DL- Project
1.5 WILDLIFE MONITORING

Muhammad Hammad Yousaf - 26100387
Muhammad Tayyab Haider - 26100275

Object detection is a fundamental task in computer vision that aims to identify and locate objects within an image or video. Traditional object detection models perform well in structured environments where objects exhibit clear contrast against their backgrounds. However, in complex real-world scenarios, certain objects are intentionally or naturally concealed within their surroundings, making detection significantly more challenging. This leads to the specialized task of Camouflaged Object Detection (COD), which focuses on identifying objects that blend into their backgrounds due to similar texture, color, or patterns. In this specific case we are focusing on snow leopard detection in snowy areas.

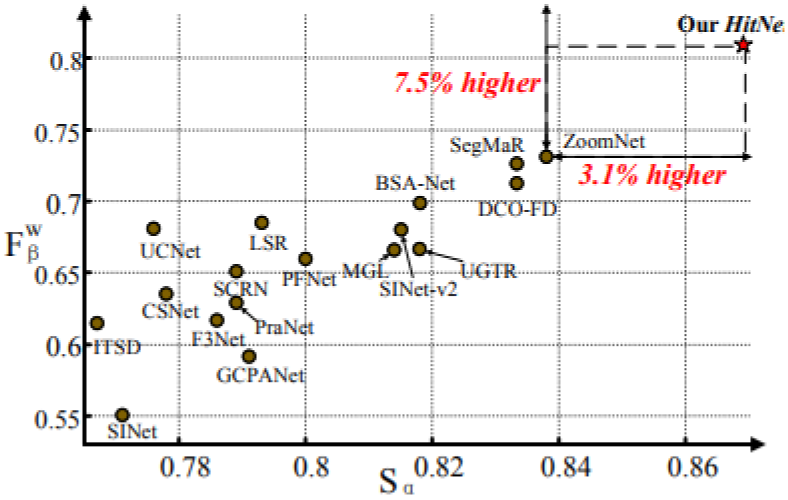
As far as standard Object detection in the wildlife is concerned, Existing AI models struggle with the following, Small/distant animals due to limited resolution in whole-image classification as well as imaging capabilities (motion sensor / heat sensor range) .as well as change in environments that the models are not trained in lead to bad classification, lack of images for all possible environments, to avoid this [4] trains an efficientv2 classifier with mega Detector v5a based on YOLO v5x6 object detcetion architecture, using datasets like WI, Lila and WildCam, proposes cropping the images after detection to be sent into the classifier, it showed how this is an effective method to improve accuracy of general wildlife detection, [1] proposes how different data capturing techniques (motion Sensor or TimelapseV2) can help with better accuracy.

Cod is an extension to the object detection problem with some further complexities of no visible boundries, Many Models try to solveThe COD problem but we will look at one of the best performing models amongst the sota models, PUENet addresses challenges in camouflaged object detection (COD), where models often fail due to biases in training data and labeling noise. Specifically, "model bias" arises from datasets where camouflaged objects are generally very centered, limiting generalization to off-center objects as they would mostly be in raw wildlife images, while "data bias" stems from ambiguous object boundaries that lead to noisy annotations. To tackle these issues, PUENet integrates a Bayesian Conditional Variational Auto-Encoder (BCVAE) (this is a bayesian cnn and a conditional Vae) to jointly estimate model uncertainty and data uncertainty, providing insights into prediction reliability. It introduces a Predictive Uncertainty Approximation (PUA) module to eliminate computationally expensive sampling during inference. Additionally, its Selective Attention Module (SAM) refines hierarchical features to better handle complex backgrounds and ambiguous edges. Evaluated on Datasets like CAMO ,COD10K, NC4K. PUENet outperforms state-of-the-art methods in accuracy and F1 measure (refer to the table below). , in our case pueNet can help massively as wildlife images are not highly centered ot processed images and this allows for a higher accuracy

