IEEE Int. Conf. on Multimedia and Expo (ICME), Niagara Falls, Canada, July 2024

Fuzzy Boundary-Guided Network for Camouflaged Object Detection

1st Qi Jia
Dalian University of Technology
Dalian, China
jiaqi@dlut.edu.cn

2nd Shuilian Yao
Dalian University of Technology
Dalian, China
Shuilian_Yao@mail.dlut.edu.cn

3rd Youcan Xu
Dalian University of Technology
Dalian, China
youcanxv@163.com

4th Yu Liu*
Dalian University of Technology
Dalian, China
liuyu8824@dlut.edu.cn

5th Dehao Kong
Dalian University of Technology
Dalian, China
20202241050@mail.dlut.edu.cn

6th Longin Jan Latecki
Temple University
Philadelphia, USA
latecki@temple.edu

***Abstract*—Camouflaged object detection (COD) is a challenging task that identifies camouflaged objects from highly similar backgrounds. Existing methods typically treat the whole object equally while neglecting the indistinguishable regions that require more attention than other regions. In this paper, we propose a Fuzzy Boundary-Guided Network (FBG-Net) for camouflaged object detection, which mimics the human behavior that pays more attention to these low-confidence regions when observing objects. Specifically, we devise two main building blocks: (1) Mixed Semantics Aggregation Module (MSAM) to integrate boundary and texture features cumulatively in the high-to-low scales, and (2) Fuzzy Boundary-Guided Module (FBGM) to locate and enhance the low-confidence regions under the guidance of fuzzy boundary. Extensive experiments demonstrate the effectiveness of FBG-Net with superior performance to existing state-of-the-art methods. https://github.com/YAOSL98/FBG-Net**

***Index Terms*—camouflaged object detection, fuzzy boundary, low-confidence regions**

I. Introduction

Identifying a camouflaged object from its background, known as camouflaged object detection (COD), is a fundamental task that facilitates various applications, such as animal conservation [1], medical image analysis [2], and image synthesis [3]. Recent works [4]–[6] have achieved a new state-of-the-art performance on all COD benchmarks. However, existing detecting results still suffer from incomplete and even false boundaries, as objects are highly similar to the background in ambiguous regions, especially when objects are small, or heavily obscured.

The reason of low detection accuracy in indistinguishable regions is that these methods treat the whole object equally while ignoring that different regions have different detection difficulty levels. In fact, a biological study validates that when humans detect camouflaged objects, they spend more search time in the "difficult visual environment" [7]. In other words, humans pay more attention to the indistinguishable regions. They tend to zoom in and out to find discriminative clues on different image scales.

* Corresponding author

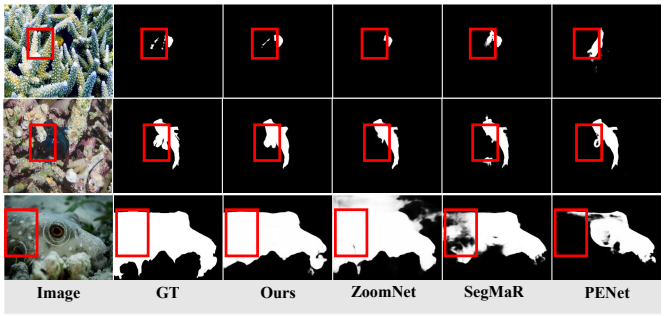

Fig. 1. Detection results with low-confidence regions. Our method exhibits more accurate results compared with the state-of-the-art methods (ZoomNet [8], SegMaR [6], and PENet [9]), as shown in red boxes.

To improve the detection accuracy in these uncertain regions, some methods introduce additional network or training data. UGTR [10] employs an additional network to estimate the uncertainties, while these uncertain regions may be inconsistent with that of the final prediction network. UJSC [11] introduces salient object detection (SOD) training dataset and an adversarial learning framework between the SOD task and COD task to impose greater penalties on the inconsistent detection results. Similarly, ZoomNet [8] proposes an uncertainty-aware loss. However, these methods fail to explicitly detect and enhance the uncertain regions that coincide with the detection result.

Fig. 1 demonstrates the cutting-edge approaches successfully localize objects, but still fail to clearly detect the camouflaged objects on blurred boundaries or hard-to-distinguish regions, labeled by red boxes. Consequently, our work aims to solve a question: *how to locate and focus on the indistinguishable regions explicitly at different scales for accurate detection?*

In this paper, we propose a Fuzzy Boundary-Guided Network (FBG-Net) for camouflaged object detection, which explicitly explores uncertain regions by fuzzy boundary detection and clustering. Specifically, to detect and preserve

