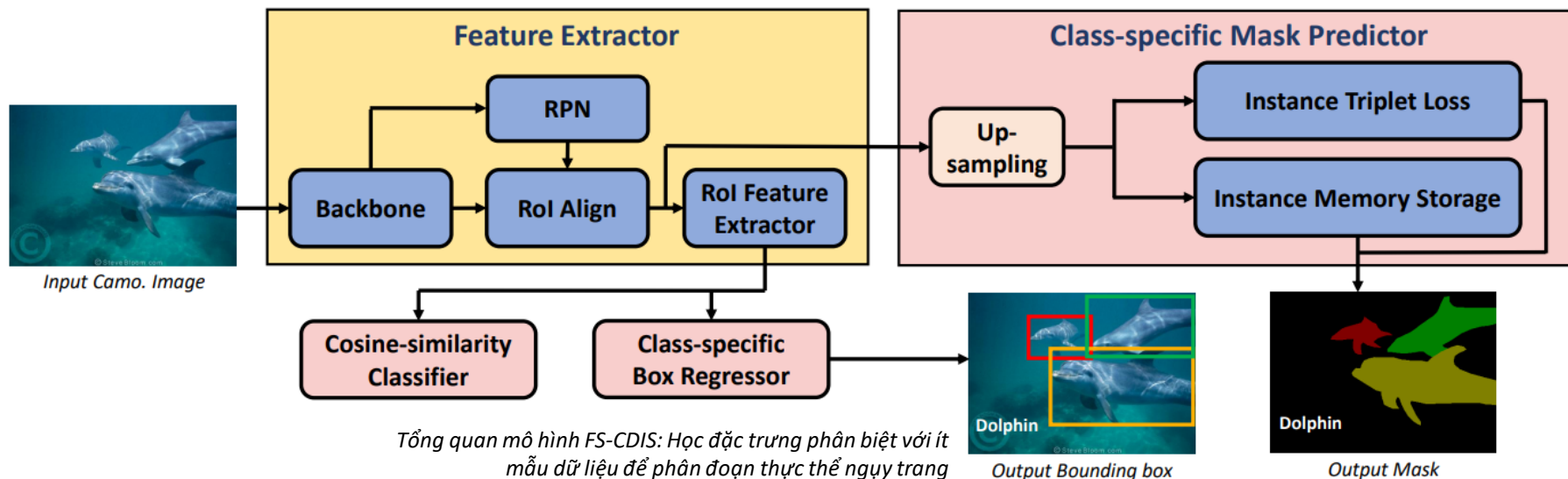
*Chỉ so sánh dữ liệu ảnh có chứa thực thể nguy trang



Nội dung

1. Giới thiệu đề tài
2. Các công trình liên quan
3. Phương pháp CE-OST khai thác đặc trưng biên cạnh
4. Tập dữ liệu CAMO-FS
- 5. Phương pháp **FS-CDIS** học đặc trưng phân biệt với ít mẫu**
6. Kết luận và hướng phát triển

