BANGALORE INSTITUTE OF TECHNOLOGY
K. R. Road, V. V. Pura, Bengaluru – 560004

Department of Computer Science and Engineering

Camouflaged Object Sensing Using Deep Learning Approach

By:
Kruthi N Raj (1BI18CS068)
Under the Guidance
of
Prof. Manasa T P
Assistant Professor
Dept. of CSE

INTRODUCTION

- Object detection is the fundamental component of optical systems
- However, it is very challenging to apply object detection techniques in harsh or extreme situations that are even challenging to the naked eye
- A typical example is to identify species with camouflage capabilities from images acquired by non-invasive sensors i.e. camera traps
- Limited imaging quality of the sensors and the illuminance conditions, objects often show similarities in color and texture with the background
- Challenging tasks like these are solved using camouflaged object detection (COD)
- Camouflage is the phenomenon of visual concealment that exists extensively in both natural and artificial objects
- COD focuses on targets that are less likely to capture human attention or attempt to deceive visual perception systems in an adversarial manner

Image of camouflaged leopard

LITERATURE REVIEW

SL.No	Title	Author	Year of Publication	Adopted Methodology	Limitations
1	**Anabranch Network for Camouflaged Object Segmentation**	T.N. Le, T.V. Nguyen, Z. Nie, M.T. Tran, Sugimoto	2019	This paper proposes an ANet model that uses an additional classification networks to refine the prediction results of traditional target segmentation networks. The proposed network leverages the strength of both image classification and semantic segmentation tasks for camouflaged object segmentation	Its a two-stream structure that is based on the traditional convolutional network structure and, thus, cannot provide the perceptual ability required by the COD task

LITERATURE REVIEW

SL.No	Title	Author	Year of Publication	Adopted Methodology	Limitations
2	**Simultaneously Localize, Segment and Rank the Camouflaged Objects**	Yunqiu Lv, Jing Zhang, Yuchao Dai, Aixuan Li, Bowen Liu, Nick Barnes, Deng-Ping Fan	2021	This paper proposes the RankNet architecture that generates saliency prediction by instance-level ranking-based region. RankNet uses the localization model to find the discriminative regions and the segmentation model to segment the full scope of the camouflaged objects	Saliency prediction deals with model saliency and salient objects tend to be visually distinct from the surroundings. Thus it can't find the vague boundaries of objects and are not competent to accurately detect camouflaged objects

LITERATURE REVIEW

SL.No	Title	Author	Year of Publication	Adopted Methodology	Limitations
3	**Camouflaged object detection**	Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, Ling Shao	2020	In this paper the authors propose a SINet architecture that utilizes a cascaded network, which divides the network into a Search Module (SM) and an Identification Module (IM), to hierarchically refine the prediction map. SINet also uses cascaded partial decoder as decoders	PDC mixes features by addition and concatenation which is not robust enough to deal with low signal-to-noise ratio features obtained due to the lack of semantic orientation from shallow encoder layers

LITERATURE REVIEW

SL.No	Title	Author	Year of Publication	Adopted Methodology	Limitations
4	**A Bayesian Approach to Camouflaged Moving Object Detection**	Xiang Zhang, Ce Zhu, Shuai Wang, Yipeng Liu	2016	The paper focuses on the detection of moving camouflaged objects. It proposes camouflage modeling (CM) to identify camouflaged foreground pixels. Since, moving object is usually composed of both camouflaged and non-camouflaged areas, CM and discriminative modeling (DM) are fused in a bayesian framework to perform complete object detection	Utilizes low-level features like texture, edge, brightness, and color to discriminate objects. These features are incapable of detecting all the sophisticated camouflage strategies in the real application scenarios

