Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Lightweight Boundary-region Activation Network for Camouflaged Object Detection

Mengjiao LU
Nanjing Normal University
Zerui Jin
Nanjing University
Wanjun Wang
Nanjing Normal University
Mingyong Pang
panion@netease.com

Nanjing Normal University

Research Article

Keywords:

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11768954/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Lightweight Boundary-region Activation Network for Camouflaged Object Detection

Mengjiao LU[1], Zerui Jin[2], Wanjun Wang[1], Mingyong Pang[1*]

[1]Institute of EduInfo Science and Technology, Nanjing Normal University, No. 122 Ninghai Road, Nanjing, 210024, Jiangsu, China.
[2]School of Electronic Science and Engineering, Nanjing University, Nanjing, 210033, China.

*Corresponding author(s). E-mail(s): panion@netease.com;
Contributing authors: lumengjiao@nnu.edu.cn;

Abstract

Camouflaged object detection (COD) aims to segment objects that closely resemble their surroundings, where ambiguous boundaries and distracting background textures complicate accurate object delineation. Within lightweight constraints, effective refinement requires not only structural boundary evidence but also a spatial decision on where additional feature enhancement is needed. To this end, we propose LBANet, a lightweight boundary-region activation network that converts structural cues and coarse prediction states into learnable binary spatial guidance. Specifically, a wavelet edge prior module (WEPM) integrates Haar high-frequency responses with deep semantic features to construct structural boundary evidence. A spike-inspired binary activation module (SBAM) further combines this evidence with prediction uncertainty and mask gradients to generate a binary boundary-region gate through a straight-through estimator. The gate modulates shallow features, which are re-aggregated by a shared lightweight aggregation module and subsequently processed by a dual refinement module (DRM) for coordinated mask and edge refinement. Experiments on four COD benchmarks demonstrate that LBANet achieves competitive performance with only 2.05M parameters and 2.66G FLOPs, demonstrating the effectiveness of boundary-region activation for accurate and lightweight COD. Code is available at https://github.com/AaHa123/LBANet.

Keywords: Computer vision, Camouflaged object detection, Lightweight segmentation, Region and boundary exploration

1 Introduction

Camouflaged object detection (COD) aims to segment foreground objects that exhibit strong visual similarity to their surrounding backgrounds. Indistinguishable colors, repetitive textures, and ambiguous foreground-background transitions make accurate object localization and region recovery particularly challenging. These difficulties become more obvious in lightweight COD, where compact feature representations must preserve both semantic object information and local structural details. COD has attracted increasing attention in computer vision and has been applied to medical segmentation [1–3], industrial defect inspection [4, 5], agricultural pest monitoring [6, 7] and artistic image synthesis [8].

Recent deep COD methods have substantially improved detection accuracy through multi-scale feature aggregation [9, 10], boundary-guided representation learning [11–13] and frequency-domain attention [14–18]. Boundary-aware approaches, such as BGNet [19] and GBNet [20] demonstrate the importance of structural information for recovering object contours and preserving fine details. While frequency-aware methods exploit complementary high-frequency representations for structural enhancement [14–18]. Meanwhile, lightweight frameworks, including TinyCOD [21], FINet [22] and FMLNet [23] improve the balance between segmentation accuracy and model complexity through compact architectures and feature interactions. These advances motivate a related question: how can structural evidence and the current prediction state jointly determine where additional feature refinement should be applied? In camouflaged scenes, high-frequency responses may originate from object contours, background textures, or internal object patterns. Therefore, combining boundary-related evidence with the current prediction state provides a basis for selecting spatial regions for additional feature updates.

Motivated by bio-inspired visual attention, we address this problem from the perspective of prediction-conditioned boundary-region activation. To translate boundary information into region-level refinement guidance, we combine structural evidence with the current segmentation state to construct an explicit spatial activation decision. As shown in Fig. 1, selective visual attention enables the visual system to prioritize informative regions based on visual evidence and contextual information [24]. Contour integration suggests how local structural responses contribute to coherent boundary perception [25], while figure-ground segregation highlights the role of contextual boundary information in distinguishing objects from their surroundings [26, 27]. We further propose LBANet, a lightweight boundary-region activation network for camouflaged object detection in a point–line–region–object progression. Specifically, a lightweight backbone and a lightweight aggregation module (LAM) first generate a coarse object prediction. A wavelet edge prior module (WEPM) then extracts Haar high-frequency components from shallow features and integrates them with high-level semantics to construct boundary evidence. Based on this representation, a spike-inspired binary boundary activation module (SBAM) combines edge evidence, prediction uncertainty, and mask gradients to produce a binary boundary-region gate. A straight-through estimator enables end-to-end optimization of the discrete activation decision. The gate modulates low-level features, which are subsequently re-aggregated by the shared LAM to enhance boundary-associated representations



Fig. 1 Illustration of the motivation about proposed boundary-region activation framework. From local structural cues to boundary-region refinement.

without introducing a separate aggregation network. Finally, a dual refinement module (DRM) jointly refines region and edge predictions to promote object completeness and contour consistency. Through this design, LBANet establishes a unified progression from local structural perception to boundary-region activation and object-level refinement.