Phương pháp FS-CDIS

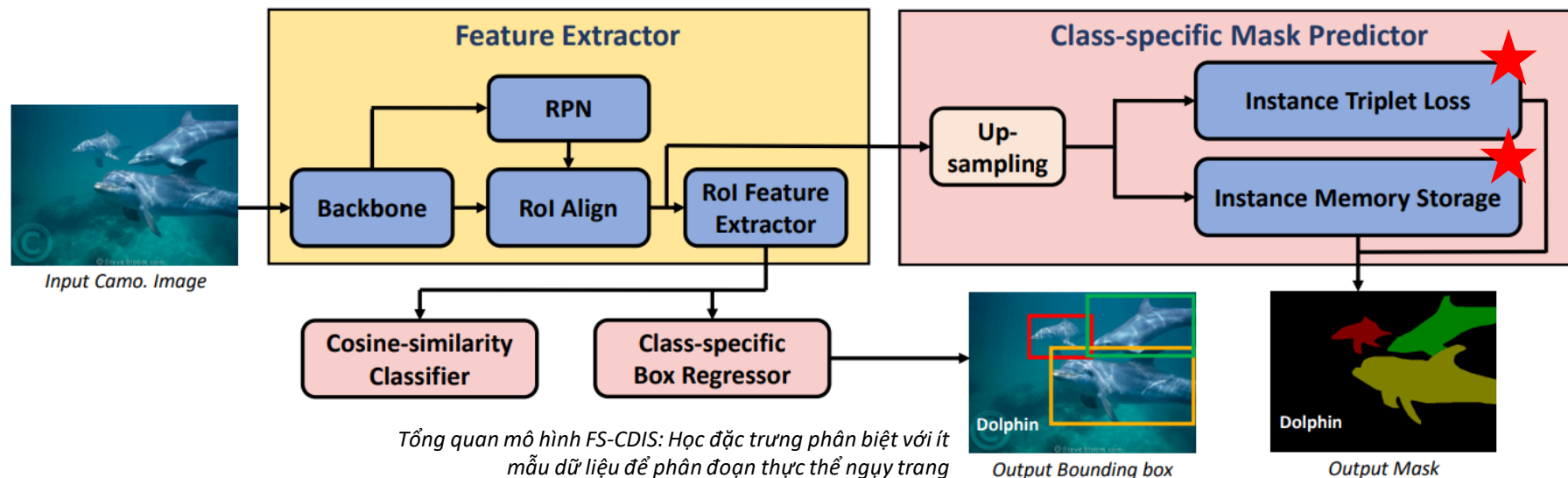


Vấn đề: Dữ liệu ngụy trang gán nhãn **có giới hạn**

→ Cần mô hình có **khả năng học hiệu quả đặc trưng phân biệt** từ ít dữ liệu

Giải pháp: **học tương phản với ít mẫu**

Phương pháp FS-CDIS



FS-CDIS kế thừa framework mTFA* với kiến trúc 2 giai đoạn

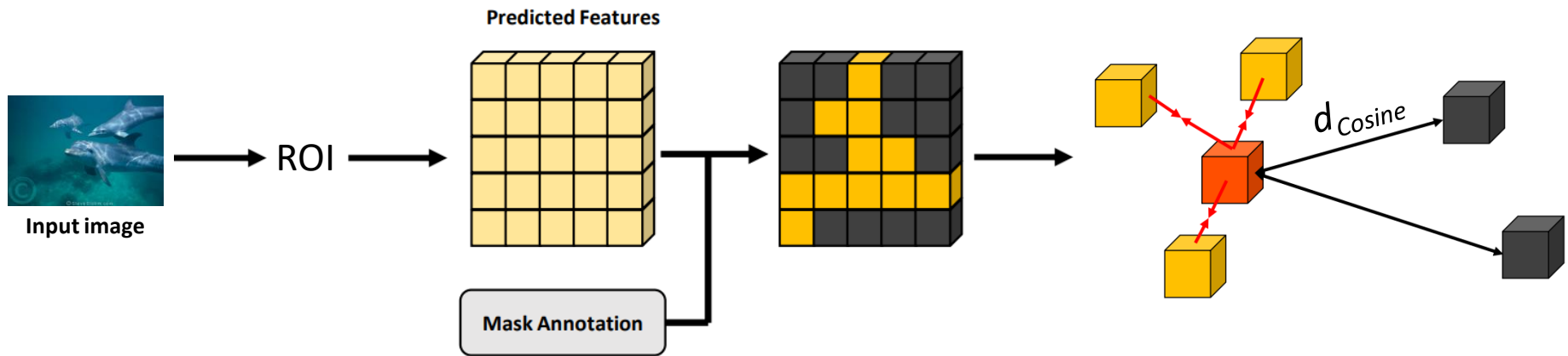
Mục tiêu: học tương phản với các cải tiến sau

- **Instance Triplet Loss** – Hàm mất mát ba thành phần cấp độ thực thể
- **Instance Memory Storage** – Bộ nhớ lưu trữ thực thể