LITERATURE REVIEW

SL.No	Title	Author	Year of Publication	Adopted Methodology	Limitations
5	**Foreground Detection in Camouflaged Scenes**	Shuai Li, Dinei Florencio, Yaqin Zhao, Chris Cook, Wanqing Li	2017	This paper proposes a texture guided weighted voting (TGWV) to detect foreground objects in camouflaged scenes. The proposed method uses the stationary wavelet transform to decompose the image into frequency bands shows that the small and hardly noticeable differences between foreground and background in images can be captured in certain wavelet frequency bands. To make the final foreground decision, a weighted voting scheme is developed based on intensity and texture of all the wavelet bands with weights carefully designed.	The computational complexity of the proposed method is higher due to the multiple level wavelet decomposition

LITERATURE REVIEW

SL.No	Title	Author	Year of Publication	Adopted Methodology	Limitations
6	**Mutual Graph Learning for Camouflaged Object Detection**	Qiang Zhai, Xin Li, Fan Yang, Chenglizhao Chen, Hong Cheng, Deng-Ping Fan	2021	The paper proposes a Mutual Graph Learning model which decouples an image into two task-specific feature maps - one for roughly locating the camouflaged target and the other for accurately capturing its boundary details and exploits the mutual benefits by recurrently reasoning their high-order relations through graphs	The paper considers edge disruption as one of the key factors for camouflage which may not be the case in all situations

LITERATURE REVIEW

SL.No	Title	Author	Year of Publication	Adopted Methodology	Limitations
7	**Mirror-net: Bio-inspired adversarial attack for camouflaged object segmentation**	Jinnan Yan, Trung-Nghia Le, Khanh-Duy Nguyen, Minh-Triet Tran, Thanh-Toan Do, Tam V. Nguyen	2021	The paper proposes a MirrorNet architecture that uses instance segmentation and mirror stream for the camouflaged object segmentation. The proposed network possesses two segmentation streams: the main stream and the mirror stream corresponding with the original image and its flipped image, respectively. The output from the mirror stream is fused into the main stream's result for the final camouflage map to boost up the segmentation accuracy	Its a bio-inspired model that mimics the perception and cognition of observers. However, it ignores an important attribute, the time that observers spend on searching for the camouflaged object varies in wide range and heavily depends on the effectiveness of camouflage. Therefore, the model fails to consider that the features employed to detect the objects are also different when they have different camouflage degrees

LITERATURE REVIEW

SL.No	Title	Author	Year of Publication	Adopted Methodology	Limitations
8	**Camouflaged Object Segmentation with Distraction Mining**	Haiyang Mei, Ge-Peng Ji, Ziqi Wei, Xin Yang	2021	This paper proposes Positioning and Focus Network (PFNet), which mimics the process of predation in nature. PFNet contains two key modules, the positioning module (PM) and the focus module (FM). The PM is designed to mimic the detection process in predation for positioning the potential target objects from a global perspective and the FM is then used to perform the identification process in predation for progressively refining the coarse prediction via focusing on the ambiguous regions. In the FM, distraction mining strategy is used for the distraction discovery and removal, to benefit the performance of estimation	In case of higher-level prediction, the performance of the network would decline to some extent as indiscriminately mining distractions from the input features increases the difficulty of the distraction discovery and thus hinders the effective distraction removal

LITERATURE REVIEW

SL.No	Title	Author	Year of Publication	Adopted Methodology	Limitations
9	**Camouflage Performance Analysis and Evaluation Framework Based on Features Fusion**	Feng Xue, Chengxi Yong, Shan Xu, Hao Dong, Yuetong Luo, Wei Jia	2016	This paper proposes a framework that uses nonlinear fusion of multiple image features to quantitatively evaluate the degree to which the target and surrounding background differ with respect to background-related and internal features. Background-related features are first formulated as a measure of saliency, which is calculated and quantized by SOD, whereas internal features refer to the interior saliency of camouflage textures, such as lines and other regular patterns. These two features are fused to identify and evaluate the camouflage effect	SOD requires objects to be visually distinct from the background and does not specialize in finding the vague boundaries of objects. Considering internal features of camouflaged objects may not be robust in case of sophisticated camouflage strategies in the real application scenarios

