



Unlocking Attributes' Contribution to Successful Camouflage: A Combined Textual and Visual Analysis Strategy

Hong Zhang, Yixuan Lyu, Qian Yu, Hanyang Liu, Huimin Ma, Yuan Ding, Yifan Yang*



The image displays a creature resembling algae, integrating seamlessly with the marine environment, exhibiting an intricate structure of appendages, positioned against a blurry ocean floor background.





The image shows a crab camouflaged against a sandy background. Its color and texture blend with the surrounding sand, making it difficult to distinguish.



Repo: https://github.com/lyu-yx/ACUMEN Paper: https://arxiv.org/abs/2408.12086

1. Introduction

1.1 Problem Statement

General detection/segmentation algorithms presented slowing of performance lift.

Main reason: have troubles when object feature become less representative or camouflaged by themselves.





1. Introduction

1.2 Related Works

- Approaches based on hand-crafted features and deep learning methods have achieved astonishing performance in **visual modalities**.
- However, the combined use of **textual and visual modalities** to **enhance performance and understanding of camouflage patterns** has not yet been explored.

1.3 Our Solution

• We commence by **collecting a dataset** enriched with image descriptions and attribute contributions. Subsequently, we construct a **bifurcated multimodal framework** that merges textual and visual analyses seamlessly.

2. Motivation

- 1. From cognitive science point of view, merging **textual and visual** information synergistically boosts cognitive understanding[1-2].
- 2. Evolutionary biology highlights the significance of camouflage pattern creation (by prey) and its identification (by predators) in evolutionary progress, underlining the necessity to analyze camouflage from both **granular attribute insights (designing)** and a wider **object detection (breaking)** standpoint.



- Environmental Pattern Matching;
- Color Matching;
- Shading.



- Shape Mimicry;
- Environmental Textures;
- Color Matching.

Mayer, R.E.: Multimedia learning. In: Psychology of learning and motivation, vol. 41, pp. 85–139 (2002)
 Paivio, A.: Imagery and verbal processes. Psychology Press (2013)

3. COD-TAX Dataset





The image displays a creature resembling algae, integrating seamlessly with the marine environment, exhibiting an intricate structure of appendages, positioned against a blurry ocean floor background.



Surrounding Factors (5 sub attr), Camouflaged Object-Self Reasons (6 sub attr), and Imaging Quality Reasons (6 sub attr). 17 attributes in total.



- The range of maximum values extends from 0.21 to 0.55.
- The average values fluctuate between 0.004 and 0.21.
- Average description length of 26.52 words.

All the descriptions and attributions are generated by GPT-4V first and finetuned by more than 30 volunteers.

4. Network Overall



ACUMAN presented a dual-branch architecture, consisting of a textual branch (in green) and a visual branch (in cyan).

During the inference, **the textual branch is omitted** to eliminate dependency on LVLMs like GPT4, thereby making the inference process solely reliant on visual cues.