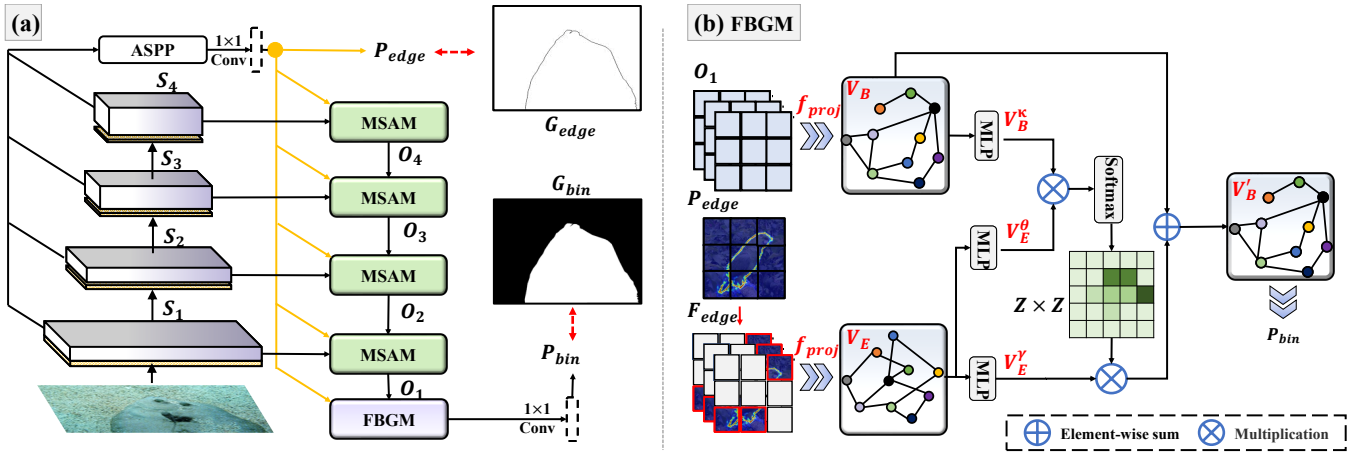

Fig. 2. (a) Overview of our FBG-Net and (b) the illustration of FBGM . G_{bin} and G_{edge} represent the ground truth of the object and the corresponding boundary, respectively.

subtle boundaries in different feature scales, we aggregate the boundary information coupled with texture features from high to low scales by the proposed Mixed Semantics Aggregation Module (MSAM), which imitates the coarse-to-fine observation process of humans. Moreover, the embedded multi-scale Feature Magnification and Fusion Unit (FMFU) is designed to explore fine-grained edge structures for indistinguishable details. Finally, we design a Fuzzy Boundary-Guided Module (FBGM) to locate and strengthen the low-confidence regions by graph-based techniques. This targeted enhancement strategy enables the proposed FBG-Net to focus on low-confidence regions during training, rendering fine-grained detection of camouflaged objects.

Our contributions are summarized as follows:

- We propose a Fuzzy Boundary-Guided Network (FBG-Net), which brings uncertainty guidance into COD task.
- We develop a new Fuzzy Boundary-Guided Module (FBGM) to explicitly capture fuzzy boundaries and utilize graph-based techniques to focus on low-confidence regions.
- Our FBG-Net achieves new records on three commonly used benchmarks, i.e., CAMO [12], NC4K [13], and COD10K [3].

II. Related Work

A. Camouflaged Object Detection

In recent decades, camouflaged object detection (COD), as a task of identifying camouflaged objects from their surroundings, has gained great attention from the computer vision community. Valuable attempts can be summarized into three aspects. The first kind of method constructs advanced network architectures or modules to explore distinguishable features for COD [14]. The second kind excavates extra and valuable clues, such as edge or fixation-relative masks, from the shared features and combines them into the joint learning or multi-task learning frameworks using cross-modality fusion techniques [5], [12], [13]. The last kind belongs to bio-inspired approaches, and the network design or learning strategy is inspired by the behavior of predators or human visual perception [3], [15]. In addition, there are also works that notice the impact of uncertain region on camouflage object detection. UGTR [16] employs a probabilistic representational model to gain an uncertain map, while further in combination with transformer to explicitly reason the final detection result. Li *et al.* [17] introduce a similarity measure to explicitly model the contradicting attribute of SOD and COD, rendering an adversarial learning network to estimate the pixel-wise confidence as uncertainty. However, both methods generate the uncertain regions upon the whole image, easily leading to false enhancement in the background regions. Our method belongs to the bio-inspired approach, which imitates the behavior of humans when observing hard-to-distinguish regions of camouflage objects.

In contrast to previous approaches, our method detects the low-confidence region by locating the fuzzy boundaries, and targeted strengthen the hard-to-distinguish regions at different feature scales.

B. boundary-guided Network

Boundary guidance aims to extract extra and valuable edge features and incorporate them into camouflage object features using cross-modality fusion techniques. The ideas have been widely used in object detection, including salient object detection [18] and camouflaged object detection [4]. As the boundary prior can contribute to localization and segmentation for object detection [19], [20], recent SOD methods [18], [20] and COD methods [4], [5] extract the essential edge features and incorporate them into object features to accurately localize, segment, and detect objects. Nevertheless, these methods all treat the target object boundary as a whole, neglecting the indistinguishable regions, which require more attention.

Different from the existing entire-boundary guidance network, our FBG-Net only focuses on part of the boundaries, which is more concentrated on guiding and enhancing relative features for high-precision object detection.