Our main contributions are summarized as follows:

- We propose LBANet, a lightweight COD framework that formulates boundary guidance as prediction-conditioned boundary-region activation. By combining structural evidence with current segmentation state, LBANet identifies spatial regions for selective feature refinement.
- We design WEPM and SBAM to integrate wavelet-based structural cues, high-level semantics, prediction uncertainty, and mask gradients into a learnable binary boundary-region gate, connecting local boundary evidence with region-level feature modulation.
- We introduce DRM to coordinate mask and edge refinement for coherent object segmentation. Experiments on four COD benchmarks demonstrate a complexity and accuracy trade-off with 2.05M parameters and 2.66G FLOPs.

2 Related Work

2.1 Lightweight Camouflaged Object Detection

Lightweight camouflaged object detection (COD) aims to preserve discriminative object representations under constrained model capacity and computational budgets. Early efficient designs include TinyCOD [21], which combines adjacent-scale feature fusion with edge-area focusing, and FINet [22], which injects high- and low-frequency cues into compact backbone features. Recent studies have extended this direction through more structured feature interactions. CSFIN [28] enhances cross-stage representations using bidirectional feature interaction, while LiteCOD [29] integrates local and global information through efficient multi-scale fusion. ULCOD-Net [30] combines gated multi-level fusion with boundary–region constraints and progressive feature refinement. FMLNet [23] employs bidirectional frequency–RGB interaction for compact cross-domain representation learning. Complementary efficiency strategies include dynamic sparse attention and multi-level knowledge distillation in DS-PRNet [31]. These methods establish compact feature representation, structured interaction, and model compression as complementary routes to efficient COD.

2.2 Bio-inspired COD and Refinement

Bio-inspired COD methods draw on predator behavior and human visual perception to facilitate object discovery and recognition. SINet [32] mimics the search-and-identification process of predators, while SINet-V2 [33] enhances this paradigm through group-reversal attention. ZoomNet [34] imitates human zooming behavior to integrate multi-scale representations. PreyNet [35] further models the sensory and cognitive stages of predation, incorporating uncertainty-guided feature calibration to enhance ambiguous regions. Beyond object localization, recent approaches explore biologically motivated structural perception and progressive refinement. FLRNet [36] adopts a filtering–localization–refinement strategy to progressively recover object details. BPNet [37] draws inspiration from visual cortical cells to enhance boundary perception, while PONet [38] coordinates central-region semantics and peripheral high-frequency details through partitioned visual observation. These studies demonstrate the value of integrating local structural cues with contextual information for region-level perception and refinement. Following this perspective, LBANet organizes structural evidence into a point-to-line-to-region-to-object progression, converting boundary and prediction cues into learnable binary spatial activation for targeted feature refinement.

2.3 Prior-Guided Camouflaged Object Detection

Boundary priors. Boundary-aware COD methods incorporate structural evidence through edge supervision, feature guidance, and joint region–edge learning. BGNet [19] introduces boundary guidance for camouflaged object representation, while ERRNet [12] employs edge-based reversible recalibration. EAMNet [39] formulates edge detection and mask segmentation as mutually guided refinement tasks. BCLNet [40] further coordinates boundary and object features through collaborative learning and boundary-guided selective fusion. GBNet [20] introduces a boundary gate to filter background interference and a boundary-aware decoder to inject structural features into segmentation. These approaches establish boundary guidance, selective filtering, and mask–edge interaction as effective mechanisms for structural recovery.

Frequency priors. Frequency-aware methods capture complementary structural information through spectral decomposition and cross-domain feature learning. Frequency Perception Network [14] integrates frequency cues into COD representations, while FEDER [41] combines feature decomposition with edge reconstruction. FINet [22] efficiently injects frequency components into lightweight features, and FMLNet [23] models bidirectional interactions between frequency and RGB representations. LaFINet [42] explores Laplacian-pyramid frequency injection for boundary-sensitive detection. Building on these prior-guided directions, LBANet integrates Haar-derived structural responses with semantic boundary evidence, prediction uncertainty, and mask gradients to construct a binary boundary-region gate. The resulting activation modulates low-level features for subsequent shared aggregation and joint mask–edge refinement.