5. **Results**

	5 5	5	\$ -	\$.		atoto	5	5	5	Š 🕺			\$	ş
	Ť		5		5			H		Ť		Ķ	,	Ý
		Jel		X		X		S.	14	Ľ	J.	N.	-J.C	Z
			7							2				
Input	GT	Ours	;	MRR]	FSPNet		PopNet	Ex	plicit	FE	DER	FPN	Net
Input	GT	Ours		MRR] мо	FSPNet		PopNet	Ех	plicit	FE	DER	FPN	Net
Input Methods	GT	Ours	$S_{\alpha} \uparrow$	$\frac{MRR}{CA}$	$\frac{MO}{F^{\omega}_{\beta}}\uparrow$	FSPNet M↓	$S_{\alpha} \uparrow$	$\frac{\text{COI}}{E_{\phi}} \uparrow$	Explock $F^{\omega}_{\beta} \uparrow$	M ↓	FE	$\frac{\text{DER}}{E_{\phi}}$	FPN $\frac{24K}{F_{\beta}^{\omega}} \uparrow$	Net M↓
Input Methods PopNet [47]	GT Publication ICCV2023	Ours Size	S_{α} \uparrow 0.806	$\frac{MRR}{E_{\phi}\uparrow}$ 0.859*	$\frac{MO}{F_{\beta}^{\omega}\uparrow}$ 0.744*	M ↓	$S_{\alpha} \uparrow$ 0.827	PopNet COI $E_{\phi} \uparrow$ 0.910^*	E_{β}	<i>M</i> ↓ 0.031	FE $S_{\alpha} \uparrow$ 0.852	$\frac{\text{DER}}{E_{\phi} \uparrow}$ 0.909*	FPN $C4K$ $F^{\omega}_{\beta} \uparrow$ 0.802^{*}	<i>M</i> ↓ 0.043
Input Methods PopNet [47] CFANet [52]	GT Publication ICCV2023 ICME2023	Ours 512 ² 416 ²	$S_{\alpha} \uparrow$ 0.806 0.815	$\begin{array}{c} \text{MRR} \\ \hline \\ CA \\ E_{\phi} \uparrow \\ 0.859^{*} \\ 0.876 \end{array}$	$\frac{\text{MO}}{F_{\beta}^{\omega}}\uparrow$ 0.744^{*} 0.761	FSPNet M ↓ 0.073 0.073	$S_{\alpha} \uparrow$ 0.827 0.834	PopNet COT $E_{\phi} \uparrow$ 0.910^{*} 0.905	Ex $\frac{D10K}{F_{\beta}^{\omega}} \uparrow$ 0.757^{*} 0.730	<i>M</i> ↓ 0.031 0.031	FE $S_{\alpha} \uparrow$ 0.852 0.848	EDER $E_{\phi} \uparrow$ 0.909^* 0.906	FPN $\frac{C4K}{F_{\beta}^{\omega}} \uparrow$ 0.802^{*} 0.791	M ↓ 0.043 0.046
Input Methods PopNet [47] CFANet [52] MFFN [54]	GT Publication ICCV2023 ICME2023 WACV2023	Ours 512 ² 416 ² 384 ²	$S_{\alpha} \uparrow$ 0.806 0.815 †	$\begin{array}{c} \text{MRR} \\ \hline \\ E_{\phi} \uparrow \\ 0.859^{*} \\ 0.876 \\ \dagger \end{array}$	$\frac{\text{MO}}{F_{\beta}^{\omega} \uparrow}$ 0.744^{*} 0.761 \dagger	M ↓ 0.073 0.073 †	$S_{\alpha} \uparrow$ 0.827 0.834 0.846	PopNet COI $E_{\phi} \uparrow$ 0.910* 0.905 0.897*	Exp F_{β}^{ω} \uparrow 0.757* 0.730 0.745	M ↓ 0.031 0.028	Ff $S_{\alpha} \uparrow$ 0.852 0.848 0.856	EDER $E_{\phi} \uparrow$ 0.909* 0.906 0.902*	FPN $F_{\beta}^{\omega} \uparrow$ 0.802* 0.791 0.791	M ↓ 0.043 0.046 0.042
Input Methods PopNet [47] CFANet [52] MFFN [54] FEDER [12]	GT Publication ICCV2023 ICME2023 WACV2023 CVPR2023	Ours 512 ² 416 ² 384 ² 384 ²	$S_{\alpha} \uparrow$ 0.806 0.815 † 0.807	MRR CA $E_{\phi} \uparrow$ 0.859* 0.876 \dagger 0.873	$\frac{MO}{F_{\beta}^{\omega}\uparrow}$ 0.744* 0.761 † 0.738*	M ↓ 0.073 0.073 † 0.069	$S_{\alpha} \uparrow$ 0.827 0.834 0.846 0.823	PopNet $E_{\phi} \uparrow$ 0.910* 0.897* 0.900	Ex F_{β}^{ω} \uparrow 0.757* 0.730 0.745 0.716*	M ↓ 0.031 0.028 0.032	Ff $S_{\alpha} \uparrow$ 0.852 0.848 0.856 0.846	EDER $E_{\phi} \uparrow$ 0.909* 0.906 0.902* 0.905	FPN $F_{\beta}^{\omega} \uparrow$ 0.802* 0.791 0.791 0.789*	M ↓ 0.043 0.046 0.042 0.045