C. Multi-scale Feature Fusion Network

Multi-scale feature fusion network aims to explore object-related clues in multi-scale features and fuse these features for camouflage object prediction. This idea has been widely used in SOD [21] and COD [8]. Existing multi-layer pyramid feature extraction structures [22], [23] are prone to lose many texture and structure details, which are unsuitable for dense prediction tasks [24], [25] that emphasize the integrity of regions and edges. Therefore, recent CNN-based COD methods [3], [5], [14], [15] and SOD methods [26], [27] employ the inter-layer features to enhance the feature representation. To enhance boundary guidance in COD tasks, recent methods blend both edge and texture features to learn discriminative mixed-scale semantics [4].

Unlike them, our work magnifies the coupled edge and texture features upon each feature scale, integrates the mixed features in the current scale and accumulative scales separately to preserve fine-grained edge structure, and consolidates the corresponding features.

III. Proposed method

Overview. The overall architecture of our FBG-Net is illustrated in Fig. 2 (a). Given an RGB image, we first feed it into the backbone to extract multi-level features $S_i(i = 1, 2, 3, 4)$. These features are further fed into an Atrous Spatial Pyramid Pooling (ASPP) [28] for boundary prediction map P_{edge}. Then, the Mixed Semantics Aggregation Modules (MSAM) couples both boundary and texture clues cumulatively along the high-to-low feature scale. Finally, the Fuzzy Boundary-Guided Module (FBGM) leverages the boundary prediction P_{edge} and aggregated feature O_1 to explore fuzzy boundaries and enhance the uncertain regions for the final prediction map P_{bin}. This section details the proposed modules and loss function.

A. Mixed Semantics Aggregation Module (MSAM)

To aggregate the boundary features and multi-scale texture features, we design a Mixed Semantics Aggregation Module (MSAM) to gather these features cumulatively along high-to-low scale features, which imitates the human observation process from global to local detail. As shown in Fig. 3, MSAM is composed of two branches. The boundary prediction map P_{edge} is used to enhance the features S_i in the current scale and the features O_{i+1} in previous accumulative scales, respectively. We employ concatenation followed by channel attention to explore valuable feature channels and improve feature representation for the accumulative features O_i as output. The output O_i of MSAM can be formulated as

$$O_i = CA(CAT(F_{conv}(S_i) \oplus S_i', F(O_{i+1}) \oplus O_{i+1}')), \quad (1)$$

where $CA(\cdot)$ is a channel attention, $CAT(\cdot)$ is concatenation, $F_{conv}(\cdot)$ is a 3×3 convolution, and $\oplus$ is an element-wise sum.

For each branch, we design FMFU to magnify the texture feature and boundary prediction map P_{edge} into different scales. Fig. 3 illustrates the detailed structure of FMFU. Especially, for the branch with input $S_i(i = 1, 2, 3, 4)$ and P_{edge}, we evenly divide S_i into four feature maps $S_i^j(j = 1, 2, 3, 4)$ along channel dimension. Then four feature groups are resized to 1.25x, 1.50x, 1.75x, and 2.00x scales respectively.

To enhance the boundary-related regions, we resize P_{edge} to corresponding scales, as labeled by the red dots in Fig. 3, and strengthen S_i^j by bit-wise multiplication. Finally, we feed these aggregated features of each scale into different 3×3 dilated convolutions followed by down-sampling. The dilation rate is $d(d = 1, 3, 5, 7)$, and final output is $S_i^{'}$.

The output S_i' of FMFU can be formulated as

$$S_i' = CAT(D(F_{conv}^d(S_e \otimes U(S_i^j)))), \quad (2)$$

where $CAT(\cdot)$ is concatenation, $D(\cdot)$ is down-sampling, $F_{conv}^d(\cdot)$ is 3×3 dilation convolution with a dilation rate of d, $\otimes$ is element-wise multiplication and $U(\cdot)$ is up-sampling operation. Similarly, we can obtain O_{i+1}' by FMFU.

B. Fuzzy Boundary-Guided Module (FBGM)

FBGM aims to locate and enhance the hard-to-distinguish regions under the guidance of fuzzy boundary, which is illustrated in Fig. 2 (b). Firstly, we generate a fuzzy boundary mask by the confidence level in the predicted boundary map P_{edge}. Then, we project the features of the local uncertain region and the whole image into graph space, represented as $\mathcal{V}_E$ and $\mathcal{V}_B$, to establish global-range dependency relationships between them. Finally, cross-graph interaction is employed to output the enhanced features $\mathcal{V}_B^{'}$ and final prediction P_{bin}.

1) Fuzzy Boundary Extraction: We divide the boundary prediction map P_{edge} into high-confidence and low-confidence regions, where the latter is the region that needs to be enhanced.