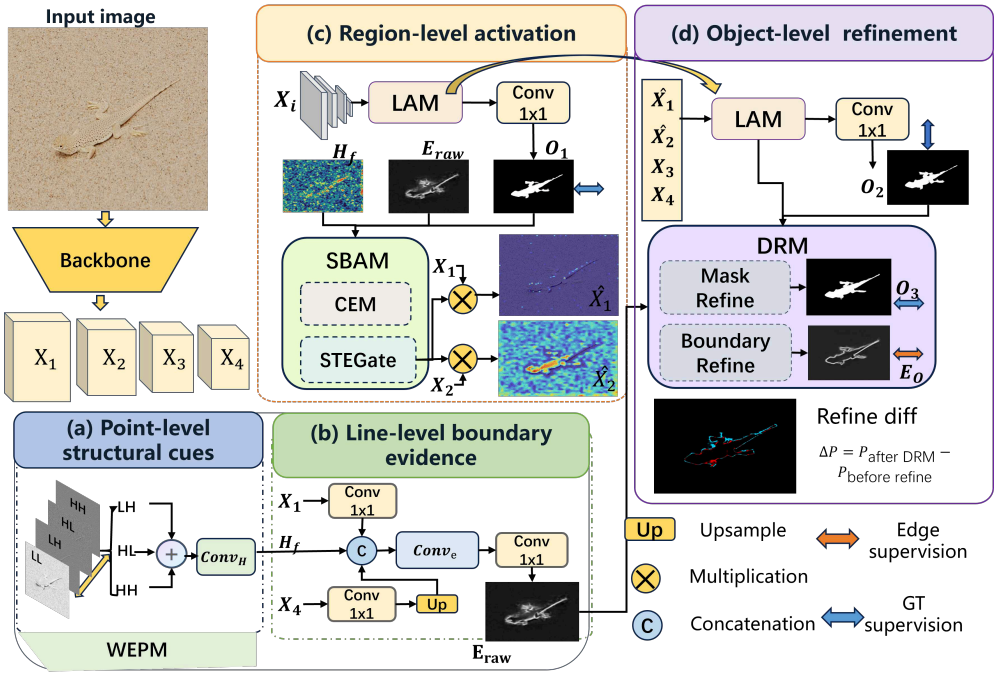

Fig. 2 Overall architecture of LBANet.

3 Method

3.1 Overview

The central idea of LBANet is to transform boundary-related structural evidence and coarse prediction states into binary spatial activation decisions, thereby identifying where additional feature refinement should be applied. Given an image $\mathbf{I} \in \mathbb{R}^{3\times H\times W}$, lightweight MobileViT backbone [43] extracts four hierarchical feature maps $\{X_i\}_{i=1}^{4}$, where shallow features preserve local structural details and deep features provide semantic localization. As shown in Fig. 2, the wavelet edge prior module (WEPM) combines shallow high-frequency responses with deep semantic guidance to generate a raw edge prior E_{raw}. In parallel, $\{X_i\}_{i=1}^{4}$ pass through the lightweight aggregation module (LAM), shown in Fig 3, produces a coarse object prediction O_1. Spike-inspired boundary activation module (SBAM) then integrates structural evidence with the uncertainty and probability gradients derived from O_1, producing a binary boundary-region gate $\hat{G}$. This gate selectively modulates structure-related features to produce $\hat{X}_1$ and $\hat{X}_2$, which are then re-aggregated with X_3 and X_4 by the shared LAM to obtain an updated object prediction O_2. Dual refinement module (DRM) performs boundary-guided mask refinement and mask-informed edge refinement, yielding the final segmentation O_3 and an auxiliary edge prediction E_o.

3.2 Wavelet Edge Prior Module

WEPM constructs compact boundary-related structural evidence from complementary shallow and deep features. Shallow features provide local texture and structural details, while deep features supply semantic information for object localization. By combining these representations, WEPM in Fig. 2, transforms local high-frequency responses into a semantic-guided boundary prior for subsequent region activation.

Shallow feature X_1 is decomposed by Haar discrete wavelet transform (DWT), yielding four subbands$X_1^{LL}, X_1^{LH}, X_1^{HL}, X_1^{HH}$. Three high-frequency subbands are then aggregated and projected as follows:

$$H_f = Conv_h\left(Up(X_1^{LH} + X_1^{HL} + X_1^{HH})\right), \tag{1}$$

where $Up(\cdot)$ denotes upsampling and $Conv_h(\cdot)$ consists of depthwise-separable convolution blocks followed by a 1×1 projection. Then, H_f is combined with projected shallow features and deep semantic features to predict the raw edge prior:

$$E_{\text{raw}} = Conv_e\left(Cat[Conv^1(X_1), Up(Conv^4(X_4)), H_f]\right), \tag{2}$$

where $Cat[\cdot]$ denotes channel-wise concatenation, $Conv^1(\cdot)$ and $Conv^4(\cdot)$ are 1×1 convolution layers. $Conv_e(\cdot)$ consists of two 3×3 depthwise-separable convolution with batch normalization and ReLU activation function, and a 1×1 convolution layer.

3.3 Spike-inspired Boundary Activation Module

SBAM converts complementary structural evidence and coarse prediction states into binary spatial activation decisions. Inspired by threshold-triggered discrete activation in spiking neural networks [44, 45], SBAM selectively modulates low-level features at boundary-related locations. As shown in Fig. 3, it consists of a Confidence Estimation Module (CEM) and a spike-inspired straight-through estimator gate (STEGate). CEM estimates the spatial activation response, while STEGate converts it into a hard binary decision for feature refinement.

3.3.1 Confidence Estimation Module

CEM estimates a spatial activation descriptor by jointly modeling boundary-related structural evidence and current segmentation state. Given raw edge prior E_{raw}, we generate edge attention map P_e by sigmoid activation function:

$$P_e = \sigma\left(Up(E_{\text{raw}})\right). \tag{3}$$

where $\sigma(\cdot)$ denotes sigmoid function. Coarse prediction O_1 is first converted into a probability map, from which the uncertainty and probability-gradient maps are computed:

$$\begin{cases} P = \sigma\left(\text{Up}(O_1)\right), \\ \mathcal{U} = 1 - |2P - 1|, \\ \mathcal{B} = |\Delta_x P| + |\Delta_y P| \end{cases} \tag{4}$$