*Ganea, D.A. et al., “Incremental few-shot instance segmentation”, CVPR 2021



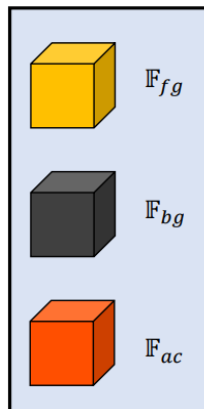
FS-CDIS – Instance Triplet Loss



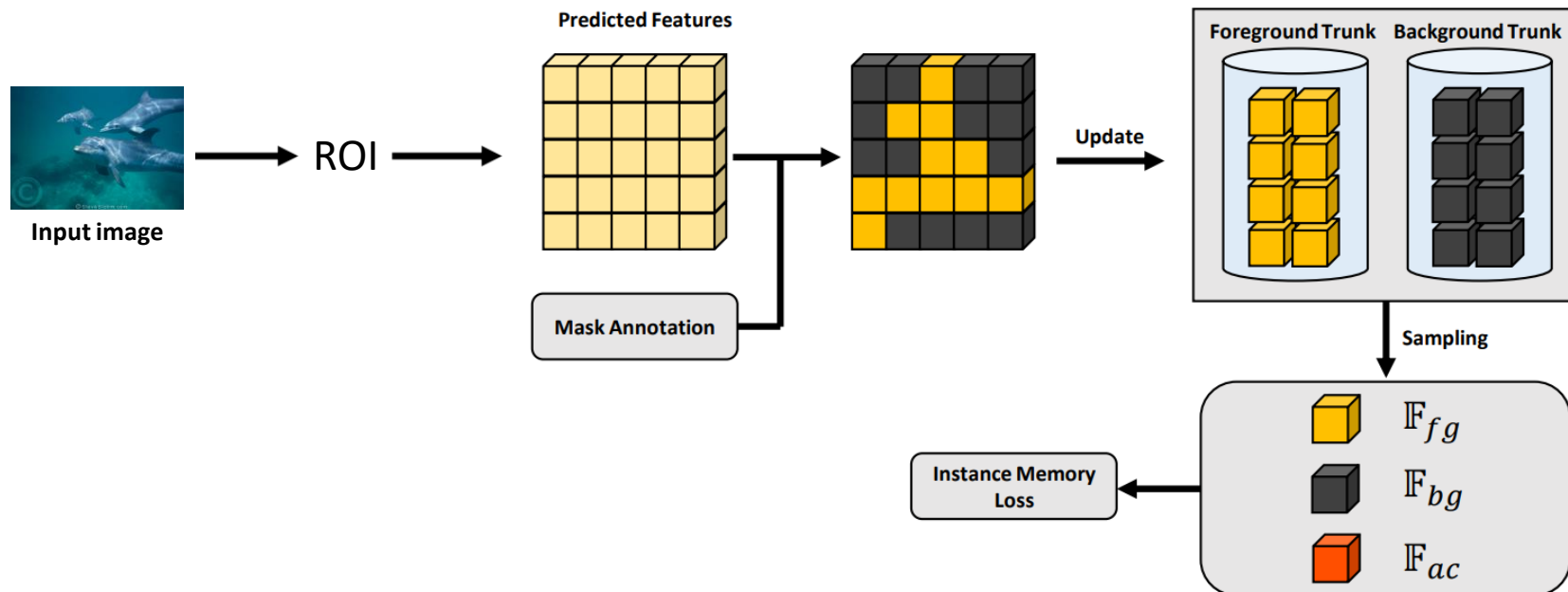
Minh họa ý tưởng Instance Triplet Loss

Instance Triplet Loss: giúp phân biệt vùng nền và vùng thực thể ngụy trang

- Tối thiểu hóa khoảng cách giữa các điểm vùng thực thể (\mathbb{F}_{fg})
- Tối đa hóa khoảng cách giữa các điểm vùng thực thể (\mathbb{F}_{fg}) và vùng nền (\mathbb{F}_{bg})
- Sử dụng độ đo Cosine để tính khoảng cách



FS-CDIS – Instance Memory Storage



Minh họa ý tưởng Instance Memory Storage

Instance Memory Storage: Lưu trữ các đặc trưng **vùng thực thể** và **vùng nền** theo mỗi lớp ngữ nghĩa để **tính toán sự khác biệt** của các đặc trưng trích xuất từ các vùng đề xuất