Specifically, for each pixel $x_k, (k = 1, 2, ..., K)$, where K is the number of pixels in P_{edge}, we normalize the probability value $p(x_k)$ into the range of $[0, 1]$. A high value indicates a high confidence of the boundary pixel. Then we divide these pixels into different clusters N_s by

$$N_s = \{x_m, x_n \in N_s | dist(\frac{x_n}{e^{p(x_n)}}, \frac{x_m}{e^{p(x_m)}}) \leq \epsilon\}, \quad (3)$$

where $m, n \in \{1, 2, ..., K\}$, x_n and x_m represent 2D pixel coordinates, $s \in \{1, 2, ..., S\}$, S is the total number of clusters, $dist$ is Euclidean distance, and ϵ is a threshold, $\epsilon = 160$.

Eq. 3 clusters pixels with higher probability and close position into the same class. Once the number of pixels in N_s is larger than a threshold γ (γ=15 in our experiment), N_s is recognized as the high-confidence class, otherwise N_s is recognized as the low-confidence class. Moreover, we divide the P_{edge} into $P \times P$ patches ($P = 12$ in our experiments). A patch containing pixels from a low confidence class is identified as an uncertain region as demonstrated by red boxes in Fig. 2 (b). We preserve the uncertain regions and mask off the high-confidence regions, and use 1×1 convolution to render the fuzzy boundary detection feature $F_{edge} \in \mathbb{R}^{h \times w \times C}$.

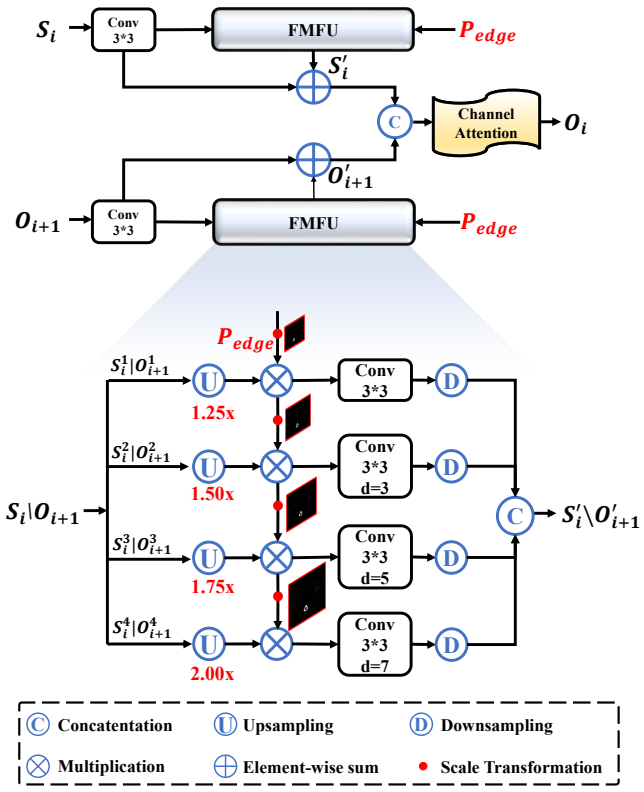

Fig. 3. The architecture of the proposed MSAM and FMFU.

2) Graph Projection f_{proj}: We project the feature of fuzzy boundary $F_{edge} \in \mathbb{R}^{h\times w\times C}$ and feature of the whole image $O_1 \in \mathbb{R}^{h\times w\times C}$ into graph node representations $\mathcal{V}_E \in \mathbb{R}^{C\times Z}$ and $\mathcal{V}_B \in \mathbb{R}^{C\times Z}(Z = h \times w)$ by graph projection f_{proj} [29]. The vertices of the graph is defined as cluster centers of feature maps, and the graph edges measure the similarity between these clusters in a feature space. Graph Projection renders larger receptive fields than traditional convolution learning networks, establishing relations between the whole features.

3) Cross-Graph Interaction: To enhance the features in the uncertain region, we employ non-local attention operation to establish the relations between two graphs $\mathcal{V}_E$ and $\mathcal{V}_B$. We use multi-layer perceptions (MLPs) to transform $\mathcal{V}_E$ to the key graph $\mathcal{V}_E^{\theta}$ and the value graph $\mathcal{V}_E^{\gamma}$, while $\mathcal{V}_B$ is transformed to the query graph $\mathcal{V}_B^{\kappa}$. The similarity matrix $\mathbb{A}^{inter}_{\mathcal{V}_E\rightarrow\mathcal{V}_B} \in \mathbb{R}^{(Z\times Z)}$ is calculated by a matrix multiplication as

$$\mathbb{A}^{inter}_{\mathcal{V}_E\rightarrow\mathcal{V}_B} = f_{norm}({\mathcal{V}_B^{\kappa}}^T \times \mathcal{V}_E^{\theta}), \tag{4}$$

where f_{norm} represents the softmax operation. Consequently, we obtain the enhanced feature $\mathcal{V}'_B$ by

$$\mathcal{V}'_B = \mathbb{A}^{inter}_{\mathcal{V}_E\rightarrow\mathcal{V}_B} \times {\mathcal{V}_E^{\gamma}}^T + \mathcal{V}_B. \tag{5}$$

Finally, we map the graph representations $\mathcal{V}'_B$ back to the original feature space, and use 1×1 convolution operation to generate the final prediction map P_{bin}.