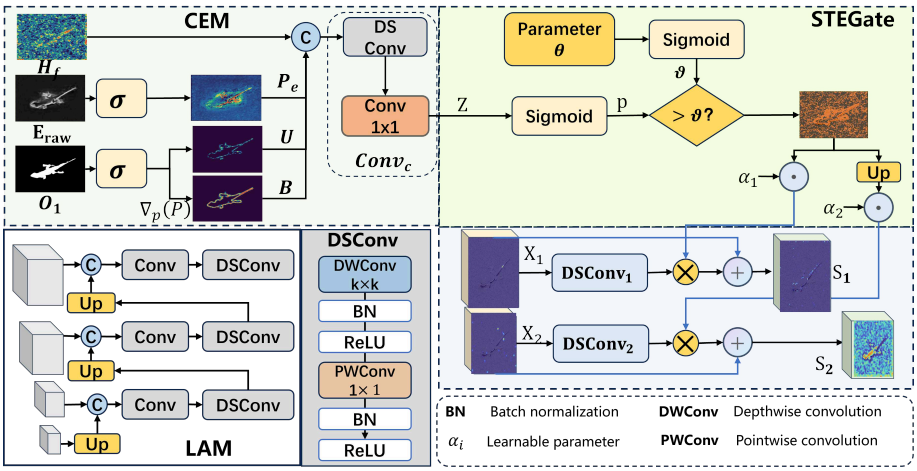

Fig. 3 Architecture of Spike-inspired Boundary Activation Module.

Δ_x and Δ_y denote horizontal and vertical finite differences, respectively. Predictions with probabilities close to 0.5 exhibit higher uncertainty.

CEM then integrates four complementary cues that are concatenated and projected into a boundary confidence descriptor:

$$\mathcal{Z} = Conv_c\left(\mathrm{Cat}[H_f, P_e, \mathcal{U}, \mathcal{B}]\right), \tag{5}$$

where $\mathrm{Cat}[\cdot]$ denotes channel-wise concatenation and $Conv_c(\cdot)$ is implemented by convolutional blocks with 3×3 and 1×1. Specifically, H_f captures local structural variations, P_e supplies edge evidence, $\mathcal{U}$ identifies uncertain predictions, and $\mathcal{B}$ captures mask-level transitions. Their joint representation provides a prediction-conditioned spatial selection signal for boundary-region activation.

3.3.2 Spike-inspired straight-through estimator gate

STEGate converts the continuous spatial activation logits into a hard binary gate. Given the activation probability $p = \sigma(\mathcal{Z})$, the forward activation decision is defined as:

$$G = \mathbb{I}[p > \vartheta], \tag{6}$$

where ϑ denotes the activation threshold and $\mathbb{I}[\cdot]$ is the indicator function. To enable gradient-based optimization of confidence estimator, we employ a straight-through estimator(STE) [46]:

$$\hat{G} = \mathrm{sg}(G) - \mathrm{sg}(p) + p, \tag{7}$$

where $\mathrm{sg}(\cdot)$ denotes the stop-gradient operation. The forward output remains binary, $\hat{G} = G$, while the backward pass uses the sigmoid probability as a surrogate. Accordingly, the gradient with respect to $\mathcal{Z}$ is approximated by

$$\frac{\partial \mathcal{L}}{\partial \mathcal{Z}} \approx \frac{\partial \mathcal{L}}{\partial \hat{G}} p(1-p). \tag{8}$$

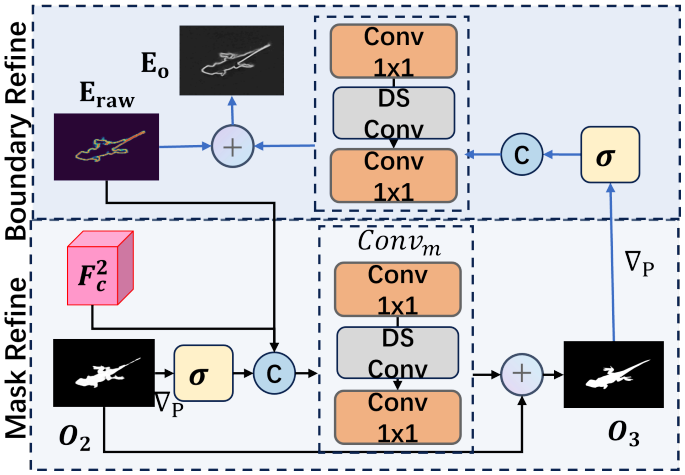

Fig. 4 Architecture of the Dual Refinement Module.

Finally, the gate is applied to low-level features through residual modulation:

$$\hat{X_1} = X_1 + \alpha_1 \cdot DSConv^1(X_1) \odot \hat{G}, \tag{9}$$

$$\hat{X_2} = X_2 + \alpha_2 \cdot DSConv^2(X_2) \odot \mathcal{U}(\hat{G}), \tag{10}$$

where $DSConv^1(\cdot)$ and $DSConv^2(\cdot)$ are convolutional block, $\odot$ denotes element-wise multiplication, and α_1 and α_2 are learnable scaling factors. Through this residual formulation, SBAM applies additional feature modulation at selected spatial locations while preserving the original feature representations.

3.4 Dual Refinement Module

Following boundary-region activation and shared LAM, DRM further improves object recovery by coupling boundary-guided mask refinement with mask-informed edge refinement. As illustrated in Fig. 4, DRM consists of two sequential residual refinement steps.