The High-Resolution Iterative Feedback Network (HiNet) is a state-of-the-art camouflaged object detection (COD) that solves the longstanding issue of object segmentation that visually blends with its environment—a problem particularly critical in detecting snow leopards on snowy or rocky landscapes where their coat patterns form almost flawless camouflage. Conventional COD models, including SINet or ZoomNet, tend to experience detail loss in low-resolution (LR) feature maps, resulting in fuzzy edges and incomplete segmentation of camouflage targets. The innovation of HiNet is its iterative feedback process, which recursively refines LR features with the aid of high-resolution (HR) information in a global loop-based framework. This mechanism is based on a Pyramid Vision Transformer (PVT) backbone, selected for its capacity to extract multi-scale features efficiently with minimal GPU memory usage necessary for handling high-resolution images common to camera traps. The PVT backbone produces multi-scale features at sizes of ¼ , ⅛, 1/16, and 1/32 of the input dimension, which are used as input to a Multi-Resolution Feedback Refinement (RIR) module. Here, the iterative feedback linkages combine LR and HR attributes through a Feedback Block (FB) that applies channel-wise concatenation and up-sampling in order to retain edge information like the subtle lines of a snow leopard's pelt against a snowy background. To make the training stable, HiNet involves an iterative feedback loss that corrects each step's output under a weighted method, guaranteeing continuous refinement of segmentation masks. This structure is complemented by an Adaptive Feature Fusion (AFF) module, which dynamically fuses features across iterations with learnable coefficients, improving resistance to occlusions or partial observability prevailing in wildlife images. HiNet is evaluated on four COD benchmarks: COD10K (10,000 images), CAMO (2,500 images), CHAMELEON (76 images), and NC4K (4,121 images). On COD10K, it attains a weighted F-measure of 0.804, outperforming ZoomNet (0.729) and SINet-V2 (0.680), and decreases the Mean Absolute Error (MAE) by 16.9%, which represents better pixel-level accuracy. The structural integrity of the model is evident in its S-measure of 0.869 and E-measure of 0.936, which are essential for applications that need accurate localization, like separating a snow leopard's outline from textured backgrounds. Qualitatively, HiNet performs better in segmenting thin edges (e.g., whiskers, paw edges) and occlusions (leopards behind rocks), solving important pain points in wildlife surveillance where partial detections may result in false negatives. HiNet also performs significantly well on sparse datasets such as CAMO and CHAMELEON.

Measures such as F-measure would be applicable (highlighting recall to minimize missed detections) and MAE (for accurate localization) should dictate assessment, prioritizing the project objective of minimizing human-wildlife conflict by providing reliable early warnings.

