SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

June 2023, Vol. 66 162403:1–162403:9
https://doi.org/10.1007/s11432-022-3592-8

Double-branch fusion network with a parallel attention selection mechanism for camouflaged object detection

Junjiang XIANG[1,2], Qing PAN[1,2*], Zhengrong ZHANG[3], Songnian FU[1,2] & Yuwen QIN[1,2]

[1]*Advanced Institute of Photonics Technology, School of Information Engineering, Guangdong University of Technology, Guangzhou* 510006*, China;*
[2]*Guangdong Provincial Key Laboratory of Photonics Information Technology, Guangdong University of Technology, Guangzhou* 510006*,China;*
[3]*Guangxi Key Laboratory of Multimedia Communications and Network Technology, School of Computer Electronics and Information, Guangxi University, Nanning* 530004*, China*

Received 7 May 2022/Revised 9 August 2022/Accepted 23 September 2022/Published online 4 April 2023

Abstract To meet the challenge of camouflaged object detection (COD), which has a high degree of intrinsic similarity between the object and background, this paper proposes a double-branch fusion network (DBFN) with a parallel attention selection mechanism (PASM). In detail, a schismatic receptive field block (SRF) combined with an attention mechanism for low-level information is performed to learn texture features in one branch, and an integration of the SRF, a hybrid attention mechanism (HAM), and a depth feature polymerization module (DFPM) is employed for high-level information to extract detection features in the other branch. Then, both texture features and detection features are input into the PASM to acquire selective expression matrices. Eventually, the final result is obtained after further selective matrix optimization with atrous spatial pyramid pooling (ASPP) and a residual channel attention block (RCAB) being applied serially. Experimental results on three public datasets verify that our method outperforms the state-of-the-art methods in terms of four evaluation metrics, i.e., mean absolute error (MAE), weighted F_β measure (F_β^ω), structural measure (S_α), and E-measure (E_φ)

Keywords camouflaged object detection, attention mechanism, feature extraction, feature aggregation, texture information, fuzzy boundary

Citation Xiang J J, Pan Q, Zhang Z G, et al. Double-branch fusion network with a parallel attention selection mechanism for camouflaged object detection. Sci China Inf Sci, 2023, 66(6): 162403, https://doi.org/10.1007/s11432-022-3592-8

1 Introduction

Camouflaged object detection (COD) aims to identify objects whose peculiar textures are similar to those of the environmental background [1]. It is also the biggest difference between COD and conventional object detection. That is to say, compared with salient object detection (SOD) and other methods, COD is more challenging [2]. Because they are usually hidden in the environment, the camouflaged objects need more semantic segmentation for detection, which is a dichotomous problem to classify camouflage objects and backgrounds. Thus, feature extraction network design and feature aggregation technology are the focus in this field. Currently, research on COD is applied in a wide range of fields [3]. In addition to its academic value, it also helps to promote the invasion of locusts in agricultural remote sensing [2], the search and detection of camouflage concealed objects in the military [4,5], the judgment of disease in the medical field [6] and so on.

To conduct comprehensive research on COD, scholars at home and abroad have performed much work. Fan et al. [1] not only created the COD10k dataset that contains 10000 images and divides them into 78

* Corresponding author (email: panqing@gdut.edu.cn)

different categories to address the lack of large datasets in COD but also proposed a search identification network (SINet) solution. Ren et al. [7] presented residual refinement blocks (RRBs) to refine different levels of feature maps to enhance details and remove background noise to distinguish between background and camouflaged objects. Zhai et al. [8] advanced region-induced graph reasoning (RIGR) to mine the high-level dependencies between rough detection objects and edges. Subsequently, the edge-constricted graph reasoning module (ECGR) is employed to fuse boundary features and detection features. PFNet was proposed by Mei et al. [9], in which focusing modules (FM) are used to aggregate the features of adjacent layers. Then, the interference of both false positives and false negatives is gradually removed to improve the accurate segmentation of camouflaged objects. Li et al. [10] achieved a joint learning network that can extract the features in both camouflaged objects and salient objects. Then, the above two kinds of features are fused through a similarity measure to improve COD performance. A ranking network was proposed by Lv et al. [11], in which the camouflage decoding module is performed for feature aggregation to generate corresponding maps to enhance the COD accuracy.