For mask refinement, given the re-aggregated feature F_c^2 and the updated coarse prediction O_2, DRM first incorporates the raw edge prior and the probability gradient of O_2 to refine the object mask:

$$O_3 = O_2 + \mathrm{Conv_m}\left(\mathrm{Cat}\left[F_c^2, \sigma(E_{\mathrm{raw}}), \nabla_p(\sigma(O_2))\right]\right), \tag{11}$$

where $\nabla_p(\cdot)$ denotes the probability-gradient operator, and $\mathrm{Conv_m}(\cdot)$ consisting of two 1×1 convolution and a depthwise separable convolution block. The residual formulation preserves the coarse region prediction while using edge-aware cues to refine weak contours and local details. The final object prediction is obtained from O_3.

For edge refinement, DRM updates the raw edge prior using both the boundary activation gate and the boundary cue from the refined mask. The refined edge

prediction is computed as

$$E_o = E_{\text{raw}} + \text{Conv}_{\text{o}}\left(\text{Cat}\left[F_c^2, \nabla_p(\sigma(O_3))\right]\right), \tag{12}$$

here, $\text{Conv}_{\text{o}}(\cdot)$ is an independent edge-refinement head with the same lightweight structure as $\text{Conv}_{\text{m}}(\cdot)$. DRM uses boundary-related evidence to refine the object region and then exploits the updated mask structure to improve the auxiliary edge representation.

3.5 Loss Function

Following the training strategy of FEDER [41], LBANet is trained with mask supervision, edge supervision, and uncertainty-aware regularization. Mask predictions $\{O_k\}_{k=1}^3$, and refined edge prediction E_o are supervised by segmentation ground truth (GT_s) and edge ground truth (GT_e), respectively. During training, all prediction maps are upsampled to the same spatial resolution as their corresponding ground truths. For mask supervision, we adopt the weighted binary cross-entropy loss $\mathcal{L}_{\text{BCE}}^w$ and the weighted IoU loss $\mathcal{L}_{\text{IoU}}^w$, which are commonly used in COD to emphasize hard pixels and object-region overlap. For edge supervision, we use the Dice loss $\mathcal{L}_{\text{dice}}$ [47]. In addition, uncertainty-aware loss $\mathcal{L}_{\text{ual}}$ is applied to the final mask prediction O_3 to suppress ambiguous responses. The overall loss function is formulated as follows:

$$\begin{aligned}\mathcal{L} = &\sum_{k=1}^{3} 2^{k-3}\left(\mathcal{L}_{\text{BCE}}^w(O_k, GT_s) + \mathcal{L}_{\text{IoU}}^w(O_k, GT_s)\right) \\ &+ \mathcal{L}_{\text{dice}}(E_o, GT_e) + \lambda\,\mathcal{L}_{\text{ual}}(O_3),\end{aligned} \tag{13}$$

where $\mathcal{L}_{\text{ual}}$ denotes the uncertainty-aware regularization applied to the final mask prediction, and λ is set to 2 in our implementation.

4 Experiments

4.1 Experimental setup

4.1.1 Datasets

We evaluated LBANet on four public COD benchmarks: CHAMELEON [48], CAMO [49], COD10K [32] and NC4K [50]. Following common setting, model was trained on CAMO and COD10K training dataset and tested on all four benchmark test sets. CHAMELEON [48] contains 76 camouflaged animal images collected from the Internet and is commonly used to evaluate generalization in natural camouflage scenes. CAMO [49] contains 1,250 camouflaged images with challenging factors such as complex backgrounds, occlusion, small objects, and irregular object shapes, among which 1,000 images are used for training and 250 images for testing. COD10K [32] is a large-scale COD dataset with 3,040 training images and 2,026 testing images covering diverse camouflaged object categories. NC4K [50] contains 4,121 high-resolution camouflaged images with complex scenes and large scale variations.

4.1.2 Experiment details

Experiments were implemented in PyTorch on a single NVIDIA GeForce RTX 4060 Laptop GPU with 8 GB memory. MobileViT-XS [43] was used as the lightweight encoder and the decoder width was set to 64 channels. Images, masks and edge maps were resized to 384×384. The network was trained for 100 epochs using Adam with an initial learning rate of 2×10^{-4} and a batch size of 20. Automatic mixed precision and gradient clipping with a threshold of 0.5 were used. The learning-rate schedule combined step decay and cosine annealing with a minimum learning rate of 1×10^{-5}.

4.1.3 Evaluation Metrics

We adopt four evaluation metrics: S-measure (S_α), weighted F-measure (F_β^w), mean absolute error ($\mathcal{M}$), adaptive E-measure (E_ϕ^{ad}). S_α evaluates object-level and region-level structural similarity between prediction maps and ground-truth masks. F_β^w jointly considers precision and recall while assigning different weights to spatial errors. $\mathcal{M}$ measures the pixel-wise absolute difference between prediction maps and ground truths, where lower values indicate better performance. E_ϕ^{ad} evaluates both local pixel matching and global structural consistency. We further report model parameters, FLOPs and FPS following [21].