PROBLEM STATEMENT

Camouflaged Object Sensing using Deep Learning

Salient object detection (SOD) and Camouflaged object detection (COD) were two traditional methodologies for camouflage object sensing . COD is foreground detection, which uses hand-crafted characteristics computed by edges, brightness, corner points, texture, or temporal information to differentiate the camouflaged target from the background. SOD finds objects that attract human attention. SOD's fundamental flaw is that it performs model saliency, which means that salient objects are visually distinguishable from their surroundings. As a result, SOD is unable to recognize unclear object boundaries and so is unable to reliably detect camouflaged objects, and COD's hand-crafted features are incapable of recognizing all advanced camouflage tactics in real-world circumstances.

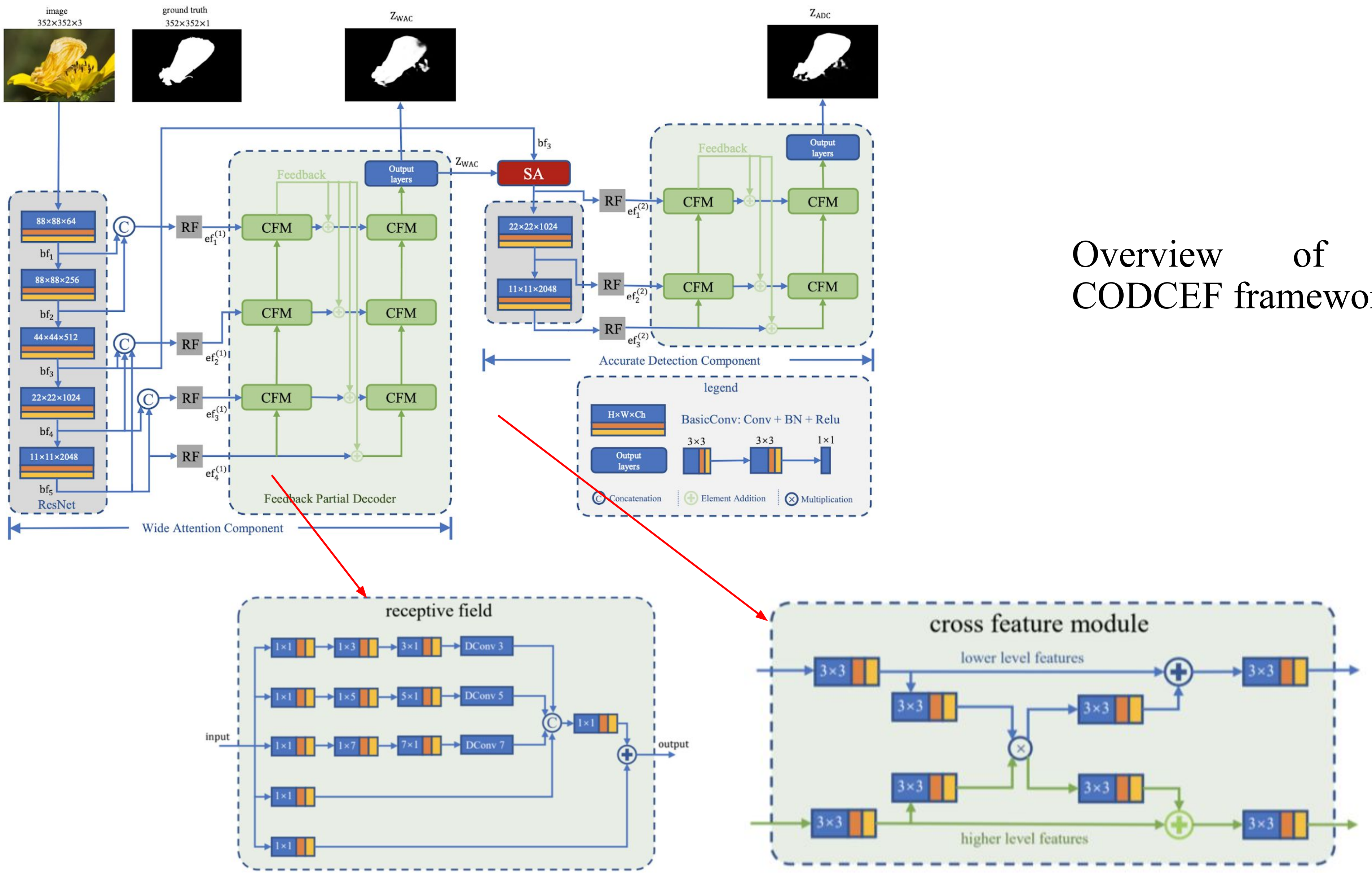
Objectives:

1. To identify camouflaged objects in images
2. To reduces the loss of information and the interference of noises in the process of feature interweaving during COD, a major drawback in the traditional approaches
3. To pay more attention to local pixels that might become the edges of an object which helps obtain better results when compared to traditional approaches

PROPOSED SYSTEM

- The proposed framework is the Camouflaged Object Detection with Cascade and FEedback Fusion (CODCEF) and the corresponding optimization strategy
- CODCEF is composed of two cascaded network components, the Wide Attention Component (WAC) to obtain an approximation of the detected outline and the Accurate Detection Component (ADC) to refine the edge of previous prediction and eliminate residual noise
- In each of the 2 modules, Feedback Partial Decoder (FPD) based on Cross feature module (CFM) which is used to deal with the structural details and semantic information in the multi-level features in the images
- The proposed framework divided into multiple parts with clear responsibilities allows to obtain intermediate results which help in developing the loss function, Pixel PerceptionFusion Loss (PPF)
- PPF gives extra weight to sharply changing pixels to focus the attention of framework on possible object boundaries

ARCHITECTURE



Overview of the CODCEF framework

ALGORITHM / METHODOLOGY

Data Acquisition

- COD10K, CAMO and CHAMELEON are the source of the basic dataset

Model

- Camouflaged Object Detection with Cascade and Feedback Fusion (CODCEF) is a 2 module deep learning framework, the modules being WAC and ADC
- It is based on an RGB optical sensor that leverages a cascaded structure with Feedback Partial Decoders (FPD) instead of a traditional encoder–decoder structure
- Through a selective fusion strategy and feedback loop, FPD reduces the loss of information and the interference of noises in the process of feature interweaving
- Pixel Perception Fusion (PPF) loss is introduced, which aims to pay more attention to local pixels that might become the edges of an object

Training Data

- The training set, COD10K and CAMO, are combined obtaining a training set containing close to 7000 images. This training set covers a variety of targets from salient targets to difficult camouflaged targets

Training Implementation

- The network is trained using Adam optimizer with batch-size = 32 and learning rate = 0.0001 and obtained the best results in 55 training epochs