From the view of feature extraction network design, which is the former focus of COD, each of the above methods proposes a new network structure suitable for COD. However, from the perspective of feature aggregation, side aggregation is adopted in the state-of-the-art methods mentioned above. That is, the features extracted from the backbone network are fused through the lateral aggregation module to realize the combination of high-level semantic features and low-level detailed features. Furthermore, compared with bidirectional aggregation for saliency detection [12, 13], the side aggregation method has a lower computation cost and seems more suitable for COD. However, the side aggregation approach does not consider the differences between low-level features and high-level features. This implies that the texture information of non-detection object features will be introduced into the detection object as noise, which results in a fuzzy boundary if all of the low-level detail features are mapped to high-level semantic features [14]. In addition, Liu et al. [15] proposed the adaptively spatial feature fusion (ASFF) module in the network structure to improve the problem of fuzzy boundaries, but this aggregation method is only applied in conventional detection.

To overcome the drawback of fuzzy boundaries in COD, a dual-branch aggregation structure was proposed. First, both the low-level information and the high-level information are performed by a schismatic receptive field block (SRF) to expand the receptive fields. Then, the former is directly concatenated to extract the enhanced texture features, while the latter is fused through a depth feature polymerization module (DFPM) after the hybrid attention mechanism (HAM) [16] operation to obtain the detection features. On the one hand, considering the incompatibility between low-level features and high-level features, it is necessary to design a reasonable mechanism to integrate low-level features and high-level features simultaneously. In this paper, an attention mechanism was used to assist the further fusion of high-level features and low-level features. In general, the spatial attention mechanism (SA) [17] and channel attention mechanism (CA) [18] are often employed to extract the spatial features with the greater expressive ability and extract the channel features with greater detecting ability, respectively. This means that if either of the two attention mechanisms is used individually, the features associated with the other one will be ignored. On the other hand, there is only one input interface in both SA and CA, which results in the inability to integrate the spatial features and the channel features at the same time when a serial framework is applied to connect the two attention mechanism modules [19]. Therefore, the structure of the parallel attention selective mechanism (PASM) is presented. Specifically, SA and CA simultaneously accept both the high-level features and the low-level features extracted from the previous module. After the concatenation and normalization by softmax, the selective expression matrix is generated. Then, with the product and superposition of the matrix and its input features, both the low-level features and the high-level features are obtained. Finally, the two kinds of features are fused by concatenation.

Taken together, the contributions made by this paper can be summarized as follows:

(1) A DFPM is presented to aggregate high-level features, in which the number of feature channels is increased to improve the low-dimension defect of single-channel features. Therefore, the information of the output characteristic channel is fully retained as much as possible.

(2) To improve the incompatibility between texture features and detection features, a PASM was constructed by CA and SA in parallel. In this structure, both texture features and detection features are simultaneously fed into CA and SA to enhance their spatial attributes and channel attributes to select the information to help detect.

(3) A double-branch fusion network (DBFN) with an attention mechanism was constructed, in which one branch is used to enhance texture information by SRF and SA, while DFPM combined with SRF and

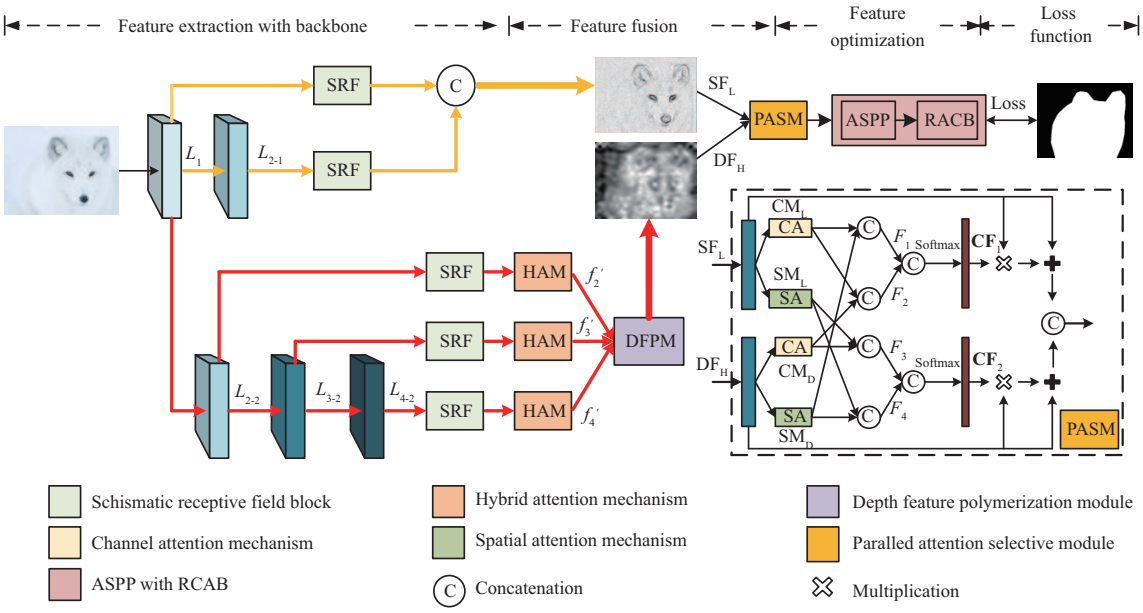

Figure 1 (Color online) The framework of a double-branch fusion network.

