Camouflaged Object Based on Attention Mechanism

Bing Zhang, Baoshan Sun

***Abstract*— The camouflage object detection(COD) aims to identify the goals that fully integrate into the surrounding environment, and have a wide range of application scenarios. The key challenge of COD is high similarity between the goals to be recognized and the surrounding background. In this article, we have proposed a framework based on attention mechanism, searching and focusing network (FDNET), which simulates the predation process of nature. Our FDNET model uses attention mechanisms, and on the basis of enhancing the characteristics of the characteristic semantics, the interference in the background area and foreground areas is gradually refined. The experimental results show that our sfnet is better than other methods at this stage for camouflage target detection.**

***Index Terms*—COD; FDNET; attention mechanisms**

I. Introduction

The camouflage goal is to integrate the purpose of the color, texture, texture, and other methods to achieve the purpose of hiding themselves[1]. Biologists call this phenomenon background matching camouflage (BMC). The task of camouflage target detection is to distinguish these camouflage objects hidden in the surrounding environment and mark them. The camouflage object test is widely used in different areas such as medical diagnosis (polyps[2], lung infection[3]), industry (inspection of unqualified products on automatic production lines), and agriculture (detection of locusts to prevent invasion).

Research into camouflaged objects detection, which has had a tremendous impact on advancing our knowledge of visual perception, has a long and rich history in biology and art. Two remarkable studies on camouflaged animals from Abbott Thayer[4] and Hugh Cott[5] are still hugely influential.

In recent years, camouflage object detection based on deep learning technology has become a research hotspot in the field of target testing. More and more deep learning camouflage target detection algorithms and models are proposed, the accuracy of camouflage target detection, timeliness, etc. are all obtained It continued to improve.

Most of the existing camouflage target detection methods based on deep learning are first adopted by convolutional neural networks, such as VGG [6] (Visual Geometry Group), Residual Neural Network, Res2Net [7]. Features, then use different strategies such as thick to fine, multi -task learning, confident perception learning, multi -source information fusion, Transformer and other strategies to further enhance the features, thereby improving the target detection performance. The current camouflage object detection methods are studied from the five perspectives of strategies, multi -tasking learning, confident perception learning, multi -source information integration, and Transformer strategy。

Cod is a fundamentally challenging task due to the fact that the camouflage strategy works by deceiving the visual perceptual system of the observer and thus a significant amount of visual perception knowledge [8] is required to eliminate the ambiguities caused by the high intrinsic similarities between the target object and the background.

In summary. This article proposes a model of a base injection mechanism -FDNET, which uses spatial attention and channel attention mechanism to enhance semantic information and characteristics, and uses a distraction to dig out the prospects and background interference to improve network performance Experiments show that the model proposed in this article should be better than the existing camouflage target detection model.

II. FDNET MODEL

The FDNET model proposed in this article mainly includes RFB modules, CSM modules and DMM modules. The specific structure is shown in Figure 1:

A. *RFB MODEL*

After the formal neuroscience experiments, human visual systems have a group of groups to feel wild (PRFS) to help a positive help for highlighting areas close to the center of the retina. Receptive Field Block) The input features expand the operation of the input characteristics to achieve the purpose of gaining enough field. The structure of the improved RFB module is shown in the figure 2 below.

Compared with the standard experience of the wild module (RFB), the improved RFB module uses asymmetric convolution to replace the original standard convolutional layer. And add a larger branch of the expansion rate to further increase the feeling of feeling. The improved RFB module uses a set of 1x3 and 3X1 convolution layers to replace the original 3X3 convolution. The goal of doing so is to reduce the calculation volume to increase the operation speed. We send the entered images $I \in R^{W*H*3}$ to the resonet-50 backbone for the extraction and extraction features of multi-level features f_k , $k \in \{1,2,3,4,5\}$, f_k including the diverse characteristics from high-resolution but weak semantics to low-resolution but strong semantics, and each each time The resolution of this feature is $H/2^k \times W/2^k$.