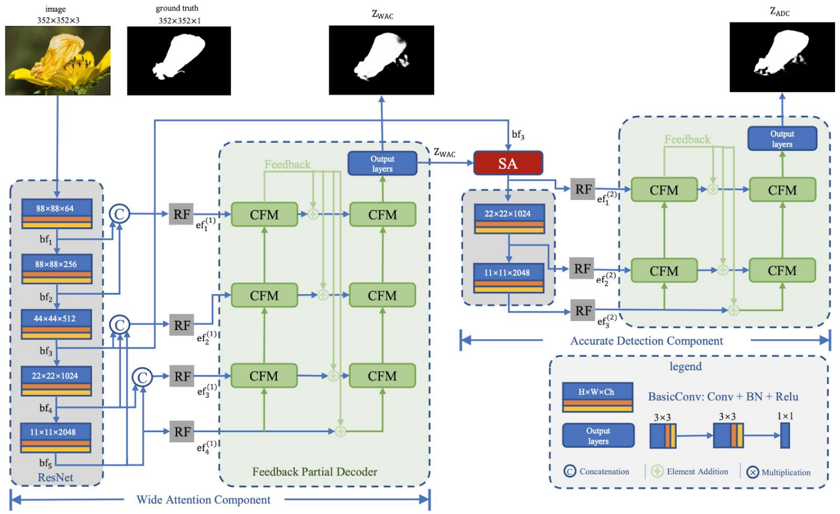
Testing Implementation

- To unify the different images, all input images were resized to a resolution of 352 × 352 and normalized. For the evaluation of the results, the prediction maps were up-sampled to the original resolution

ALGORITHM / METHODOLOGY

Implementation

- WAC as a relatively independent module, takes, as input, the original RGB image and outputs a prediction that can be used to calculate loss of network.
- The ADC combines the output of results with the middle-level features of the original image to screen out possible misleading information and noise.
- ResNet-50 pre-trained model is used as the encoder of CODCEF to extract basic features at different levels
- Up-sampling and down-sampling operations are used to normalize the resolution of the basic features obtained to the maximum resolution, obtaining four hybrid features.
- Receptive Fields Block module (RFB) are used to reduce some loss in the feature discriminability as well as robustness and transform hybrid features into enhanced features
- Feedback Partial Decoder (FPD) with three feedback loops interweave and merge features into a phased result, denoted
- as ZWAC in architecture figure
- FPD comprises of CFMs which effectively suppress the background noise of the feature and sharpen the boundary of the prediction map.
- Search Attention function (SA) function summarizes the prediction results of WAC. It multiplies a preliminary prediction by the middle-level feature, which contains most of the features of the original image with low noise, generating the attention map A
- Gaussian filter to actively blur the boundary
- The attention map goes through a shallow convolutional network to extract certain features and the features are enhanced by using RF
- To holistically obtain the final prediction map, FPD with two layers of feedback loops is used