Weighted F-measure vs. Structure-measure
of top 17 models from 35 SOTA methods



Evaluation Chart for SOTA COD Models

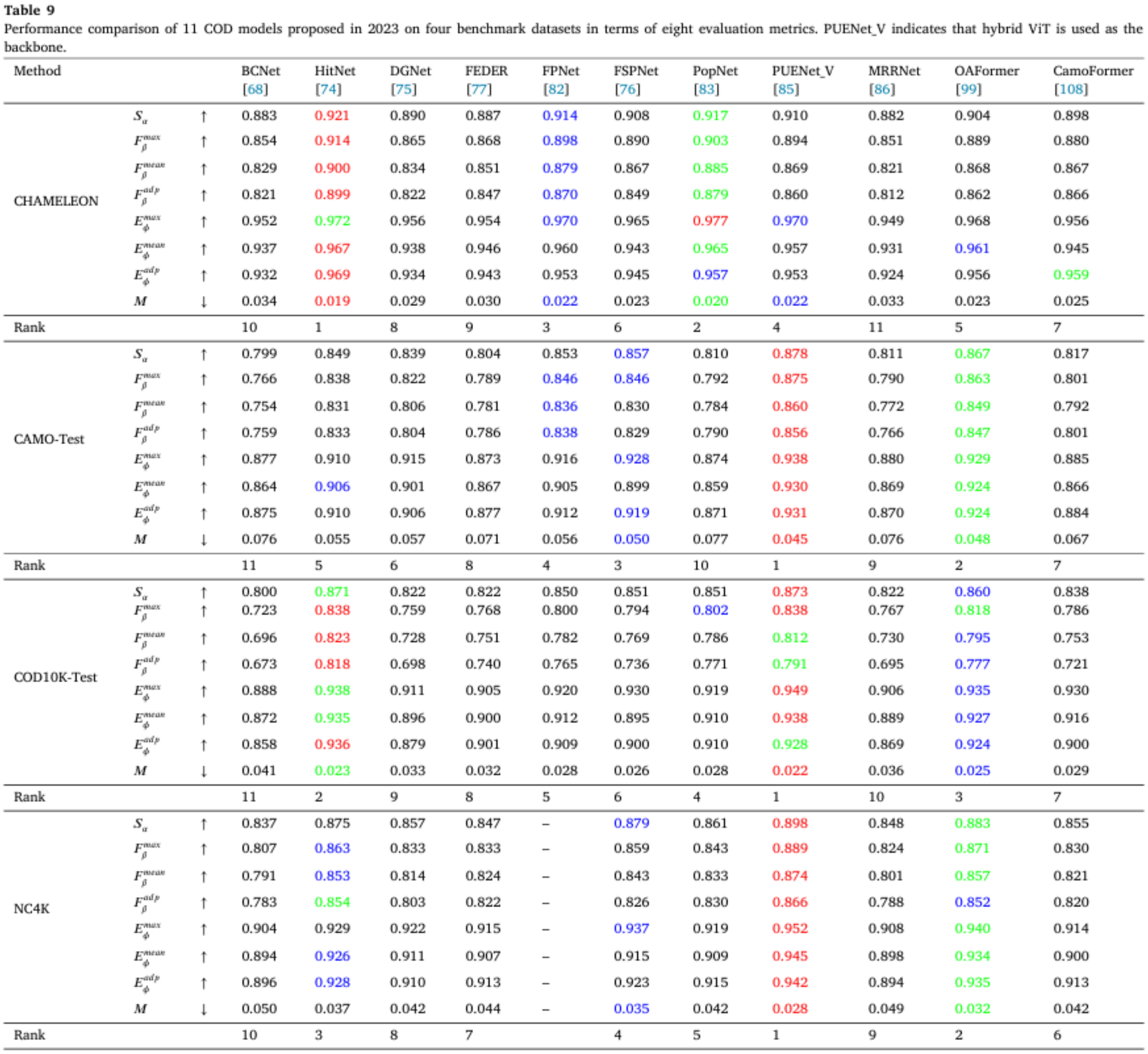


Table 9

Performance comparison of 11 COD models proposed in 2023 on four benchmark datasets in terms of eight evaluation metrics. PUENet_V indicates that hybrid ViT is used as the backbone.