HAM is used to extract detection features from high-frequency feature maps in another branch, compared with the low-dimensional.

Section 2 describes our proposed network. Section 3 introduces our experiment and the experimental results. Finally, we summarize the work and look forward to future work in Section 4.

2 Proposed network

As shown in Figure 1, two kinds of features are extracted with the input image passing through backbone network [20] with the double branch. That is, texture features are enhanced by SRF in one branch, and the detection features can be acquired by the SRF combined with DFPM after HAM in the other branch. Subsequently, the detection information and texture information will be selected for better detection after PASM, which is composed of CA and SA in parallel construction. Finally, an atrous spatial pyramid pooling (ASPP) [21] and a residual channel attention block (RCAB) [22] are set as the optimization modules.

2.1 Schismatic receptive field block

Although the receptive field (RF) block [23] is used to expand the receptive fields of the object area and enhance texture information, the fixed receptive field expansion rate leads to the fuzzy problem in object detection. As shown in Figure 2, the input feature X can be divided into x_n at first,

$$x_1, x_2, x_3, x_4 = \text{Divided } (\boldsymbol{X}), \tag{1}$$

where x_n represents the nth sub-feature after the feature is cut, for $n \in [1, 2, 3, 4]$. Then, through convolution with kernels of different sizes, multiscale features are acquired,

$$F_n = \begin{cases} C^1_{(3,3)}(x_n), & n = 1, \\ C^{2n-1}_{(3,3)}(C^1_{(2n-1,1)}(C^1_{(1,2n-1)}(C^1_{(1,1)}(x_n)))), & n = 2, 3, 4, \end{cases} \tag{2}$$

where $C^k_{(i,j)}$ represents a convolution operation with kernel size $= (i, j)$ and dilation rate $= k$ to obtain multiple-scale features. To optimize the SRF module, a residual mechanism can be introduced as

$$\text{SRF}(\boldsymbol{X}) = \text{Res}(\boldsymbol{X}) + C^1_{(3,3)}\left(\text{Cat}^4_{n=1}(F_n)\right), \tag{3}$$

where, $\text{Cat}(\cdot)$ indicates a concatenation operation, and $\text{Res}(\cdot)$ indicates the residual structure [20]. A fast connection channel is designed by Res $(\cdot)$. Generally speaking, a saturated network is defined as

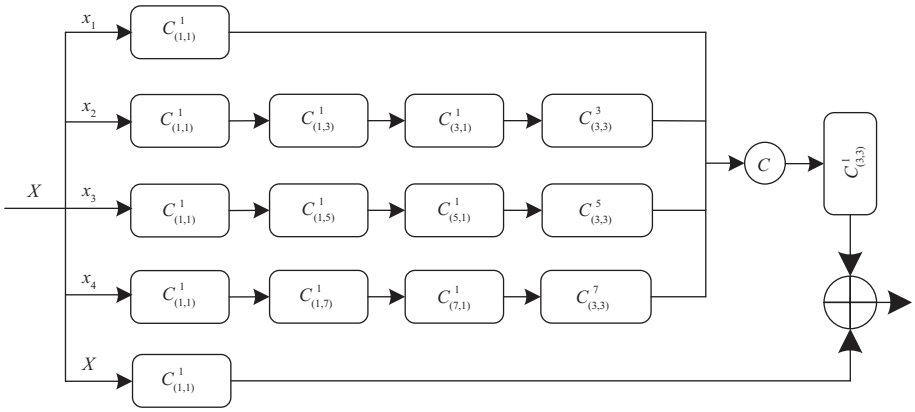

Figure 2 Schismatic receptive field block. © indicates a concatenation operation. ⊕ indicates an add operation.