RESULTS

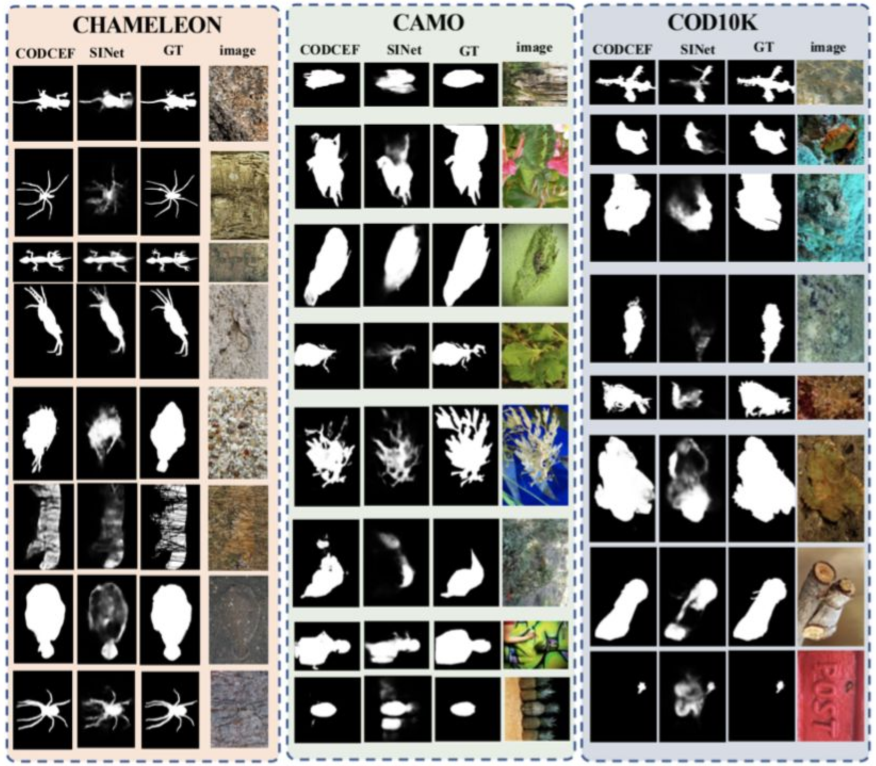
Evaluation Metrics

- Mean Absolute Error (MAE) - to calculate the difference between prediction maps and the ground truth
- S-measure - to evaluates region-aware and object-aware structural similarity
- F-measure - to evaluate structural similarity
- E-measure - to evaluate pixel-level matching

Evaluation

- CODCEF is compared with10 previous methods, including FPN, BASNet, PFANet, CPD, ANet, CSNet, SINet, RankNet, and R-MGL

Models	CHAMELEON [49]				COD10K [20]				CAMO [18]			
	$M \downarrow$	$S_\alpha \uparrow$	$F_\beta \uparrow$	$E_m \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$F_\beta \uparrow$	$E_m \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$F_\beta \uparrow$	$E_m \uparrow$
FPN^{2017} [34]	0.075	0.794	0.590	0.783	0.075	0.697	0.411	0.691	0.131	0.684	0.483	0.677
BASNet^{2019} [26]	0.118	0.687	0.474	0.721	0.105	0.634	0.365	0.678	0.159	0.618	0.413	0.661
PFANet^{2019} [22]	0.144	0.679	0.378	0.648	0.128	0.636	0.286	0.618	0.172	0.659	0.391	0.622
CPD^{2019} [25]	0.052	0.853	0.706	0.866	0.059	0.747	0.508	0.770	0.115	0.726	0.550	0.729
CSNet^{2019} [51]	0.051	0.819	0.759	0.859	0.048	0.745	0.615	0.808	0.106	0.704	0.633	0.753
F3Net^{2020} [27]	0.047	0.848	0.770	0.894	0.051	0.739	0.593	0.795	0.109	0.711	0.616	0.741
ANet^{2019} [18]	-	-	-	-	-	-	-	-	0.126	0.682	0.484	0.685
SINet^{2020} [20]	0.044	0.869	0.740	0.891	0.051	0.771	0.551	0.806	0.100	0.751	0.606	0.771
RankNet^{2021} [19]	0.046	0.842	0.794	0.896	0.045	0.760	0.658	0.831	0.105	0.708	0.645	0.755
R-MGL^{2021} [50]	0.030	0.893	0.813	0.923	0.035	0.814	0.666	0.865	0.088	0.775	0.673	0.847
CODCEF(Ours)	0.030	0.875	0.825	0.932	0.043	0.766	0.667	0.854	0.092	0.736	0.685	0.797



RESULTS

Analysis

- CODCEF demonstrated strong competitiveness in the prediction accuracy and model size
- SOD domain model still lagged behind the COD model by a large margin indicating that the challenge of the COD task is, indeed, different than that of the traditional SOD task
- To locate the object when the camouflage degree of the target is close to the limit of what the naked eye can detect, a COD model is required
- COD models, SINet and RankNet, CODCEF showed more powerful camouflaged target positioning capabilities and more accurate object boundary perception capabilities, outperforming in most of metrics
- In terms of the prediction accuracy,CODCEF method is indeed slightly worse than R-MGL
- on $S\alpha$ and Em.
- However, we must note that the typical structure of R-MGL contains 444M parameters, while CODCEF only needs half (213M), which makes it more suitable for running in edge devices with small memory
- CODCEF focuses on enhancing the robustness of features with a low signal-to-noise ratio without significantly increasing the complexity and size of the network
- On the COD10K dataset, CODCEF produces accurate predictions with sharp object boundaries in cases where SINet fails
- On the CAMO dataset, CODCEF performs well even on extremely low feature signal-to-noise ratio images
- On CHAMELEON dataset, CODCEF outperformed all 10 SOD and COD models in four metrics