C. Loss Function

Our FBG-Net is trained end-to-end by two loss terms: the dice loss L_{dice} and the structure loss L_{str} [36]. L_{str} includes a weighted binary cross entropy loss L_{wbce} and a IoU loss L_{wiou}, defined as $L_{str} = L_{wbce} + L_{wiou}$. The overall loss is

$$L_{total} = L_{dice}(P_{edge}, G_{edge}) + L^w_{str}(P_{bin}, G_{bin}), \tag{6}$$

where G_{bin} and G_{edge} denote object binary ground truth annotations and corresponding boundary ground truth, respectively.

IV. Experiments

A. Experiment Setting

Our FBG-Net is implemented with PyTorch on an NVIDIA A40 GPU 48G. We employ pre-trained Swin Transformer [37] on ImageNet as our backbone. We resize all the input images to 384×384. The whole training stage takes about 2 hours with batch sizes of 16 and 30 epochs. We evaluate our FBG-Net on three widely used benchmarks: CAMO [12], COD10K [3] and NC4K [13]. Our training set includes 1,000 images from the CAMO and 3,040 images from COD10K, and the test set merges 250 images from CAMO, 2,026 images from COD10K, and 4,121 images from NC4K.

We employ four evaluation metrics, including mean absolute error (M), weighted F-measure (F^w_β) [38], structure-measure (S_α) [39] and mean E-measure(E_m) [40]. M is defined as the element-wise difference between the prediction map and binary ground truth. S_α is defined as $S_\alpha = \alpha S_o + (1-\alpha)S_r$, where S_o is object-aware structural similarity and S_r is region-aware structural similarity. E_m evaluates the pixel-level similarity and image-level statistic, which is related to human visual perception. F^w_β is a measure of both precision and recall. F^w_β is more comprehensive and reliable than F-measure.

B. Comparison with State-of-the-arts

We compare FBG-Net with 16 state-of-the-art COD methods, including the CNN-based methods and the Transformer-based methods. For a fair comparison, all the results are provided by their authors or computed using released codes, and we train and test them with the same evaluation protocol.

Tab. I shows that FBG-Net outperforms other methods on three benchmarks. Specifically, FBG-Net outperforms the best CNN-based method PENet [9] by an average of 14.33% on M, and the best Transformer-based method EVP [35] by an average of 10.43% on M.

Fig. 4 shows a visual comparison of FBG-Net with other two boundary-guided methods BGNet [4] and MGL [5]. We leverage red arrows to indicate fuzzy boundaries in Fig. 4 (c) and compare the prediction maps and GT for clear visualization in the following columns, where the red color indicates false predictions. By contrast, wrong predictions in red typically appear around the fuzzy boundaries for all methods, validating the uncertain regions need additional attention in COD task. Moreover, due to our targeted enhancement strategy, FBG-Net exhibits the fewest red regions compared to BGNet and MGL.

TABLE I

QUANTITATIVE COMPARISON WITH OTHER COD STATE-OF-THE-ART METHODS ON THREE BENCHMARKS USING FOUR WIDELY USED EVALUATION METRICS (I.E., S_α, E_m, F_β^w, AND M). "↑" / "↓" INDICATES THAT LARGER/SMALLER IS BETTER. THE BEST RESULTS ARE HIGHLIGHTED IN RED.

Method	Pub.'Year	CAMO-Test (250 images)				COD10K-Test (2,026 images)				NC4K (4,121 images)			
		$S_\alpha \uparrow$	$E_m \uparrow$	$F_\beta^w \uparrow$	$\mathcal{M} \downarrow$	$S_\alpha \uparrow$	$E_m \uparrow$	$F_\beta^w \uparrow$	$\mathcal{M} \downarrow$	$S_\alpha \uparrow$	$E_m \uparrow$	$F_\beta^w \uparrow$	$\mathcal{M} \downarrow$
CNN-Based Models													
SINet [3]	CVPR'20	0.745	0.804	0.644	0.092	0.776	0.864	0.631	0.043	0.808	0.871	0.723	0.058
MGL [5]	CVPR'21	0.772	0.806	0.664	0.089	0.811	0.844	0.654	0.037	0.829	0.862	0.731	0.055
LSR [13]	CVPR'21	0.787	0.838	0.696	0.080	0.804	0.880	0.673	0.037	0.840	0.895	0.766	0.048
PFNet [15]	CVPR'21	0.782	0.841	0.695	0.085	0.800	0.877	0.660	0.040	0.829	0.887	0.745	0.053
UJSC [11]	CVPR'21	0.800	0.873	0.728	0.073	0.809	0.891	0.684	0.035	0.842	0.907	0.771	0.047
BASNet [30]	AAAI'22	0.794	0.851	0.717	0.079	0.817	0.891	0.699	0.034	0.841	0.897	0.771	0.048
SegMaR [6]	CVPR'22	0.805	0.864	0.724	0.072	0.813	0.880	0.682	0.035	0.844	0.905	0.773	0.047
BGNet [4]	IJCAI'22	0.812	0.870	0.749	0.073	0.831	0.901	0.722	0.033	0.851	0.907	0.788	0.044
ZoomNet [8]	CVPR'22	0.820	0.892	0.752	0.066	0.838	0.911	0.729	0.029	0.853	0.912	0.784	0.043
FEDER-R50 [31]	CVPR'23	0.807	0.873	0.785	0.069	0.823	0.900	0.740	0.032	0.846	0.905	0.817	0.045
PENet [9]	IJCAI'23	0.828	0.890	0.771	0.063	0.831	0.908	0.723	0.031	0.855	0.912	0.795	0.042
Transformer-Based Models													
VST [32]	ICCV'21	0.807	0.848	0.713	0.081	0.820	0.879	0.698	0.037	0.845	0.893	0.767	0.048
UGTR [10]	ICCV'21	0.785	0.822	0.685	0.086	0.818	0.852	0.667	0.035	0.839	0.874	0.746	0.052
TPRNet [33]	TVCJ'22	0.807	0.861	0.725	0.074	0.817	0.887	0.683	0.036	0.846	0.898	0.768	0.048
ICON [34]	PAMI'22	0.838	0.894	0.769	0.058	0.818	0.904	0.688	0.033	0.847	0.911	0.784	0.045
EVP [35]	CVPR'23	0.846	0.895	0.777	0.059	0.843	0.907	0.742	0.029	0.863	0.919	0.800	0.041
Ours	-	0.855	0.916	0.809	0.051	0.845	0.919	0.747	0.028	0.872	0.927	0.820	0.036