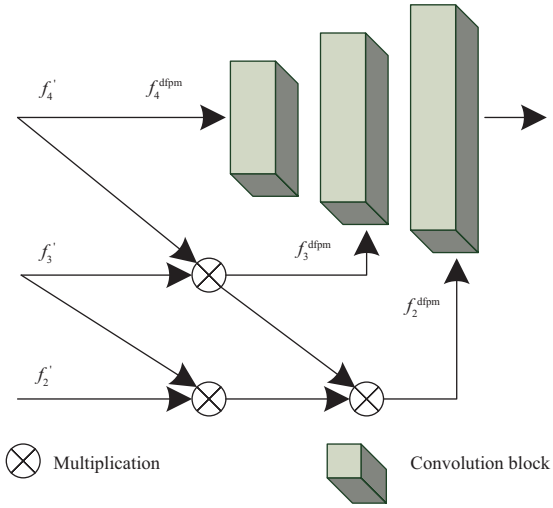

Figure 3 (Color online) Depth feature polymerization module.

$H(x) = x$, however, with the deepening of the network, the accuracy of the model will be saturated, and the model cannot continue to converge. The addition of a fast connection channel breaks the input mode of the network, and $H(x)$ can be defined as $H(x) = \text{Res}(x) + x$, that is to say, $\text{Res}(x) = H(x) - x$. The purpose of the residual structure is to make $H(x)$ infinitely close to x.

2.2 Depth feature polymerization module

In [2], all the features are mapped into one channel. However, this approach greatly limits the feature expression and is not conducive to subsequent feature extraction. Therefore, a DFPM is proposed in this paper, as shown in Figure 3. That is, only four layers of the backbone network are used as the generation feature network, so the DFPM is specifically expressed as

$$\begin{cases} f_4^{\text{dfpm}}[\text{ inSize; }] = f_4', \\ f_3^{\text{dfpm}} = f_3' \otimes \text{BC}[\text{up}_{\uparrow}^2(f_4')\,; \boldsymbol{W}^1], \\ f_2^{\text{dfpm}}[;\text{ outSize }] = f_2' \otimes \text{BC}[\text{up}_{\uparrow}^2(f_3^{\text{dfpm}}); \boldsymbol{W}^2] \otimes \text{BC}[\text{up}_{\uparrow}^2(f_3'); \boldsymbol{W}^3], \end{cases} \tag{4}$$

where $\text{up}_{\uparrow}^2$ denotes the up-sampling operation (e.g., 2 times), $\text{BC}[\,; \boldsymbol{W}^u]$ denotes a 3×3 convolution operation and a batch normalization operation, $\boldsymbol{u} \in [1, 2, 3]$ is defined as the order of $\boldsymbol{W}$, $\boldsymbol{W}$ represents the feature maps of different layers, and $f_k^{\text{dfpm}}[;]$ represents the feature aggregation function. inSize represents the input channel of the module, and outSize represents the output channel of the module. After the feature aggregation is finished, multi-dimensional features are fed into PASM for feature selection. This approach avoids the problem of insufficient information retention for a single-channel feature.

2.3 Parallel attention selection module

A parallel attention mechanism of feature selection is proposed to mine the influence of feature attributes on the network and aggregate features. The low-level features are defined as $\mathrm{SF_L}$, and the high-level features are defined as $\mathrm{DF_H}$. Then, the symmetrical double-branch network structure is designed. Using the low-level features as an example, $\mathrm{SF_L}$ is first fed into a convoluted block with kernel size $= 3 \times 3$, and then we enhance the spatial attributes by the SA and channel attributes by the CA; that is, the feature passes through both the spatial attention mechanism and channel attention mechanism. The features are defined as $\mathrm{SM_L}$ and $\mathrm{CM_L}$:

$$\mathrm{SM_L} = \mathrm{SA}\left(\mathrm{BConv}\left(\mathrm{SF_L}\right)\right), \tag{5}$$

$$\mathrm{CM_L} = \mathrm{CA}\left(\mathrm{BConv}\left(\mathrm{SF_L}\right)\right). \tag{6}$$

By using unhide partitioning, the spatial attention feature of detection features and the channel attention feature of detection features can be generated. To fuse the spatial features and channel features effectively, the concatenation method is used to generate four groups of fusion features: F_1, F_2, F_3, and F_4.