Kết quả thực nghiệm FS-CDIS



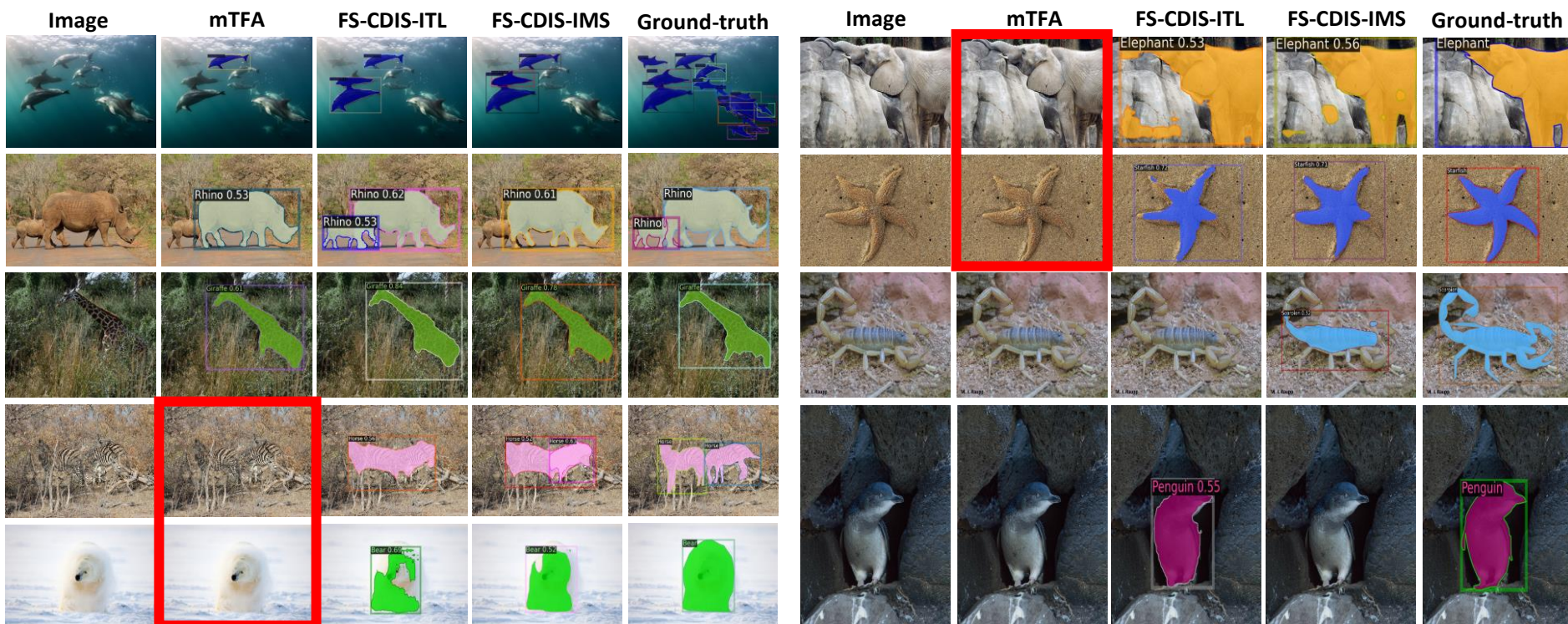
So sánh hiệu năng các phương pháp trên tập dữ liệu CAMO-FS.

“IMS” đại diện cho Instance Memory Storage , “ITL” đại diện cho Instance Triplet Loss.

Mô hình		AP giai đoạn tinh chỉnh									
Phương pháp	Mô hình cơ sở	Phân đoạn thực thể					Phát hiện đối tượng				
		1	2	3	5	Avg.	1	2	3	5	Avg.
MTFA [20]	COCO-80 ResNet-50	2.48	6.67	5.81	6.40	5.34	1.98	6.47	5.82	6.17	5.11
M-RCNN [†] [28]		4.08	6.79	6.90	8.29	6.52	2.82	5.09	5.46	6.18	4.89
iFS-RCNN [56]		4.17	6.26	5.73	6.38	5.64	3.92	6.06	5.47	6.60	5.51
MTFA [20]	COCO-80 ResNet-101	3.66	6.21	6.16	5.95	5.50	2.93	5.90	5.84	5.84	5.13
M-RCNN [†] [28]		4.39	7.69	7.94	10.09	7.53	3.03	5.80	6.20	7.79	5.71
iFS-RCNN [56]		4.27	6.55	6.07	7.80	6.17	3.79	6.28	6.01	8.08	6.04
Các phương pháp đề xuất (↑)											
FS-CDIS-ITL	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	MTFA	5.46	6.95	7.36	9.61	7.35	4.50	6.95	7.55	10.36	7.34
FS-CDIS-ITL	ResNet-101	5.73	7.97	8.52	9.92	8.04	5.08	7.56	7.85	9.67	7.34
FS-CDIS-IMS	M-RCNN	5.52	7.84	8.65	9.82	7.96	4.92	7.39	7.96	9.52	7.45
FS-CDIS-ITL	ResNet-101	5.35	6.01	7.80	6.23	6.35	4.71	5.66	7.10	6.06	5.88
FS-CDIS-IMS	iFS-RCNN	2.99	6.83	6.14	9.03	6.25	2.74	6.39	5.94	8.44	5.88