Method			BCNet [68]	HitNet [74]	DGNet [75]	FEDER [77]	FPNet [82]	FSPNet [76]	PopNet [83]	PUENet_V [85]	MRRNet [86]	OAFormer [99]	CamoFormer [108]
CHAMELEON	S_α	↑	0.883	0.921	0.890	0.887	0.914	0.908	0.917	0.910	0.882	0.904	0.898
	F_β^{max}	↑	0.854	0.914	0.865	0.868	0.898	0.890	0.903	0.894	0.851	0.889	0.880
	F_β^{mean}	↑	0.829	0.900	0.834	0.851	0.879	0.867	0.885	0.869	0.821	0.868	0.867
	F_β^{adp}	↑	0.821	0.899	0.822	0.847	0.870	0.849	0.879	0.860	0.812	0.862	0.866
	E_ϕ^{max}	↑	0.952	0.972	0.956	0.954	0.970	0.965	0.977	0.970	0.949	0.968	0.956
	E_ϕ^{mean}	↑	0.937	0.967	0.938	0.946	0.960	0.943	0.965	0.957	0.931	0.961	0.945
	E_ϕ^{adp}	↑	0.932	0.969	0.934	0.943	0.953	0.945	0.957	0.953	0.924	0.956	0.959
	M	↓	0.034	0.019	0.029	0.030	0.022	0.023	0.020	0.022	0.033	0.023	0.025
Rank			10	1	8	9	3	6	2	4	11	5	7
CAMO-Test	S_α	↑	0.799	0.849	0.839	0.804	0.853	0.857	0.810	0.878	0.811	0.867	0.817
	F_β^{max}	↑	0.766	0.838	0.822	0.789	0.846	0.846	0.792	0.875	0.790	0.863	0.801
	F_β^{mean}	↑	0.754	0.831	0.806	0.781	0.836	0.830	0.784	0.860	0.772	0.849	0.792
	F_β^{adp}	↑	0.759	0.833	0.804	0.786	0.838	0.829	0.790	0.856	0.766	0.847	0.801
	E_ϕ^{max}	↑	0.877	0.910	0.915	0.873	0.916	0.928	0.874	0.938	0.880	0.929	0.885
	E_ϕ^{mean}	↑	0.864	0.906	0.901	0.867	0.905	0.899	0.859	0.930	0.869	0.924	0.866
	E_ϕ^{adp}	↑	0.875	0.910	0.906	0.877	0.912	0.919	0.871	0.931	0.870	0.924	0.884
	M	↓	0.076	0.055	0.057	0.071	0.056	0.050	0.077	0.045	0.076	0.048	0.067
Rank			11	5	6	8	4	3	10	1	9	2	7
COD10K-Test	S_α	↑	0.800	0.871	0.822	0.822	0.850	0.851	0.851	0.873	0.822	0.860	0.838
	F_β^{max}	↑	0.723	0.838	0.759	0.768	0.800	0.794	0.802	0.838	0.767	0.818	0.786
	F_β^{mean}	↑	0.696	0.823	0.728	0.751	0.782	0.769	0.786	0.812	0.730	0.795	0.753
	F_β^{adp}	↑	0.673	0.818	0.698	0.740	0.765	0.736	0.771	0.791	0.695	0.777	0.721
	E_ϕ^{max}	↑	0.888	0.938	0.911	0.905	0.920	0.930	0.919	0.949	0.906	0.935	0.930
	E_ϕ^{mean}	↑	0.872	0.935	0.896	0.900	0.912	0.895	0.910	0.938	0.889	0.927	0.916
	E_ϕ^{adp}	↑	0.858	0.936	0.879	0.901	0.909	0.900	0.910	0.928	0.869	0.924	0.900
	M	↓	0.041	0.023	0.033	0.032	0.028	0.026	0.028	0.022	0.036	0.025	0.029
Rank			11	2	9	8	5	6	4	1	10	3	7
NC4K	S_α	↑	0.837	0.875	0.857	0.847	–	0.879	0.861	0.898	0.848	0.883	0.855
	F_β^{max}	↑	0.807	0.863	0.833	0.833	–	0.859	0.843	0.889	0.824	0.871	0.830
	F_β^{mean}	↑	0.791	0.853	0.814	0.824	–	0.843	0.833	0.874	0.801	0.857	0.821
	F_β^{adp}	↑	0.783	0.854	0.803	0.822	–	0.826	0.830	0.866	0.788	0.852	0.820
	E_ϕ^{max}	↑	0.904	0.929	0.922	0.915	–	0.937	0.919	0.952	0.908	0.940	0.914
	E_ϕ^{mean}	↑	0.894	0.926	0.911	0.907	–	0.915	0.909	0.945	0.898	0.934	0.900
	E_ϕ^{adp}	↑	0.896	0.928	0.910	0.913	–	0.923	0.915	0.942	0.894	0.935	0.913
	M	↓	0.050	0.037	0.042	0.044	–	0.035	0.042	0.028	0.049	0.032	0.042
Rank			10	3	8	7		4	5	1	9	2	6

[2]