Model	Params	Infer Time
SINet	198M	32 ms
R-MGL	444M	48 ms
CODCEF	212M	37 ms

APPLICATIONS

- Camouflaged object detection (COD) is a critical object detection technique in biological, security, and military scenarios
- COD is used to identify species with camouflage capabilities from images acquired by non-invasive sensors
- COD is beneficial for applications in the fields of computer vision (for search- and-rescue work, or rare species discovery)
- COD is used for medical image segmentation (e.g., polyp segmentation, lung infection segmentation)
- COD is used in the agricultural domain for applications like locust detection
- COD is used in the art field for photo-realistic blending, recreational art, etc.

CONCLUSION

Summary of the proposed system

- CODCEF consists of two relatively independent and cascaded perceptual modules
- Compared with traditional single encoder–decoder structures, the architecture showed stronger detection accuracy and robustness
- To undertake the feature decoding task of CODCEF, a Cross Feature Module to build a Feedback Partial Decoder (FPD), which effectively reduces misleading information brought about by camouflage images was used
- In addition, a novel loss function Pixel Perception Fusion Loss (PPF) to mitigate possible target edges was proposed
- The experiments showed that CODCEF achieved state-of-the-art performance on three benchmark datasets of camouflaged object detection on four evaluation metrics

Limitations and scope of the system

- The main limitation of CODCEF lies in the lack of generalization ability caused by the training samples
- For target types that do not appear in large-scale training samples, such as human military camouflage, CODCEF's prediction accuracy is limited

Future enhancements

- To introduce semi-supervised learning to deal with the lack of target training data in certain fields

REFERENCES

[1] Zhang, X.; Zhu, C.; Wang, S.; Liu, Y.; Ye, M. A Bayesian approach to camouflaged moving object detection. IEEE Trans. Circuits Syst. Video Technol. 2016, 27, 2001–2013.

[2] Li, S.; Florencio, D.; Zhao, Y.; Cook, C.; Li, W. Foreground detection in camouflaged scenes. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 4247–4251. doi:10.1109/ICIP.2017.8297083.

[3] Pike, T.W. Quantifying camouflage and conspicuousness using visual salience. Methods Ecol. Evol. 2018, 9, 1883–1895.

[4]Le, T.N.; Nguyen, T.V.; Nie, Z.; Tran, M.T.; Sugimoto, A. Anabranch network for camouflaged object segmentation. Comput. Vis. Image Underst. 2019, 184, 45–56.

[5] Lv, Y.; Zhang, J.; Dai, Y.; Li, A.; Liu, B.; Barnes, N.; Fan, D.P. Simultaneously localize, segment and rank the camouflaged objects. arXiv 2021, arXiv:2103.04011.

[6] Fan, D.P.; Ji, G.P.; Sun, G.; Cheng, M.M.; Shen, J.; Shao, L. Camouflaged object detection. In Proceedings of the IEEE/CVF. Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 2777–2787.

[7] F. Xue, C. Yong, S. Xu, H. Dong, Y. Luo, and W. Jia, "Camouflage per- formance analysis and evaluation framework based on features fusion," Multimedia Tools and Applications, vol. 75, no. 7, pp. 4065–4082, Apr 2016.

[8] Haiyang Mei, Ge-Peng Ji, Ziqi Wei, Xin Yang, Xiaopeng Wei, and Deng-Ping Fan. Camouflaged object segmentation with distraction mining. In CVPR, 2021.

[9] Jinnan Yan, Trung-Nghia Le, Khanh-Duy Nguyen, Minh- Triet Tran, Thanh-Toan Do, and Tam V Nguyen. Mirror- net: Bio-inspired adversarial attack for camouflaged object segmentation. arXiv, 2020.