Input Methods PopNet [47] CFANet [52] MFFN [54] FEDER [12] Explicit [26]	GT Publication ICCV2023 ICME2023 WACV2023 CVPR2023 CVPR2023	Ours 512 ² 416 ² 384 ² 384 ² 352 ²	$S_{\alpha} \uparrow$ 0.806 0.815 \dagger 0.807 0.846	MRR $E_{\phi} \uparrow$ 0.859* 0.876 † 0.873 0.895	$\begin{array}{c} \text{MO} \\ \hline F_{\beta}^{\omega} \uparrow \\ 0.744^{*} \\ 0.761 \\ \dagger \\ 0.738^{*} \\ 0.777 \end{array}$	M ↓ 0.073 0.073 † 0.069 0.059	$S_{\alpha} \uparrow$ 0.827 0.834 0.846 0.823 0.843	PopNet COI $E_{\phi} \uparrow$ 0.910* 0.905 0.897* 0.900 0.907	Ex F_{β}^{ω} \uparrow 0.757* 0.730 0.745 0.716* 0.742	M↓ 0.031 0.031 0.032 0.032	Ff $S_{\alpha} \uparrow$ 0.852 0.848 0.856 0.846 †	EDER $E_{\phi} \uparrow$ 0.909* 0.906 0.902* 0.905 †	FPN $F_{\beta}^{\omega} \uparrow$ 0.802* 0.791 0.791 0.789* †	M ↓ 0.043 0.046 0.042 0.045 †
Input Methods PopNet [47] CFANet [52] MFFN [54] FEDER [12] Explicit [26] FSPNet [18]	GT Publication ICCV2023 ICME2023 WACV2023 CVPR2023 CVPR2023 CVPR2023	Ours 512 ² 416 ² 384 ² 384 ² 352 ² 384 ²	$S_{\alpha} \uparrow$ 0.806 0.815 \uparrow 0.807 0.846 0.856	MRR $E_{\phi} \uparrow$ 0.859* 0.876 \dagger 0.873 0.895 0.899	$\begin{array}{c} \text{MO} \\ \hline F_{\beta}^{\omega} \uparrow \\ 0.744^{*} \\ 0.761 \\ \dagger \\ 0.738^{*} \\ 0.777 \\ 0.799 \end{array}$	M ↓ 0.073 0.073 † 0.069 0.059 0.050	$S_{\alpha} \uparrow$ 0.827 0.834 0.846 0.823 0.843 0.843	PopNet $E_{\phi} \uparrow$ 0.910* 0.905 0.897* 0.900 0.907 0.895	Ex F_{β}^{ω} \uparrow 0.757* 0.730 0.745 0.716* 0.742 0.735	multiplicit M↓ 0.031 0.031 0.028 0.032 0.029 0.026	Ff $S_{\alpha} \uparrow$ 0.852 0.848 0.856 0.846 \dagger 0.879	EDER $E_{\phi} \uparrow$ 0.909* 0.906 0.902* 0.905 † 0.915	FPN $F_{\beta}^{\omega} \uparrow$ 0.802* 0.791 0.791 0.789* † 0.816	M ↓ 0.043 0.046 0.042 0.045 † 0.035
Input Methods PopNet [47] CFANet [52] MFFN [54] FEDER [12] Explicit [26] FSPNet [18] MRR-Net [49]	GT Publication ICCV2023 ICME2023 WACV2023 CVPR2023 CVPR2023 CVPR2023 TNNLS2023	Ours 512 ² 416 ² 384 ² 352 ² 384 ² 384 ² 384 ² 384 ²	$S_{\alpha} \uparrow$ 0.806 0.815 \uparrow 0.807 0.846 0.856 0.826	MRR $E_{\phi} \uparrow$ 0.859* 0.876 \dagger 0.873 0.895 0.899 0.880	$\begin{array}{r} \text{MO} \\ \hline F_{\beta}^{\omega} \uparrow \\ 0.744^{*} \\ 0.761 \\ \dagger \\ 0.738^{*} \\ 0.777 \\ 0.799 \\ 0.759^{*} \end{array}$	M ↓ 0.073 0.073 † 0.069 0.059 0.050 0.070	$S_{\alpha} \uparrow$ 0.827 0.834 0.846 0.823 0.843 0.851 0.835	PopNet E_{ϕ} ↑ 0.910* 0.905 0.897* 0.900 0.907 0.895 0.901	Ex F_{β}^{ω} \uparrow 0.757* 0.730 0.745 0.716* 0.742 0.735 0.720*	M ↓ 0.031 0.028 0.032 0.029 0.026 0.032	Ff $S_{\alpha} \uparrow$ 0.852 0.848 0.856 0.846 \dagger 0.846 \dagger 0.857	EDER $E_{\phi} \uparrow$ 0.909* 0.906 0.902* 0.905 † 0.915 0.906	FPN $F\beta \uparrow$ 0.802* 0.791 0.791 0.789* † 0.816 0.786*	M↓ 0.043 0.046 0.042 0.045 † 0.035 0.044