PueNet Performance comparison with SOTA Cod Models

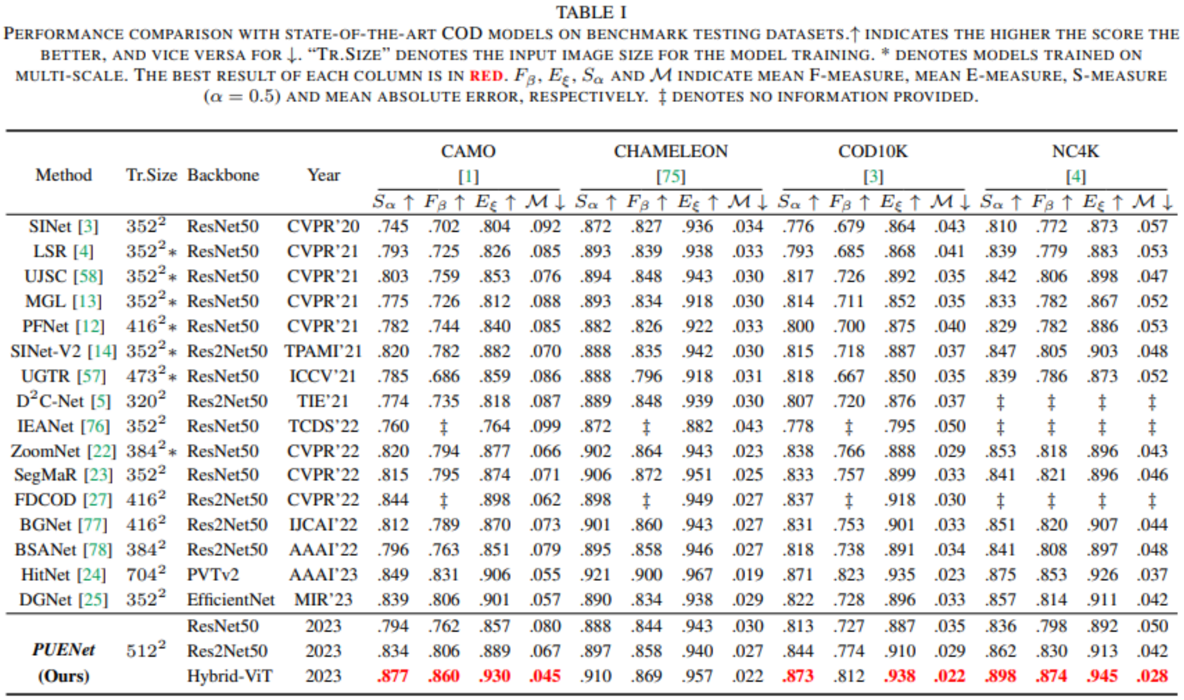


TABLE I
PERFORMANCE COMPARISON WITH STATE-OF-THE-ART COD MODELS ON BENCHMARK TESTING DATASETS.↑ INDICATES THE HIGHER THE SCORE THE BETTER, AND VICE VERSA FOR ↓. "TR.SIZE" DENOTES THE INPUT IMAGE SIZE FOR THE MODEL TRAINING. * DENOTES MODELS TRAINED ON MULTI-SCALE. THE BEST RESULT OF EACH COLUMN IS IN RED. F_β, E_ξ, S_α AND $\mathcal{M}$ INDICATE MEAN F-MEASURE, MEAN E-MEASURE, S-MEASURE ($\alpha = 0.5$) AND MEAN ABSOLUTE ERROR, RESPECTIVELY. ‡ DENOTES NO INFORMATION PROVIDED.

Method	Tr.Size	Backbone	Year	CAMO [1]				CHAMELEON [75]				COD10K [3]				NC4K [4]			
				$S_\alpha \uparrow$	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$	$S_\alpha \uparrow$	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$	$S_\alpha \uparrow$	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$	$S_\alpha \uparrow$	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$
SINet [3]	352^2	ResNet50	CVPR'20	.745	.702	.804	.092	.872	.827	.936	.034	.776	.679	.864	.043	.810	.772	.873	.057
LSR [4]	352^2*	ResNet50	CVPR'21	.793	.725	.826	.085	.893	.839	.938	.033	.793	.685	.868	.041	.839	.779	.883	.053
UJSC [58]	352^2*	ResNet50	CVPR'21	.803	.759	.853	.076	.894	.848	.943	.030	.817	.726	.892	.035	.842	.806	.898	.047
MGL [13]	352^2*	ResNet50	CVPR'21	.775	.726	.812	.088	.893	.834	.918	.030	.814	.711	.852	.035	.833	.782	.867	.052
PFNet [12]	416^2*	ResNet50	CVPR'21	.782	.744	.840	.085	.882	.826	.922	.033	.800	.700	.875	.040	.829	.782	.886	.053
SINet-V2 [14]	352^2*	Res2Net50	TPAMI'21	.820	.782	.882	.070	.888	.835	.942	.030	.815	.718	.887	.037	.847	.805	.903	.048
UGTR [57]	473^2*	ResNet50	ICCV'21	.785	.686	.859	.086	.888	.796	.918	.031	.818	.667	.850	.035	.839	.786	.873	.052
D^{2}C-Net [5]	320^2	Res2Net50	TIE'21	.774	.735	.818	.087	.889	.848	.939	.030	.807	.720	.876	.037	‡	‡	‡	‡
IEANet [76]	352^2	ResNet50	TCDS'22	.760	‡	.764	.099	.872	‡	.882	.043	.778	‡	.795	.050	‡	‡	‡	‡
ZoomNet [22]	384^2*	ResNet50	CVPR'22	.820	.794	.877	.066	.902	.864	.943	.023	.838	.766	.888	.029	.853	.818	.896	.043
SegMaR [23]	352^2	ResNet50	CVPR'22	.815	.795	.874	.071	.906	.872	.951	.025	.833	.757	.899	.033	.841	.821	.896	.046
FDCOD [27]	416^2	Res2Net50	CVPR'22	.844	‡	.898	.062	.898	‡	.949	.027	.837	‡	.918	.030	‡	‡	‡	‡
BGNet [77]	416^2	Res2Net50	IJCAI'22	.812	.789	.870	.073	.901	.860	.943	.027	.831	.753	.901	.033	.851	.820	.907	.044
BSANet [78]	384^2	Res2Net50	AAAI'22	.796	.763	.851	.079	.895	.858	.946	.027	.818	.738	.891	.034	.841	.808	.897	.048
HitNet [24]	704^2	PVTv2	AAAI'23	.849	.831	.906	.055	.921	.900	.967	.019	.871	.823	.935	.023	.875	.853	.926	.037
DGNet [25]	352^2	EfficientNet	MIR'23	.839	.806	.901	.057	.890	.834	.938	.029	.822	.728	.896	.033	.857	.814	.911	.042
***PUENet* (Ours)**	512^2	ResNet50	2023	.794	.762	.857	.080	.888	.844	.943	.030	.813	.727	.887	.035	.836	.798	.892	.050
		Res2Net50	2023	.834	.806	.889	.067	.897	.858	.940	.027	.844	.774	.910	.029	.862	.830	.913	.042
		Hybrid-ViT	2023	**.877**	**.860**	**.930**	**.045**	.910	.869	.957	.022	**.873**	.812	**.938**	**.022**	**.898**	**.874**	**.945**	**.028**