4.2 Comparison with state-of-the-art methods

We compared LBANet with 22 representative COD methods, including 18 general COD models SINet [32], R-MGL [51], LSR [50], JCOD [52], UGTR [53], C2FNet [9], PFNet [54], SINetV2 [33], FindNet [55], ERRNet [12], C2FNetV2 [56], BGNet [19], FEDER [41], DTC-Net [57], R-MGLV2 [58], MRRNet [59], DINet [60], EPFDNet [61], TinyCOD [21], FINet [22], FMLNet [23], and lightweight methodsTinyCOD [21], FINet [22], FMLNet [23]and ULCOD-Net [30]. Following [62], we just regarded models with fewer than 10M parameters as lightweight methods. For fair comparison, we used results reported in the original papers when available. When released prediction maps were provided, we evaluated them under the same protocol. Otherwise, we reproduced the methods using their official implementations and recommended settings.

4.2.1 Quantitative Comparison

Table 1 presents the quantitative results on four COD benchmarks. Among lightweight methods, LBANet achieves 13 the best results and 3 second-best performance across all 16 evaluation entries while using the fewest parameters. Compared with FML-Net [23], LBANet reduces the parameter count by 26.0%. On CAMO, it improves S_α, E_ϕ^{ad}, and F_β^w by 1.22%, 0.45%, and 1.85%, respectively, and reduces MAE by 8.96%. These results demonstrate that LBANet maintains strong segmentation accuracy within a compact parameter budget.

Compared with general COD methods, LBANet achieves competitive or superior performance with substantially fewer parameters. For instance, LBANet reduces the parameter count by 91.8% compared with SINetV2 [33], while improving S_α, E_ϕ^{ad}, and F_β^w on CAMO by 1.46%, 0.90%, and 3.50%, respectively, while reducing

Table 1 Quantitative comparison with representative COD methods. Each dataset reports S_α, E_ϕ^{ad}, F_β^w, and M. ↑ and ↓ indicate that higher and lower values correspond to better performance, respectively. For lightweight methods, the best and second-best results are shown in bold and underlined, respectively.

Method	Pub.	Param. (M)	CHAMELEON (76)				CAMO (250)				COD10K (2,026)				NC4K (4,121)			
			$S_\alpha\uparrow$	$E_\phi^{ad}\uparrow$	$F_\beta^w\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$E_\phi^{ad}\uparrow$	$F_\beta^w\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$E_\phi^{ad}\uparrow$	$F_\beta^w\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$E_\phi^{ad}\uparrow$	$F_\beta^w\uparrow$	$M\downarrow$
General COD methods (Param. > 10M)																		
SINet [32]	CVPR'20	48.95	0.869	0.899	0.740	0.044	0.751	0.834	0.606	0.100	0.776	0.867	0.631	0.043	0.808	0.883	0.723	0.058
R-MGL [51]	CVPR'21	73.70	0.893	0.923	0.812	0.031	0.776	0.848	0.673	0.088	0.814	0.864	0.666	0.035	0.833	0.884	0.741	0.052
LSR [50]	CVPR'21	28.70	0.890	0.936	0.822	0.030	0.787	0.859	0.696	0.080	0.804	0.883	0.673	0.037	0.840	0.904	0.766	0.048
JCOD [52]	CVPR'21	55.90	0.891	0.943	0.833	0.030	0.800	0.872	0.728	0.073	0.809	0.882	0.684	0.035	0.842	0.906	0.771	0.047
UGTR [53]	ICCV'21	48.87	0.888	0.920	0.796	0.031	0.785	0.861	0.686	0.086	0.818	0.850	0.667	0.035	0.839	0.890	0.747	0.052
C2FNet [9]	IJCAI'21	28.41	0.888	0.932	0.828	0.032	0.796	0.865	0.719	0.080	0.813	0.886	0.686	0.036	0.838	0.901	0.762	0.049
PFNet [54]	CVPR'21	46.50	0.882	0.942	0.810	0.033	0.800	0.855	0.695	0.085	0.800	0.868	0.660	0.040	0.829	0.894	0.745	0.053
SINetV2 [33]	TPAMI'22	24.90	0.888	0.930	0.816	0.030	0.820	0.884	0.743	0.070	0.815	0.864	0.680	0.037	0.847	0.901	0.770	0.048
FindNet [55]	TIP'22	32.44	0.895	0.944	0.839	0.027	0.800	0.862	0.725	0.077	0.811	0.883	0.688	0.036	0.841	0.895	0.769	0.048
ERRNet [12]	PR'22	67.70	0.917	0.875	0.787	0.039	0.779	0.855	0.679	0.085	0.786	0.845	0.630	0.043	0.827	0.892	0.737	0.054
C2FNetV2 [56]	TCSVT'22	25.20	0.893	0.947	0.845	0.028	0.799	0.869	0.730	0.077	0.811	0.890	0.691	0.036	0.840	0.903	0.772	0.048
BGNet [19]	IJCAI'22	79.85	0.901	0.938	0.850	0.027	0.812	0.870	0.749	0.073	0.831	0.901	0.722	0.033	0.851	0.907	0.788	0.044
FEDER [41]	CVPR'23	44.13	0.894	0.942	0.834	0.028	0.807	0.873	0.785	0.069	0.823	0.900	0.740	0.032	0.846	0.913	0.824	0.042
DTC-Net [57]	TMM'23	53.50	0.876	0.897	0.773	0.039	0.778	0.804	0.667	0.084	0.790	0.821	0.616	0.041	–	–	–	–
R-MGLV2 [58]	TIP'23	74.02	0.892	0.935	0.825	0.029	0.774	0.848	0.684	0.085	0.813	0.882	0.682	0.034	0.831	0.892	0.751	0.051
MRRNet [59]	TNNLS'24	71.82	0.882	0.931	0.799	0.033	0.812	0.869	0.772	0.076	0.822	0.889	0.730	0.036	0.848	0.898	0.801	0.048
DINet [60]	TMM'24	30.84	–	–	–	–	0.821	0.883	0.790	0.068	0.832	0.901	0.744	0.031	0.856	0.910	0.820	0.043
EPFDNet [61]	TCSVT'25	27.50	–	–	–	–	0.817	0.887	0.757	0.068	0.815	0.894	0.700	0.033	0.844	0.906	0.780	0.044
Lightweight COD methods (Param. ≤ 10M)																		
TinyCOD [21]	ICASSP'23	4.72	0.887	0.931	0.814	0.030	0.822	0.890	0.752	0.066	0.811	0.877	0.678	0.036	0.843	0.903	0.766	0.047
FINet [22]	SPL'24	3.74	0.883	0.929	0.808	0.031	0.828	0.890	0.752	0.065	0.817	0.883	0.686	0.034	0.847	0.904	0.771	0.047
FMLNet [23]	ASOC'26	2.77	0.883	0.935	0.820	0.032	0.822	0.888	0.755	0.067	0.823	0.894	0.704	0.033	0.850	0.909	0.780	0.045
ULCOD-Net [30]	ASOC'26	2.51	**0.893**	0.938	**0.837**	0.030	0.823	0.890	0.758	0.067	**0.829**	0.893	0.714	0.033	**0.854**	0.908	0.787	0.045
LBANet	Ours	**2.05**	0.889	**0.941**	0.831	**0.029**	**0.832**	**0.892**	**0.769**	**0.061**	0.828	**0.897**	**0.717**	**0.032**	**0.854**	**0.913**	**0.791**	**0.043**