Then, the four groups of features are divided into two groups. The two groups of features for output are mapped to the single feature map, and the features are normalized by the softmax function to generate a selective representation matrix of features (e.g., $\mathbf{CF}_1, \mathbf{CF}_2$). This is formulated as follows:

$$\mathbf{CF}_1 = \mathrm{SoftMax}\left(\mathrm{BConv}\left(\mathrm{Cat}\left(F_1, F_2\right)\right)\right), \tag{7}$$

$$\mathbf{CF}_2 = \mathrm{SoftMax}\left(\mathrm{BConv}\left(\mathrm{Cat}\left(F_3, F_4\right)\right)\right). \tag{8}$$

After the matrix of selective expressions is multiplied by the input feature, the selective expression of the feature is obtained. To optimize the feature selective expression network, a residual mechanism is introduced:

$$X_1 = \mathrm{Mul}\left(\mathbf{CF}_1, \mathrm{SF_L}\right) + \mathrm{SF_L}, \tag{9}$$

$$X_2 = \mathrm{Mul}\left(\mathbf{CF}_2, \mathrm{DF_H}\right) + \mathrm{DF_H}, \tag{10}$$

where $\mathrm{Mul}(\cdot)$ is the multiplication operation. Finally, the high-level features and low-level features are fused by connecting them.

3 Experiments

3.1 Datasets and evaluation metrics

In this paper, three benchmark datasets are used in COD: CHAMELEON [24], CAMO-test [25], and COD10K-test. Four common evaluation metrics are used in these experiments: the mean absolute error (MAE), weighted F_β measure (F_β^ω) [26], structural measure ($S\alpha$) [27], and E-measure (E_φ) [28]. All algorithms are operated on Windows 10, Python 3.6, and PyTorch. An RTX 3090 GPU is used for acceleration. Other settings include the Adam algorithm, 100 iterations, and 1E−4 learning rate.

3.2 Ablation study

Feature extraction of the network structure. In this subsection, the low-level features are selectively mapped into the high-level features, as shown in Table 1. Similar to the work in [15], the features obtained from three layers are put into a convolution block at first, the convolution block gives the three layers different weight coefficients, and finally feature aggregation is achieved. However, there are still some texture features of the non-detected object in the third layer network, as shown in Figure 4. Texture information will interfere with the excessive detection feature. To verify our point of view, the following schemes were designed. (1) Low-level information obtained from the 1st layer is combined with high-level features. (2) Low-level information obtained from the 1st and 2nd layers is combined with high-level features. (3) Low-level information obtained from the 1st–3rd layers is combined with high-level features. The result shows that the secondary fusion strategy performs best.

Feature selection of the attention mechanism. We explored the influence of the attention mechanism structure on the experimental results. According to Table 2, good performance occurs when we selectively use spatial and channel attention mechanisms.

Figure 4 (Color online) Feature diagram in the third layer of neural network.

Table 1 Ablation studies of the network structure

	CAMO-test			
	$S_\alpha \uparrow$	$E_\varphi \uparrow$	$F_\beta^\omega \uparrow$	MAE $\downarrow$
1	0.812	0.867	0.737	0.075
2	0.822	0.878	0.749	0.069
3	0.808	0.864	0.731	0.075

Table 2 Ablation studies of the mechanism structure

	CAMO-test			
	$S_\alpha \uparrow$	$E_\varphi \uparrow$	$F_\beta^\omega \uparrow$	MAE $\downarrow$
All CA	0.807	0.860	0.727	0.076
All SA	0.802	0.858	0.722	0.076
Ours	0.822	0.878	0.749	0.069

Table 3 Comparison with existing methods

	CHAMELEON				CAMO-test				COD10K-test			
	$S_\alpha \uparrow$	$E_\varphi \uparrow$	$F_\beta^\omega \uparrow$	MAE $\downarrow$	$S_\alpha \uparrow$	$E_\varphi \uparrow$	$F_\beta^\omega \uparrow$	MAE $\downarrow$	$S_\alpha \uparrow$	$E_\varphi \uparrow$	$F_\beta^\omega \uparrow$	MAE $\downarrow$
FPN [29]	0.794	0.783	0.590	0.075	0.684	0.677	0.483	0.131	0.697	0.691	0.411	0.075
UNet++ [30]	0.695	0.762	0.501	0.094	0.599	0.653	0.392	0.149	0.623	0.672	0.350	0.086
PFANet [19]	0.679	0.648	0.378	0.144	0.659	0.622	0.391	0.172	0.636	0.618	0.286	0.128
PoolNet [31]	0.776	0.779	0.555	0.081	0.702	0.698	0.494	0.129	0.705	0.713	0.416	0.074
CPD [32]	0.853	0.866	0.706	0.052	0.726	0.729	0.550	0.115	0.747	0.770	0.508	0.059
EGNet [33]	0.848	0.870	0.702	0.050	0.732	0.768	0.583	0.104	0.737	0.779	0.509	0.056
SINet [1]	0.869	0.891	0.740	0.044	0.751	0.771	0.606	0.100	0.771	0.806	0.551	0.051
PraNet [6]	0.860	0.907	0.763	0.044	0.769	0.824	0.663	0.094	0.789	0.861	0.629	0.045
RankNet [11]	**0.893**	0.938	–	0.033	0.793	0.826	–	0.085	0.793	0.868	–	0.041
MCIFNet [34]	–	–	–	–	0.784	0.845	0.677	0.084	0.787	0.872	0.636	0.042
TANet [7]	0.888	0.911	0.786	0.036	0.793	0.834	0.690	0.083	0.803	0.848	0.629	0.041
SINet_v2 [2]	0.888	0.942	0.816	0.030	0.820	**0.882**	0.743	0.070	0.815	0.887	0.680	0.037
Ours	0.892	**0.944**	**0.831**	**0.030**	**0.822**	0.878	**0.749**	**0.069**	**0.821**	**0.893**	**0.698**	**0.034**