Input Methods PopNet [47] CFANet [52] MFFN [54] FEDER [12] Explicit [26] FSPNet [18] MRR-Net [49] FPNet [4]	GT Publication ICCV2023 ICME2023 WACV2023 CVPR2023 CVPR2023 CVPR2023 TNNLS2023 ACM MM202	Ours 512 ² 416 ² 384 ² 384 ² 384 ² 384 ² 384 ² 384 ² 384 ² 384 ² 384 ²	$S_{\alpha} \uparrow$ 0.806 0.815 \dagger 0.807 0.846 0.856 0.826 0.852	MRR $E_{\phi} \uparrow$ 0.859* 0.876 \dagger 0.873 0.895 0.899 0.880 0.905	$\begin{array}{c} \text{MO} \\ \hline F_{\beta}^{\omega} \uparrow \\ 0.744^{*} \\ 0.761 \\ \dagger \\ 0.738^{*} \\ 0.777 \\ 0.799 \\ 0.759^{*} \\ 0.806 \end{array}$	M ↓ 0.073 0.073 † 0.069 0.059 0.050 0.070 0.056	$S_{\alpha} \uparrow$ 0.827 0.834 0.846 0.823 0.843 0.851 0.835 0.835	PopNet COI $E_{\phi} \uparrow$ 0.910* 0.905 0.897* 0.900 0.907 0.895 0.901 0.913	Ex F_{β}^{ω} \uparrow 0.757* 0.730 0.745 0.745 0.742 0.735 0.720* 0.748	M ↓ 0.031 0.032 0.032 0.029 0.032 0.032 0.032	Ff $S_{\alpha} \uparrow$ 0.852 0.848 0.856 0.846 † 0.857 †	EDER $E_{\phi} \uparrow$ 0.909* 0.906 0.902* 0.905 † 0.915 0.906 †	FPN $F_{\beta}^{\omega} \uparrow$ 0.802* 0.791 0.791 0.789* † 0.816 0.786* †	M ↓ 0.043 0.046 0.042 0.045 † 0.035 0.044 †
Input Methods PopNet [47] CFANet [52] MFFN [54] FEDER [12] Explicit [26] FSPNet [18] MRR-Net [49] FPNet [4] LSR+ ² [28]	GT Publication ICCV2023 ICME2023 WACV2023 CVPR2023 CVPR2023 CVPR2023 TNNLS2023 ACM MM203 TCSVT2023	Ours 512 ² 416 ² 384 ²	$S_{\alpha} \uparrow$ 0.806 0.815 \dagger 0.807 0.846 0.856 0.826 0.852 0.854	MRR $E_{\phi} \uparrow$ 0.859* 0.876 $\dot{1}$ 0.895 0.899 0.880 0.905 0.924	MO Fωβ ↑ 0.744* 0.761 † 0.738* 0.777 0.799 0.759* 0.806 †	M ↓ 0.073 0.073 † 0.069 0.059 0.050 0.050 0.056 0.049	$S_{\alpha} \uparrow$ 0.827 0.834 0.846 0.823 0.843 0.851 0.835 0.850 0.847	PopNet COI $E_{\phi} \uparrow$ 0.910* 0.905 0.897* 0.900 0.907 0.895 0.901 0.913 0.924	Ex F_{β}° \uparrow 0.757* 0.730 0.745 0.745 0.742 0.735 0.720* 0.748 \dagger	M↓ 0.031 0.032 0.029 0.026 0.032 0.028	Ff $S_{\alpha} \uparrow$ 0.852 0.848 0.856 0.846 \dagger 0.879 0.857 \dagger 0.857 \dagger	EDER $E_{\phi} \uparrow$ 0.909* 0.906 0.902* 0.905 † 0.915 0.906 † 0.906 †	FPN F_{β}^{ω} ↑ 0.802* 0.791 0.791 0.789* † 0.816 0.786* † †	M ↓ 0.043 0.046 0.042 0.045 † 0.035 0.044 † 0.036

5. Results

Intermediate Results:









5. **Results**

Failure Cases:



6. Conclusion & Future Works

- Presented a study on the role of **camouflage attributes** in determining the effectiveness of camouflage patterns, alongside the introduction of the COD-TAX dataset for comprehensive analysis.
- We also introduce the **ACUMEN framework**, which uniquely integrates textual and visual data for enhancing COS performance.
- For future works, we aim to refine our investigation by **assessing the camouflage level**, introducing metrics for quantifying camouflage patterns and identifying their primary influencing factors.
- In terms of broad applicability, we are eager to investigate **additional downstream applications** pertinent to COS.

Repo: https://github.com/lyu-yx/ACUMEN Paper: https://arxiv.org/abs/2408.12086