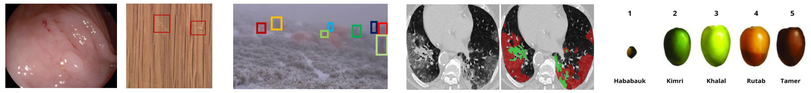
Unlocking Attributes' Contribution to Successful Camouflage: A Combined Textual and Visual Analysis Strategy

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

Introduction

Problem statement: Conventional object detection and segmentation algorithms often encounter substantial difficulties in identifying camouflaged objects, resulting in **diminished efficacy**. This challenge has spurred interest in the field of Camouflaged Object Segmentation (COS).



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.

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.

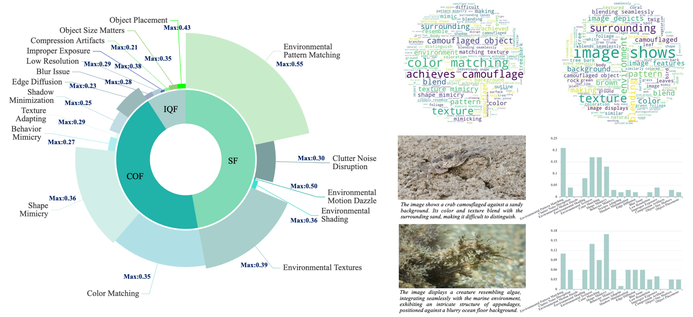
Motivation

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

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

1. Cognitive science shows that merging **textual and visual** information synergistically boosts cognitive understanding.
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.

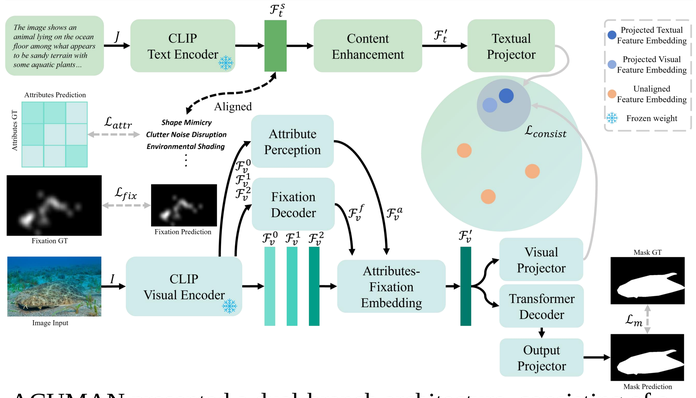
COD-TAX Dataset



- 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.

Network Overall



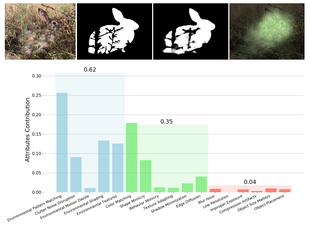
ACUMAN presented a dual-branch architecture, consisting of a textual branch (in green) and a visual branch (in cyan).

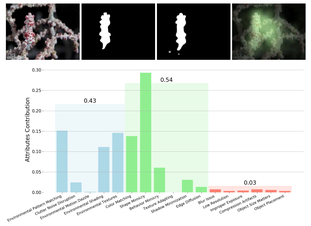
During the inference, **the textual branch is omitted** to eliminate dependency on LVLMs like GPT4, thereby making the inference process solely reliant on visual cues.

Results

Qualitative Results:

Intermediate Results:

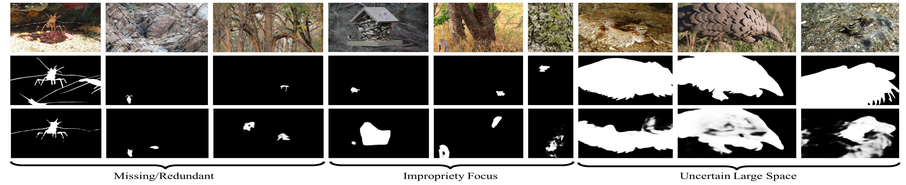




Quantitative Results:

Methods	Publication	Size	CAMO				COD10K				NC4K			
			$S_\alpha\uparrow$	$E_\phi\uparrow$	$F_\beta^w\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$E_\phi\uparrow$	$F_\beta^w\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$E_\phi\uparrow$	$F_\beta^w\uparrow$	$M\downarrow$
PopNet [47]	ICCV$_{2023}$	512^2	0.806	0.859*	0.744*	0.073	0.827	0.910*	0.757*	0.031	0.852	0.909*	0.802*	0.043
CFANet [52]	ICME$_{2023}$	416^2	0.815	0.876	0.761	0.073	0.834	0.905	0.730	0.031	0.848	0.906	0.791	0.046
MFFN [54]	WACV$_{2023}$	384^2	†	†	†	†	0.846	0.897*	0.745	0.028	0.856	0.902*	0.791	0.042
FEDER [12]	CVPR$_{2023}$	384^2	0.807	0.873	0.738*	0.069	0.823	0.900	0.716*	0.032	0.846	0.905	0.789*	0.045
Explicit [26]	CVPR$_{2023}$	352^2	0.846	0.895	0.777	0.059	0.843	0.907	0.742	0.029	†	†	†	†
FSPNet [18]	CVPR$_{2023}$	384^2	0.856	0.899	0.799	0.050	0.851	0.895	0.735	0.026	**0.879**	0.915	0.816	**0.035**
MRR-Net [49]	TNNLS$_{2023}$	384^2	0.826	0.880	0.759*	0.070	0.835	0.901	0.720*	0.032	0.857	0.906	0.786*	0.044
FPNet [4]	ACM MM$_{2023}$	512^2	0.852	0.905	0.806	0.056	0.850	0.913	0.748	0.029	†	†	†	†
LSR+^{2} [28]	TCSVT$_{2023}$	384^2	0.854	0.924	†	0.049	0.847	0.924	†	0.028	0.870	0.924	†	0.036
Ours	-	336^2	**0.886**	**0.939**	**0.850**	**0.039**	**0.852**	**0.930**	**0.761**	**0.026**	0.874	**0.932**	**0.826**	0.036

Failure Cases:



Conclusion and Future Work

- 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.