3.3 Comparison with existing methods

Several methods are compared in terms of their experimental results, namely, SINet [1], SINet_v2 [2], ParNet [6], TANet [7], RankNet [11], PFANet [19], FPN [29], UNet++ [30], PoolNet [31], CPD [32], EGNet [33], and MCIFNet [34]. Table 3 shows that for the F_β^ω, S_α, MAE, and E_φ evaluation metrics, the proposed algorithm achieves better performance than the other algorithms. Particularly in the COD10K-test and CHAMELEON datasets, the proposed algorithm improves F_β^ω by approximately 1.5%. And FLOPs and Params are used to evaluate the complexity of models in Table 4. We can find that the proposed method has superior performance under the same complexity. Figure 5 shows the performance of different COD methods in terms of the PR curve and the methods are PSPNet [35], PiCANet [36], MSRCNN [37], MaskRCNN [38] and so on. The closer the PR curve is to the upper right, the better the performance of the model. If the PR curve of the model is completely covered by the PR curve of another's, the performance of the latter is better than the former. As shown in Figure 5, the PR curve

Table 4 The complexity analysis of models

	UNet++ [30]	PFANet [19]	PoolNet [31]	CPD [32]	EGNet [33]	MCIFNet [34]	PraNet [6]	RankNet [11]	SINet [1]	SINet_v2 [2]	Ours
FLOPs (G)	65.85	68.71	92.60	17.74	294.80	30.13	13.07	17.42	19.47	12.24	14.02
Params (M)	9.16	16.29	68.26	47.85	111.65	75.72	30.49	28.72	48.94	24.92	24.72

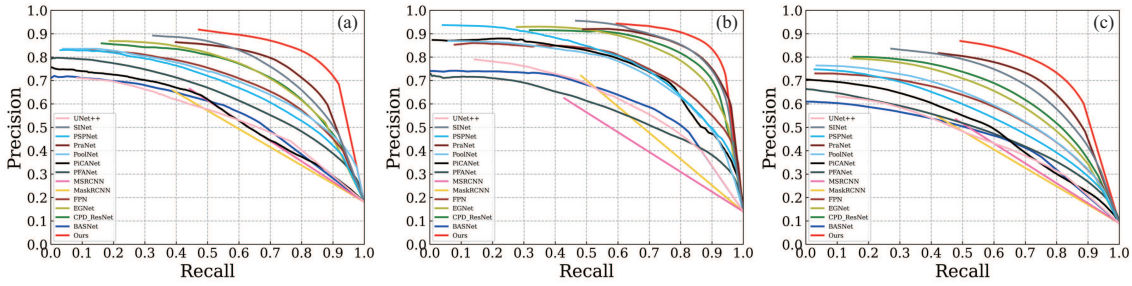

Figure 5 (Color online) PR-curve. (a) CAMO; (b) CHAMELEON; (c) COD.

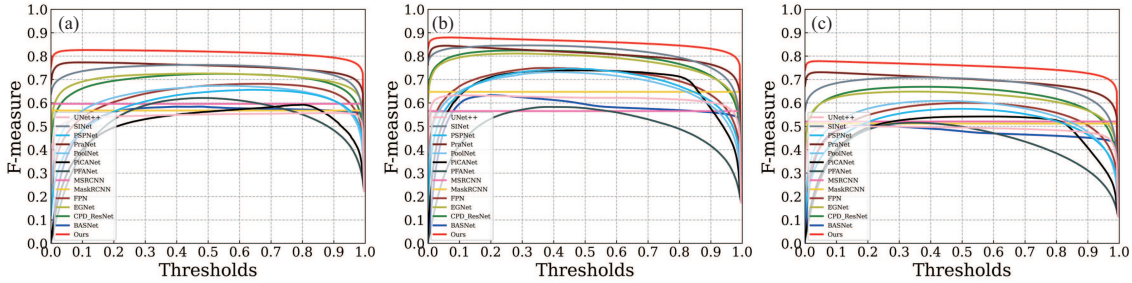

Figure 6 (Color online) F-measure curve. (a) CAMO; (b) CHAMELEON; (c) COD.