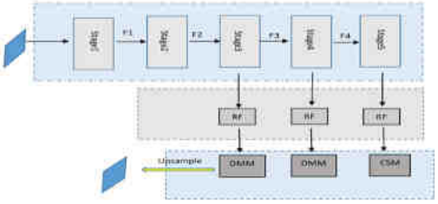


Figure 1 RFB MODEL

Manuscript received Oct 19, 2022

Bing Zhang, School of Software, Tiangong university, Tianjin ,China
Baoshan Sun, School of computer science and technology, Tiangong university, Tianjin ,China

Each RFB module after improvement contains four parallel residual branches $\{b_i, i = 1,2,3,4\}$. These four residual branches have different expansion rates and the setting of the expansion rate conforms to the HDC specifications -- $d \in \{1,3,5,7\}$ .In each branch, the first layer uses a 1x1 convolution operation to reduce the size of the channel to 32, followed by the asymmetric convolution layer of two (2i-1) X (2i-1) to reduce the amount of calculation amount Finally, a 3x3 empty convolutional layer. When I> 1, the expansion rate is set to (2i-1). Connect the first four branches $\{b_i, i = 1,2,3,4\}$ Later, we will reduce the channel size to C = 32 through a 3x3 convolutional layer, and then add an Indentity Shortcut branch and pass the entire module into a RELU function to obtain the final output feature $f_k^{'}$. The obtained output features will be passed into the next module for further processing.

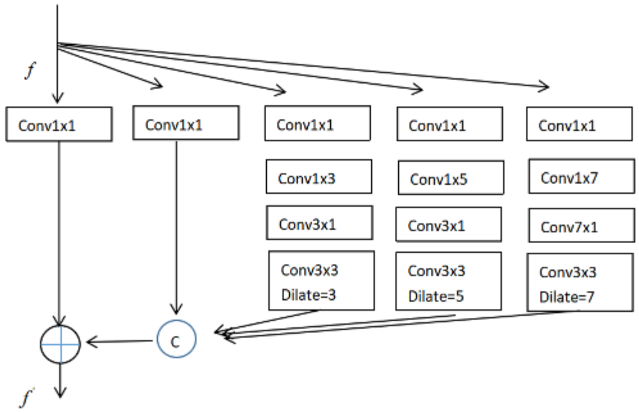


Figure 2 RFB MODEL

B. CSM MODEL

In the Channel Attention and Spatial Attention Module (CSM), we select the high -level features of the semantic enhancement of the RFB module for further processing to generate a initial segmentation diagram. The CSM module uses channel attention and spatial attention hybrid processing to capture the remote dependencies of images in the two positions of channels and space, so as to enhance the semantic representation of the highest level characteristics of the global angle. The structure of the CSM module is shown in Figure 3:

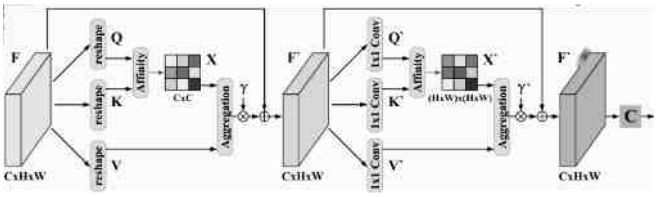


Figure 3 CSM MODEL

First of all, through the channel attention module, the shape of the input F is changed to get Q, K, V, and the size of C × N, n = H × W. Q and K's conversion execution matrix multiplication and pass the result into the SoftMax layer to get X: C × C.X and V perform a matrix multiplication, which changes the characteristics of the characteristics to obtain the characteristics of the C × H × W size. The result is multiplied to the learning proportional parameter γ (initial 1) and the connection with F is added to the output of the channel attention. Then use F' as the input of the space attention module, and the shape of three 1 × 1 convolutional layers will be changed to get Q', K ', and V'. Q', k's size becomes C/8 × N, and the size of V is C × N. The results obtained by the converter execution matrix multiplication of Q ‘and K 'are passed into the SoftMax layer calculation to get X‘: N X N. The matrix multiplication between V ′ and X ′ converter matrix changes the shape of the C × H × W size, then we use another scaling parameter γ '(initially 1) which is also added to F to get the final output F'. Finally, a convolution with a convolution kernel of 7×7 and a filling of 3 is applied to F ' 'to obtain the initial location map of the camouflage object, which is gradually refined by the subsequent module (DMM).