MAE by 12.86%. Compared with the high-capacity MRRNet [59], LBANet uses only 2.85% of its parameters and achieves competitive results on CAMO, COD10K, and NC4K. These comparisons highlight the ability of the compact architecture to preserve discriminative object representations.

Table 2 summarizes the model complexity and inference efficiency of representative COD methods. LBANet contains only 2.05M parameters and requires 2.66G FLOPs at an input resolution of 384×384. Compared with DINet [60], LBANet reduces the parameter count from 30.84M to 2.05M and FLOPs from 112.24G to 2.66G, achieving a substantial reduction in model complexity. Relative to FMLNet [23], LBANet reduces FLOPs by 11.3% and improves FPS by about 32.8%. As shown in Fig. 5, these results demonstrate a favorable balance between model compactness, computational complexity, and segmentation accuracy.

4.3 Qualitative Analysis

Fig. 6 presents qualitative comparisons on challenging COD10K samples, including low-contrast objects, cluttered backgrounds, weak boundaries, and irregular target shapes. In highly camouflaged scenes, several baselines miss parts of the objects or activate background textures, whereas LBANet preserves more coherent foreground regions. In weak-boundary cases, LBANet maintains more accurate object shapes and reduces spurious foreground expansion. For objects with complex structures, LBANet preserves separated regions and fine shape details, contributing to more complete object predictions. In cluttered backgrounds, its predictions suppress distracting responses while maintaining target localization. For large objects, LBANet also recovers more complete foreground regions and extended structures in the illustrated

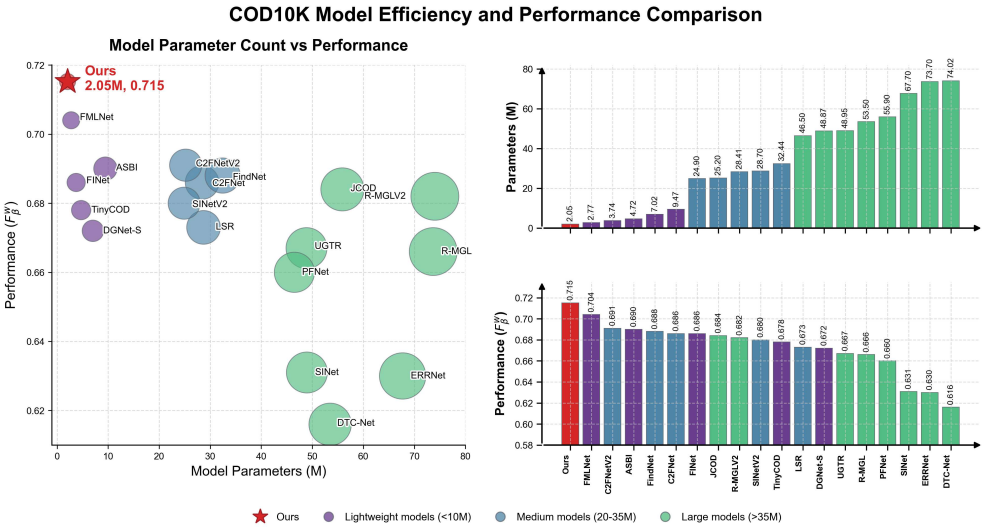

Fig. 5 COD methods on COD10K Evaluated by F_β^w and parameters

Table 2 Efficiency comparison with representative COD methods.