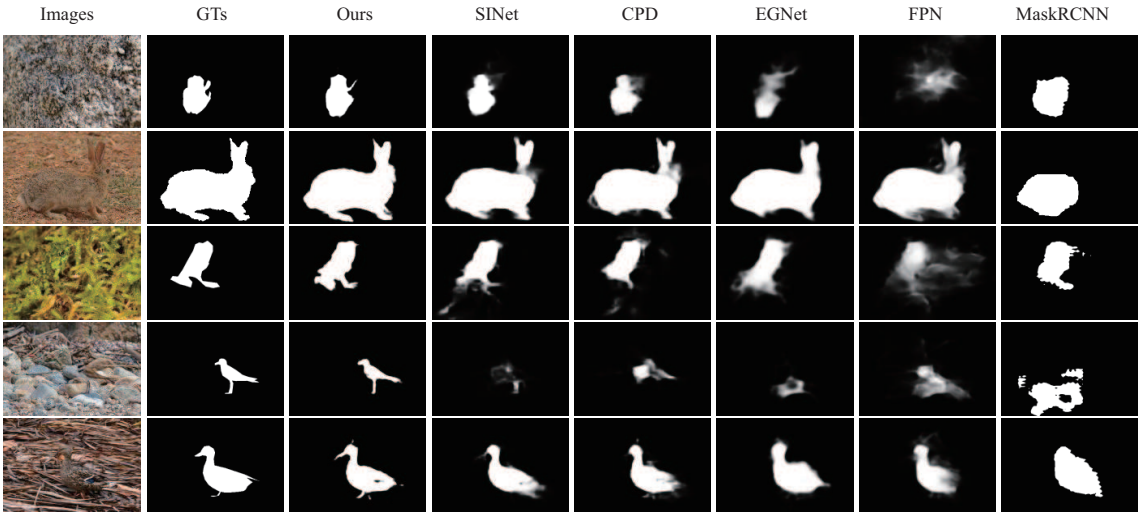

Figure 7 (Color online) Comparison with other methods.

of ours is at the top right of all models and the PR curve of ours has covered others completely. Figure 6 shows the COD performance of F-measure under different thresholds, and we can see that our scheme has superior performance under different thresholds. Figure 7 shows the detection results of our method and other algorithms. In our scheme, the shadow caused by error detection is less than others. This is because the feature selection is adopted, and only the texture features of the object can be fused into the detection features.

4 Conclusion

To solve the incompatibility between high-level features and low-level features in the fusion of camouflage object detection, this paper optimizes the algorithm from two levels: feature extraction and feature fusion. The experimental results show that our method with a feature selection mechanism can detect objects better than other methods. New techniques [39, 40], such as Transformer, and reinforcement learning, could also be explored in the feature work.

Acknowledgements This work was partially supported by National Key R&D Program of China (Grant No. 2018YFB18011001), National Natural Science Foundation of China (Grant No. 62025502), Guangdong Introducing Innovative and Entrepreneurial Teams of 'The Pearl River Talent Recruitment Program' (Grant No. 2019ZT08X340), and Guangdong Guangxi Joint Science Key Foundation (Grant No. 2021GXNSFDA076001).