- FS-CDIS **cải thiện hiệu năng so với mô hình cơ sở** (cột Avg.)
- Việc học phân biệt đặc trưng vùng thực thể và vùng nền giúp **cải thiện hiệu năng với ít mẫu huấn luyện** $K = \{1, 2, 3, 5\}$

Kết quả thực nghiệm FS-CDIS (tt)



Trực quan hóa các kết quả thực nghiệm trên tập dữ liệu CAMO-FS.

“IMS” đại diện cho Instance Memory Storage và “ITL” đại diện cho Instance Triplet Loss

- FS-CDIS +ITL/IMS **cải thiện được các trường hợp mTFA không phát hiện được** thực thể
- FS-CDIS **chưa xử lý tốt ảnh có vùng nền phức tạp**

Kết luận



- Luận văn nghiên cứu bài toán phân đoạn thực thể nguy trang với **3 đề xuất**:
 1. **Mô hình CE-OST** ^[CT1] tận dụng hiệu quả đặc trưng biên cạnh thực thể
 2. **Tập dữ liệu ảnh CAMO-FS** ^[CT2, CT3] cho hướng tiếp cận học ít dữ liệu
 3. **Mô hình FS-CDIS** ^[CT2, CT3] kỹ thuật học tương phản với ít mẫu huấn luyện

Hướng phát triển



1. Cải tiến các đặc trưng có tính phân biệt cao

Các đặc trưng có tính phân biệt cao **không giới hạn ở biên cạnh thực thể**

➔ Khai thác đặc trưng đặc thù của mỗi thực thể thuộc cùng một lớp ngữ nghĩa, đặc trưng đặc thù của vùng nền, hay yếu tố ngữ cảnh

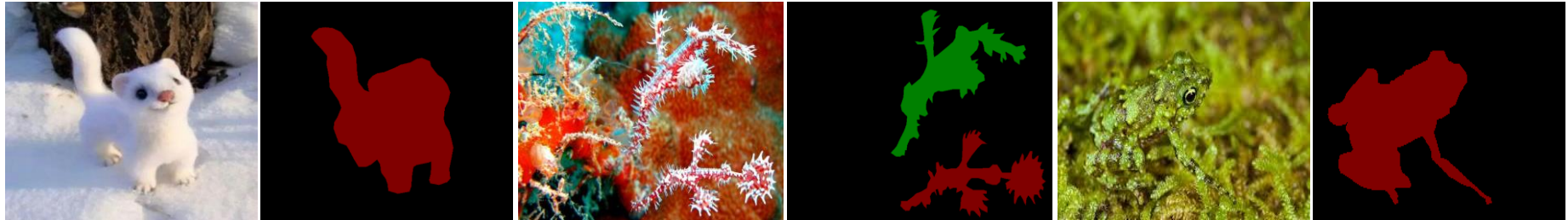
2. Áp dụng hướng tiếp cận cho bài toán trên ảnh y khoa

- Các thực thể trong ảnh y khoa như bộ phận nội tạng, các vùng bệnh, khối u cũng chứa các loại đặc trưng ngữ trang tương tự với **màu sắc, hình dáng tương đồng** với các bộ phận, tế bào khác



LUẬN VĂN THẠC SĨ
NGÀNH KHOA HỌC MÁY TÍNH

PHÂN ĐOẠN THỰC THỂ NGUY TRẠNG DỰA TRÊN ĐẶC TRƯNG CÓ TÍNH PHÂN BIỆT CAO



NGUYỄN THÀNH DANH – 210101001
GVHD: TS. NGUYỄN VINH TIỆP



VINIF