TABLE II

ABLATION ANALYSIS OF OUR MODULES. B: BASELINE. M: MIXED SEMANTICS AGGREGATION MODULE. F: FUZZY BOUNDARY-GUIDED MODULE. THE BEST RESULTS ARE HIGHLIGHTED IN RED.

Method	FLOPs (G)	Params (M)	CAMO-Test				COD10K-Test				NC4K			
			$S_\alpha \uparrow$	$E_m \uparrow$	$F_\beta^{w\uparrow}$	$\mathcal{M} \downarrow$	$S_\alpha \uparrow$	$E_m \uparrow$	$F_\beta^{w\uparrow}$	$\mathcal{M} \downarrow$	$S_\alpha \uparrow$	$E_m \uparrow$	$F_\beta^{w\uparrow}$	$\mathcal{M} \downarrow$
B	53.83	92.99	0.805	0.863	0.733	0.072	0.817	0.891	0.694	0.036	0.845	0.899	0.775	0.049
B+F	56.17	93.09	0.818	0.905	0.763	0.054	0.826	0.893	0.714	0.033	0.851	0.914	0.804	0.046
B+M	64.12	99.69	0.842	0.910	0.789	0.053	0.833	0.912	0.740	0.029	0.869	0.921	0.816	0.039
B+F+M	66.46	99.79	0.855	0.916	0.809	0.051	0.845	0.919	0.747	0.028	0.872	0.927	0.820	0.036

C. Ablation Study

We validate the effectiveness of the Fuzzy Boundary-Guided Module (FBGM) and Mixed Semantics Aggregation Module (MSAM) in Tab. II. On the first row, we only preserve the ASPP module and remove both MSAM and FBGM, where MSAM is replaced by several up-sampling and concatenation operations, and O_1 is the final prediction.

Compared to the basic setting on the first row, adding FBGM and MSAM gains 7.23% and 19.9% performance improvement on M respectively, which demonstrates the effectiveness of fuzzy boundary guidance and the rationality of integrating boundary with texture features by multi-scale operations. The last row validates the combination of both FBGM and MSAM achieves the best performance. Additionally, we report corresponding floating point operations (FLOPs) and number of parameters in Tab. II, showcasing the efficiency of our design with small parameters achieving significant performance enhancements.

To validate the effectiveness of FBG-Net on challenging images with indistinguishable regions, we select a sub-dataset from CAMO, COD10K, and NC4K test sets, where the number of uncertain patches in each image is more than 5. We use M as the evaluation metric which depicts the element-wise difference between the prediction map and binary ground truth directly. The sub-dataset contains 51, 136, and 752 images from the above datasets, respectively. Tab. III shows the sub-dataset involves higher detection difficulty with larger M than the original test sets. Compared with two competitive methods, i.e., ZoomNet and ICON, our method has remarkable improvements on the subset, which surpasses these two methods by 17.20% and 16.91% in terms of M on average, while the percentage of improvement is only 14.15% and 15.74% on the original test sets. By contrast, our FBG-Net renders superior performance on hard-to-detect images than cutting-edge methods.