C. DMM MODEL

Disguised objects usually have a similar appearance to the background, so false positive and false negative predictions naturally occur in the initial segmentation results. The DMM module is designed to find and eliminate these wrong predictions. It takes the current level features, superior features and prediction results as input, and outputs refined features and more accurate prediction results. Humans are good at discriminating between areas of distraction after careful analysis and perform contextual reasoning[9], which compares patterns in areas of ambiguity with areas of confidence, such as texture and semantics, to make final decisions. Therefore, all predicted foreground (or background) regions are contextually explored in this paper to find false-positive distraction regions (or false-negative distraction regions) in the predicted foreground (or background) regions. First of all, the prediction of higher level is up-sampled, normalized by sigmoid, and duplicated by 1 matrix subtraction and inverse, respectively, and the current layer features are multiplied to generate foreground attention features F_f And the background attention feature F_b , feed into the parallel CE model(contex exploration) block for context inference, and detect false-positive distractions F_{fpd} and false-negative distractions F_{fnd} . The structure of the CSM module is shown in Figure 4:t

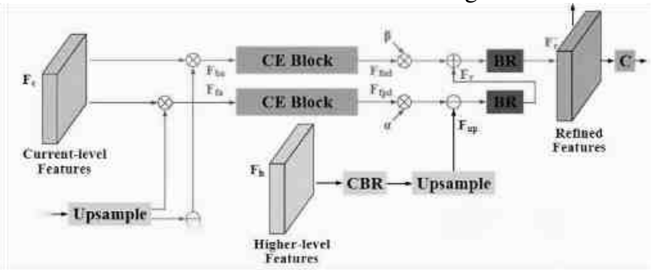


Figure 4 DMM MODEL

III. EXPERIMENTAL SETUP

A. Experimental Environment

We implement our model withthe PyTorch toolbox . An eight-core PC with an Intel Core i7-9700K 3.6 GHz CPU (with 32GB RAM) and an NVIDIA GeForce RTX 2060 GPU (with 6GB memory) is used for both training and testing.

B. Experimental data

Three datasets are used in the experiments in this chapter, namely CHAMELEON[10], CAMO[11] and CDO10K[12]. For the allocation of data sets, we used the training set of CAMO and COD10 as the training set, and the test set of CHAMELEON and CAMO and COD10K as the test set. CHAMELEON collected 76 images from the Google search engine using the keyword "camouflaged animal" and manually annotated them to achieve object-level truth maps. CAMO contains a total of 1250 camouflage images of different categories, which are divided into 1000 images for training set and 250 images for testing set. COD10K is the largest camouflaged object detection dataset to date, with a total of 10,000 images covering camouflaged objects in various natural scenes and more than 78 object categories.

C. Experimental evaluation standards

We ues four widely used and standard metrics to evaluate our method: structure-measure(Sα), e E-measure, F-measure, mean absolute error. Structure-measure(Sα) focuses on evaluating the structural information of the prediction maps, which is defined as equation (1). E-measure (Eφ) simultaneously evaluates the pixel-level matching and image-level statistics, as shown in equation (2); F-measure (Fβ) is a comprehensive measure on both the precision and recall of the prediction map, as shown in equation (3);The mean absolute errror is widely uesd in foreground-background segmentation tasks, in equation (4):