References

1 Fan D P, Ji G P, Sun G, et al. Camouflaged object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020. 2777–2787
2 Fan D P, Ji G P, Cheng M M, et al. Concealed object detection. IEEE Trans Pattern Anal Mach Intell, 2021. doi: 10.1109/TPAMI.2021.3085766
3 Zheng Y, Zhang X, Wang F, et al. Detection of people with camouflage pattern via dense deconvolution network. IEEE Signal Process Lett, 2018, 26: 29–33
4 Chu H K, Hsu W H, Mitra N J, et al. Camouflage images. ACM Trans Graph, 2010, 29: 1–8
5 Ge S, Jin X, Ye Q, et al. Image editing by object-aware optimal boundary searching and mixed-domain composition. Comp Visual Media, 2018, 4: 71–82
6 Fan D P, Ji G P, Zhou T, et al. PraNet: parallel reverse attention network for polyp segmentation. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, 2020. 263–273
7 Ren J, Hu X, Zhu L, et al. Deep texture-aware features for camouflaged object detection. IEEE Trans Circuits Syst Video Technol, 2021. doi: 10.1109/TCSVT.2021.3126591
8 Zhai Q, Li X, Yang F, et al. Mutual graph learning for camouflaged object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021. 12997–13007
9 Mei H, Ji G P, Wei Z, et al. Camouflaged object segmentation with distraction mining. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021. 8772–8781
10 Li A, Zhang J, Lv Y, et al. Uncertainty-aware joint salient object and camouflaged object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021. 10071–10081
11 Lv Y, Zhang J, Dai Y, et al. Simultaneously localize, segment and rank the camouflaged objects. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021. 11591–11601
12 Feng M, Lu H, Ding E. Attentive feedback network for boundary-aware salient object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019. 1623–1632
13 Qin X, Zhang Z, Huang C, et al. BASNet: boundary-aware salient object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019. 7479–7489
14 Chen Z, Xu Q, Cong R, et al. Global context-aware progressive aggregation network for salient object detection. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2020. 34: 10599–10606
15 Liu S, Huang D, Wang Y. Learning spatial fusion for single-shot object detection. 2019. ArXiv:1911.09516
16 Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV), 2018. 3–19
17 Chen L, Zhang H, Xiao J, et al. SCA-CNN: spatial and channel-wise attention in convolutional networks for image captioning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017. 5659–5667
18 Wang Q, Wu B, Zhu P, et al. ECA-Net: efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020. 11531–11539
19 Zhao T, Wu X. Pyramid feature attention network for saliency detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019. 3085–3094
20 Gao S H, Cheng M M, Zhao K, et al. Res2Net: a new multi-scale backbone architecture. IEEE Trans Pattern Anal Mach Intell, 2019, 43: 652–662
21 Yang M, Yu K, Zhang C, et al. DenseASPP for semantic segmentation in street scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018. 3684–3692
22 Basak H, Kundu R, Agarwal A, et al. Single image super-resolution using residual channel attention network. In: Proceedings of 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS), 2020. 219–224
23 Liu S, Huang D. Receptive field block net for accurate and fast object detection. In: Proceedings of the European Conference on Computer Vision (ECCV), 2018. 385–400
24 Skurowski P, Abdulameer H, Błaszczyk J, et al. Animal camouflage analysis: CHAMELEON database. http://kgwisc.aei.polsl.pl/datasets/CamouflageBase/animals.7z
25 Le T N, Nguyen T V, Nie Z, et al. Anabranch network for camouflaged object segmentation. Comput Vision Image Understanding, 2019, 184: 45–56
26 Margolin R, Zelnik-Manor L, Tal A. How to evaluate foreground maps? In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014. 248–255
27 Fan D P, Cheng M M, Liu Y, et al. Structure-measure: a new way to evaluate foreground maps. In: Proceedings of the IEEE International Conference on Computer Vision, 2017. 4548–4557

28 Fan D P, Gong C, Cao Y, et al. Enhanced-alignment measure for binary foreground map evaluation. In: Proceedings of International Joint Conference on Artificial Intelligence, 2018. 698–704

29 Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017. 2117–2125

30 Zhou Z, Siddiquee M M R, Tajbakhsh N, et al. UNet++: a nested U-Net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham: Springer, 2018. 3–11

31 Liu J J, Hou Q, Cheng M M, et al. A simple pooling-based design for real-time salient object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019. 3917–3926

32 Wu Z, Su L, Huang Q. Cascaded partial decoder for fast and accurate salient object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019. 3907–3916

33 Zhao J X, Liu J J, Fan D P, et al. EGNet: edge guidance network for salient object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019. 8779–8788

34 Dong B, Zhuge M, Wang Y, et al. Towards accurate camouflaged object detection with mixture convolution and interactive fusion. 2021. ArXiv:2101.05687

35 Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017. 6230–6239

36 Liu N, Han J, Yang M-H. PiCANet: learning pixel-wise contextual attention for saliency detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018. 3089–3098

37 Huang Z, Huang L, Gong Y, et al. Mask scoring R-CNN. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019. 6409–6418

38 He K, Gkioxari G, Dollar P, et al. Mask R-CNN, In: Proceedings of the IEEE International Conference on Computer Vision, 2017. 2961–2969

39 Mao Y, Zhang J, Wan Z, et al. Generative transformer for accurate and reliable salient object detection. 2022. ArXiv:2104.10127

40 Hou F Y, Sun J, Yang Q L, et al. Deep reinforcement learning for optimal denial-of-service attacks scheduling. Sci China Inf Sci, 2022, 65: 162201