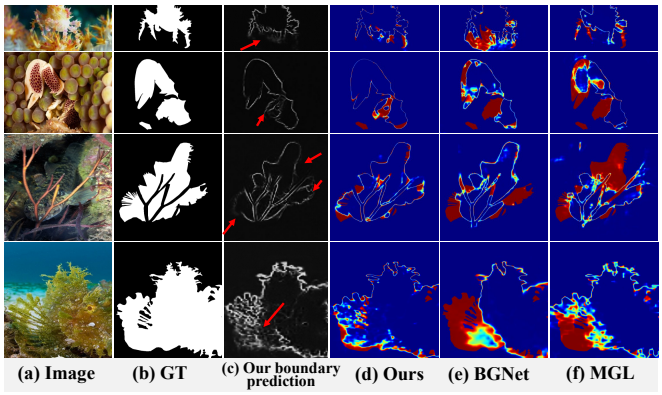

Fig. 4. Visual comparison of FBG-Net with BGNet [4] and MGL [5]. Red arrows in (c) point to fuzzy boundaries. Red color in (d) (e) (f) indicates lower similarity between GT and predictions while blue color indicates higher similarity.

TABLE III
PERFORMANCE RESULTS ON SUBSETS WHERE THE NUMBER OF UNCERTAIN PATCHES IS MORE THAN 5 ON $M \downarrow$. $(\cdot/\cdot)$ DENOTES (RESULTS ON ORIGINAL TEST SETS / RESULTS ON TEST SUB-SETS).

Methods	CAMO-Test (250 / 51)	COD10K-Test (2,026 / 136)	NC4K (4,121 / 752)
ZoomNet	0.066 / 0.093	0.029 / 0.044	0.043 / 0.055
ICON	0.058 / 0.083	0.033 / 0.045	0.045 / 0.059
Ours	0.051 / 0.067	0.028 / 0.040	0.036 / 0.047

V. CONCLUSION

To focus on the indistinguishable regions for COD, we propose an effective Fuzzy Boundary-Guided Network (FBG-Net) according to the observed behavior of humans. FBG-Net cumulatively integrates boundary and texture features in multi-scales by the designed Mixed Semantics Aggregation Module (MSAM) and explicitly locates and enhances low-confidence regions by the designed Fuzzy Boundary-Guided Module (FBGM). Extensive experiments demonstrate a superior performance of the proposed approach on three COD benchmarks.

REFERENCES

[1] R. Pérez-de la Fuente, X. Delclòs, E. Peñalver, M. Speranza, J. Wierzchos, C. Ascaso, and M. S. Engel, "Early evolution and ecology of camouflage in insects," *Proceedings of the National Academy of Sciences*, vol. 109, no. 52, pp. 21 414–21 419, 2012.

[2] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Pranet: Parallel reverse attention network for polyp segmentation," in *MICCAI*. Springer, 2020, pp. 263–273.

[3] D.-P. Fan, G.-P. Ji, G. Sun, M.-M. Cheng, J. Shen, and L. Shao, "Camouflaged object detection," in *CVPR*, 2020, pp. 2777–2787.

[4] Y. Sun, S. Wang, C. Chen, and T.-Z. Xiang, "Boundary-guided camouflaged object detection," *arXiv preprint arXiv:2207.00794*, 2022.

[5] Q. Zhai, X. Li, F. Yang, C. Chen, H. Cheng, and D.-P. Fan, "Mutual graph learning for camouflaged object detection," in *CVPR*, 2021, pp. 12 997–13 007.

[6] Q. Jia, S. Yao, Y. Liu, X. Fan, R. Liu, and Z. Luo, "Segment, magnify and reiterate: Detecting camouflaged objects the hard way," in *CVPR*, 2022, pp. 4713–4722.

[7] P. A. Todd, H. Phua, and K. B. Toh, "Interactions between background matching and disruptive colouration: Experiments using human predators and virtual crabs," *Current Zoology*, vol. 61, no. 4, pp. 718–728, 08 2015.

[8] Y. Pang, X. Zhao, T.-Z. Xiang, L. Zhang, and H. Lu, "Zoom in and out: A mixed-scale triplet network for camouflaged object detection," in *CVPR*, 2022, pp. 2160–2170.

[9] X. Li, J. Yang, S. Li, J. Lei, J. Zhang, and D. Chen, "Locate, refine and restore: A progressive enhancement network for camouflaged object detection," in *IJCAI*. ijcai.org, 2023, pp. 1116–1124. [Online]. Available: https://doi.org/10.24963/ijcai.2023/124

[10] F. Yang, Q. Zhai, X. Li, R. Huang, A. Luo, H. Cheng, and D.-P. Fan, "Uncertainty-guided transformer reasoning for camouflaged object detection," in *ICCV*, 2021, pp. 4146–4155.

[11] A. Li, J. Zhang, Y. Lv, B. Liu, T. Zhang, and Y. Dai, "Uncertainty-aware joint salient object and camouflaged object detection," in *CVPR*, 2021, pp. 10 066–10 076.

[12] T.-N. Le, T. V. Nguyen, Z. Nie, M.-T. Tran, and A. Sugimoto, "Anabranch network for camouflaged object segmentation," *CVIU*, vol. 184, pp. 45–56, 2019.

[13] Y. Lv, J. Zhang, Y. Dai, A. Li, B. Liu, N. Barnes, and D.-P. Fan, "Simultaneously localize, segment and rank the camouflaged objects," in *CVPR*, 2021, pp. 11 591–11 601.