$$S_\alpha = \alpha S_o + (1-\alpha) S_r \qquad (1)$$

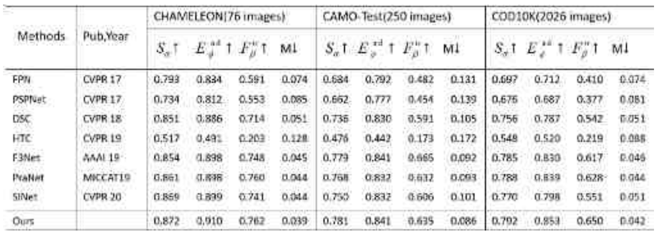
$$Q_{FM} = \frac{1}{w \times h} \sum_{x=1}^{w} \sum_{y=1}^{h} \phi_{FM}(x, y) \qquad (2)$$

$$F = (1+\beta^2) \cdot \frac{precision \cdot recall}{\beta^2 \cdot precision + recall} \qquad (3)$$

$$MAE = \frac{1}{W \times H} \sum\nolimits_{x=1}^{W} \sum\nolimits_{y=1}^{H} | S(x,y) - T(x,y) | \qquad (4)$$

C .Experimental results and analysis

Table 1 Comparison with seven other advanced methods on three datasets



Methods	Pub,Year	CHAMELEON(76 images)				CAMO-Test(250 images)				COD10K(2026 images)			
		$S_\alpha\uparrow$	$E_\phi^{ad}\uparrow$	$F_\beta^w\uparrow$	M↓	$S_\alpha\uparrow$	$E_\phi^{ad}\uparrow$	$F_\beta^w\uparrow$	M↓	$S_\alpha\uparrow$	$E_\phi^{ad}\uparrow$	$F_\beta^w\uparrow$	M↓
FPN	CVPR 17	0.793	0.834	0.591	0.074	0.684	0.792	0.482	0.131	0.697	0.712	0.410	0.074
PSPNet	CVPR 17	0.734	0.812	0.553	0.085	0.662	0.777	0.454	0.139	0.676	0.687	0.377	0.081
DSC	CVPR 18	0.851	0.886	0.714	0.051	0.736	0.830	0.591	0.105	0.756	0.787	0.542	0.051
HTC	CVPR 19	0.517	0.491	0.203	0.128	0.476	0.442	0.173	0.172	0.548	0.520	0.219	0.088
F3Net	AAAI 19	0.854	0.898	0.748	0.045	0.779	0.841	0.665	0.092	0.785	0.830	0.617	0.046
PraNet	MICCAT19	0.861	0.898	0.760	0.044	0.768	0.832	0.632	0.093	0.788	0.839	0.628	0.044
SINet	CVPR 20	0.869	0.899	0.741	0.044	0.750	0.832	0.606	0.101	0.770	0.798	0.551	0.051
Ours		0.872	0.910	0.762	0.039	0.781	0.841	0.635	0.086	0.792	0.853	0.650	0.042

Since there are not many camouflaged target detection models publicly available at present, we selected a total of 7 latest methods in related fields for comparative experiments. These methods include: Object detection method FPN[13], semantic segmentation method PSPNet[14], instance segmentation method HTC[15], shadow detection method DSC[16] medical image segmentation method PraNet[17], salient object detection method F3Net[18], and camouflaged object detection method SINet. The prediction maps for these methods are generated by retraining running open source code or are available on public websites. The experimental comparison results are shown in Table 1.

It can be seen that the method proposed in this paper has certain improvement compared with other methods in four indicators. Compared with the existing camouflage target detection model SINet, the indicators on CHAMELEON, CAMO and COD10K are improved by 2.1%, 2.9% and 9.9%, respectively. This shows that the proposed method has better performance in the current stage of camouflaged target detection. Medical image segmentation is similar to camouflaged target detection to a certain extent. Both of them need to accurately identify the target object and remove the interference influence of the surrounding environment under the condition that the object has a high similarity with the surrounding environment. We can see from the comparison of PraNet, a medical image segmentation method. Compared with PraNet, the method in this paper improved by 2%,3% and 3.2% respectively on the three data sets.

Since the features of the extended receptive field are fed into the CSM module through the RFB module and the channel attention and spatial attention mechanism are introduced, the long range semantic dependence of the camouflage object is found out, and the multi-scale initial location map of the camouflage object is provided for the DMM module. The DMM module can get the correct camouflage target area by resolving and eliminating the similar interference between the camouflaged object and the background area. Therefore, the detection effect of our method for camouflaged targets is better than that of the existing models.