[3]

A big Issue however is that most of the wildlife training data cannot be used as Training Data for the COD problem, The Datasets primarily used in the Sota Cod models are trained on COD10K (10,000 images), CAMO (2,500 images), CHAMELEON (76 images), and NC4K (4,121 images), however these datasets are not primarily images of camouflaged animals but also include camouflaged vehicles or patterns etc. On top of that, none of these datasets include snow leopards. They focus on general camouflaged animals which doesn't bind specifically to our problem, if we minimize the dataset to just snow leopards, we will end up with a highly sparse dataset,HiNet's cross-domain learning (Cross Domain Learning) approach provides a practical solution to data poverty. Utilizing CycleGAN-based transformations, we can potentially generate data with snow leopards to improve training of the models in our context, while HiNet is pre-trained on generic COD datasets (e.g., CAMO's 8 animal classes), fine-tuning with snow leopard-specific data can help cover domain shifts, e.g., specific snow texture or lighting in northern Pakistan. The feasibility of this idea needs further testing and research. The mega detector created bounding boxes to crop images, if we can access data that detects movement through heat or movement sensors we can use a cropped version of the image to firstly help with the distance / size issue discussed in [4] and secondly, help with reducing the zero bias in the images as discussed above. Further improvements can be made by substituting individual component models in the above discussed models to improve efficiency/ accuracy, but this needs further research and testing.

(STUDY THE NUMBERED ARTICLES BELOW AS KEY TO TEXT REFERENCES)

[1]
Leorna, Scott, and Todd Brinkman. "Human vs. machine: Detecting wildlife in camera trap images." *Ecological Informatics* 72 (2022): 101876.
[2]
Liang, Yanhua, et al. "A systematic review of image-level camouflaged object detection with deep learning." *Neurocomputing* 566 (2024): 127050.
[3]
Zhang, Yi, et al. "Predictive uncertainty estimation for camouflaged object detection." *IEEE Transactions on Image Processing* 32 (2023): 3580-3591.
[4]
Gadot, Tomer, et al. "To crop or not to crop: Comparing whole-image and cropped classification on a large dataset of camera trap images." *IET Computer Vision* 18.8 (2024): 1193-1208.
[5]
Hu, Xiaobin, et al. "High-resolution iterative feedback network for camouflaged object detection." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 37. No. 1. 2023.

Datasets:

WI Dataset - https://www.wildlifeinsights.org/
LILA - https://lila.science/
Camo - https://sites.google.com/view/ltnghia/research/camo
CHAMELEON - https://www.polsl.pl/rau6/chameleon-database-animal-camouflage-analysis/-
COD10k - http://dpfan.net/camouflage/
Snapshot Serengeti dataset - https://www.zooniverse.org/projects/zooniverse/snapshot-serengeti

GitHubLinks

https://github.com/visionxiang/awesome-camouflaged-object-detection?tab=readme-ov-file
https://github.com/google/cameratrapai/tree/main/speciesnet