[14] Y. Sun, G. Chen, T. Zhou, Y. Zhang, and N. Liu, "Context-aware cross-level fusion network for camouflaged object detection," *arXiv preprint arXiv:2105.12555*, 2021.

[15] H. Mei, G.-P. Ji, Z. Wei, X. Yang, X. Wei, and D.-P. Fan, "Camouflaged object segmentation with distraction mining," in *CVPR*, 2021, pp. 8772–8781.

[16] F. Yang, Q. Zhai, X. Li, R. Huang, A. Luo, H. Cheng, and D.-P. Fan, "Uncertainty-guided transformer reasoning for camouflaged object detection," in *ICCV*, 2021, pp. 4146–4155.

[17] A. Li, J. Zhang, Y. Lv, B. Liu, T. Zhang, and Y. Dai, "Uncertainty-aware joint salient object and camouflaged object detection," in *CVPR*, 2021, pp. 10 071–10 081.

[18] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, and M. Jagersand, "Basnet: Boundary-aware salient object detection," in *CVPR*, 2019, pp. 7479–7489.

[19] P. Zhang, D. Wang, H. Lu, H. Wang, and X. Ruan, "Amulet: Aggregating multi-level convolutional features for salient object detection," in *ICCV*, 2017, pp. 202–211.

[20] J.-X. Zhao, J.-J. Liu, D.-P. Fan, Y. Cao, J. Yang, and M.-M. Cheng, "Egnet: Edge guidance network for salient object detection," in *ICCV*, 2019, pp. 8779–8788.

[21] Z. Wu, L. Su, and Q. Huang, "Cascaded partial decoder for fast and accurate salient object detection," in *CVPR*, 2019, pp. 3907–3916.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *TPAMI*, vol. 37, no. 9, pp. 1904–1916, 2015.

[23] X. Zhao, Y. Pang, J. Yang, L. Zhang, and H. Lu, "Multi-source fusion and automatic predictor selection for zero-shot video object segmentation," in *ACMMM*, 2021, pp. 2645–2653.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015, pp. 234–241.

[25] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, 2015, pp. 3431–3440.

[26] Y. Pang, X. Zhao, L. Zhang, and H. Lu, "Multi-scale interactive network for salient object detection," in *CVPR*, 2020, pp. 9413–9422.

[27] W. Ji, G. Yan, J. Li, Y. Piao, S. Yao, M. Zhang, L. Cheng, and H. Lu, "Dmra: Depth-induced multi-scale recurrent attention network for rgb-d saliency detection," *TIP*, vol. 31, pp. 2321–2336, 2022.

[28] M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang, "Denseaspp for semantic segmentation in street scenes," in *CVPR*, 2018, pp. 3684–3692.

[29] Y. Li and A. Gupta, "Beyond grids: Learning graph representations for visual recognition," in *NeurIPS*, 2018.

[30] H. Zhu, P. Li, H. Xie, X. Yan, D. Liang, D. Chen, M. Wei, and J. Qin, "I can find you! boundary-guided separated attention network for camouflaged object detection," in *AAAI*, vol. 36, no. 3, 2022, pp. 3608–3616.

[31] C. He, K. Li, Y. Zhang, L. Tang, Y. Zhang, Z. Guo, and X. Li, "Camouflaged object detection with feature decomposition and edge reconstruction," in *CVPR*, 2023, pp. 22 046–22 055.

[32] N. Liu, N. Zhang, K. Wan, L. Shao, and J. Han, "Visual saliency transformer," in *ICCV*, 2021, pp. 4722–4732.

[33] Q. Zhang, Y. Ge, C. Zhang, and H. Bi, "Tprnet: camouflaged object detection via transformer-induced progressive refinement network," *TVC*, pp. 1–15, 2022.

[34] M. Zhuge, D.-P. Fan, N. Liu, D. Zhang, D. Xu, and L. Shao, "Salient object detection via integrity learning," *TPAMI*, vol. 45, no. 3, pp. 3738–3752, 2022.

[35] W. Liu, X. Shen, C.-M. Pun, and X. Cun, "Explicit visual prompting for low-level structure segmentations," in *CVPR*, 2023, pp. 19 434–19 445.

[36] J. Wei, S. Wang, and Q. Huang, "F^3net: fusion, feedback and focus for salient object detection," in *AAAI*, vol. 34, no. 07, 2020, pp. 12 321–12 328.

[37] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *ICCV*, 2021, pp. 10 012–10 022.

[38] R. Margolin, L. Zelnik-Manor, and A. Tal, "How to evaluate foreground maps?" in *CVPR*, 2014, pp. 248–255.

[39] D.-P. Fan, M.-M. Cheng, Y. Liu, T. Li, and A. Borji, "Structure-measure: A new way to evaluate foreground maps," in *ICCV*, 2017, pp. 4548–4557.

[40] D.-P. Fan, G.-P. Ji, X. Qin, and M.-M. Cheng, "Cognitive vision inspired object segmentation metric and loss function," *SSI*, vol. 6, no. 6, 2021.