CONCLUSION

This paper puts forward the model of camouflage target detection based on attention mechanism, by improving the receptive field module expanded and enhanced characteristics of the reception field said through a mixture of spatial attention and channel attention module to enhance the features of semantic rely on for a long time, and interference by eliminating the foreground region and background region accurately identify the camouflage target. The experimental results show that our method is due to other target detection models at the present stage. In the subsequent work, the semi-supervised learning strategy for camouflage target detection will be studied, and less data will be used to achieve more accurate results when the dataset is insufficient.

REFERENCES

[1] Martin Stevens and Sami Merilaita. Animal camouflage:current issues and new perspectives. Philosophical Transactions ofthe Royal Society B, 2009.

[2] Deng-Ping Fan, Ge-Peng Ji, Tao Zhou, Geng Chen, HuazhuFu, Jianbing Shen, and Ling Shao. Pranet: Parallel reverse attention network for polyp segmentation. In MICCAI, 2020.

[3] Deng-Ping Fan, Tao Zhou, Ge-Peng Ji, Yi Zhou, Geng Chen,Huazhu Fu, Jianbing Shen, and Ling Shao. Inf-net: Automatic covid-19 lung infection segmentation from ct images.IEEE TMI, 2020.

[4] G. H. Thayer and A. H. Thayer, Concealing-coloration in the Animal Kingdom: An Exposition of the Laws of Disguise Through Color and Pattern: Being a Summary of Abbott H. Thayer's Discoveries.Macmillan Company, 1909.

[5] H. B. Cott, Adaptive coloratcottion in animals. Methuen & Co., Ltd.,1940.

[6] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[C]// International Conference on Learning Representations, San Diego, May 7-9, 2015. Piscataway, NJ: IEEE Computer Society, 2015: 1-14

[7] GAO S H, Cheng M M, Zhao K, et al. Res2net: A new multi-scale backbone architecture[J]. IEEE transactions on pattern analysis and machine intelligence, 2019, 43(2): 652-662.

[8] Tom Troscianko, Christopher P Benton, P. George Lovell,David J Tolhurst, and Zygmunt Pizlo. Camouflage and visual perception. Philosophical Transactions ofthe Royal Society B, 2009.

[9] MEI H, Ji G P, Wei Z, et al. Camouflaged object segmentation with distraction mining[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, June 19-25, 2021. Pisca taway, NJ: IEEE Computer Society, 2021: 8772-8781

[10] P Skurowski, H Abdulameer, J Błaszczyk, T Depta, A Kornacki, and P Kozieł. Animal camouflage analysis:Chameleon database. Unpublished Manuscript, 2018.

[11] Trung-Nghia Le, Tam V Nguyen, Zhongliang Nie, Minh-Triet Tran, and Akihiro Sugimoto. Anabranch network forcamouflaged object segmentation. CVIU, 2019.

[12] Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng,Jianbing Shen, and Ling Shao. Camouflaged object detection. In CVPR, 2020.

[13] Tsung-Yi Lin, Piotr Doll´ar, Ross Girshick, Kaiming He,Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In CVPR, 2017.14 Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In CVPR, 2017.

[14] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In CVPR, 2017.

[15] Kai Chen, Jiangmiao Pang, Jiaqi Wang, Yu Xiong, Xiao xiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jianping Shi,Wanli Ouyang, et al. Hybrid task cascade for instance segmentation. In CVPR, 2019.

[16] Xiaowei Hu, Lei Zhu, Chi-Wing Fu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection. In CVPR, 2018.

[17] Deng-Ping Fan, Ge-Peng Ji, Tao Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Pranet: Parallel reverse attention network for polyp segmentation. In MICCAI, 2020.

[18] Jun Wei, Shuhui Wang, and Qingming Huang. F3net: Fusion, feedback and focus for salient object detection. In AAAI, 2020.