Method	Pub.	Param. (M)	FLOPs (G)	FPS
SINet [32]	CVPR'20	48.95	38.84	42
R-MGL [51]	CVPR'21	73.70	378.60	20
LSR [50]	CVPR'21	28.70	17.50	72
JCOD [52]	CVPR'21	55.90	30.30	32
UGTR [53]	ICCV'21	48.87	121.70	20
C2FNet [9]	IJCAI'21	28.41	26.24	40
PFNet [54]	CVPR'21	46.50	26.60	51
ERRNet [12]	PR'22	67.70	20.10	80
SINetV2 [33]	TPAMI'22	24.90	24.56	66
C2FNetV2 [56]	TCSVT'22	25.20	18.20	41
MRRNet [59]	TNNLS'24	71.82	17.53	80
DINet [60]	TMM'24	30.84	112.24	35
TinyCOD [21]	ICASSP'23	4.72	**1.40**	52
FINet [22]	SPL'24	3.74	**1.25**	66
FMLNet [23]	ASOC'26	2.77	3.00	58
ULCOD-Net [30]	ASOC'26	2.51	3.14	96
LBANet	Ours	**2.05**	2.66	77

examples. These observations demonstrate the effectiveness of LBANet in preserving object integrity and structural details under challenging camouflage conditions.

4.4 Ablation study

We conduct systematic ablation studies on four benchmark datasets to evaluate the contribution of each component in LBANet. All variants are trained with the same

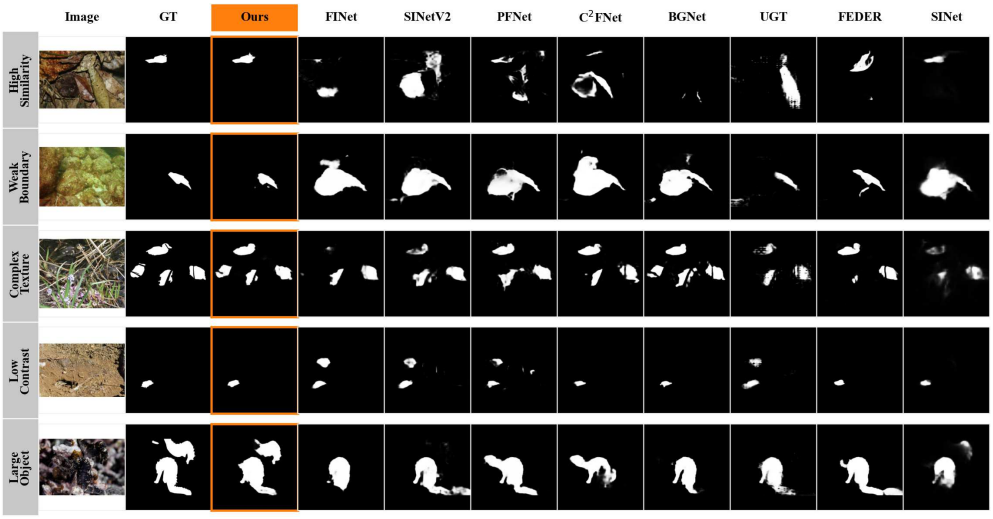

Fig. 6 Qualitative comparison on representative challenging samples.

Table 3 Ablation study on CHAMELEON, CAMO, COD10K, and NC4K. Best results are shown in bold.

No.	Method	Param. (M)	CHAMELEON (76)				CAMO (250)				COD10K (2,026)				NC4K (4,121)			
			$S_\alpha \uparrow$	$E_\phi^{ad} \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi^{ad} \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi^{ad} \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\phi^{ad} \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$
1	B	2.000	0.864	0.892	0.749	0.045	0.817	0.869	0.706	0.077	0.798	0.819	0.623	0.045	0.837	0.875	0.725	0.056
2	B+W	2.023	0.871	0.928	0.800	0.033	0.825	0.885	0.754	0.066	0.821	0.883	0.697	0.035	0.851	0.906	0.779	0.045
3	B+W+2L	2.023	0.874	0.928	0.802	0.034	0.822	0.891	0.752	0.067	0.822	0.883	0.699	0.034	0.850	0.904	0.777	0.046
4	B+W+2L+S	2.039	0.882	0.935	0.819	0.031	0.828	0.896	0.765	**0.059**	0.824	0.889	0.706	0.033	0.851	0.908	0.783	0.045
5	Only Mask	2.045	**0.889**	0.932	0.830	0.030	0.825	0.890	0.765	0.064	0.824	0.892	0.710	0.033	0.852	0.910	0.787	0.044
6	Only Edge	2.045	0.879	0.926	0.817	0.031	0.824	0.890	0.759	0.064	0.823	0.891	0.707	0.034	0.851	0.909	0.785	0.045
7	WEPM w/o H_f	2.045	0.886	0.934	0.827	0.032	0.827	**0.897**	**0.769**	0.062	0.825	0.893	0.713	0.033	0.851	0.910	0.787	0.045
8	LBANet	2.050	**0.889**	**0.941**	**0.831**	**0.029**	**0.832**	0.892	**0.769**	0.061	**0.828**	**0.897**	**0.717**	**0.032**	**0.854**	**0.913**	**0.791**	**0.043**

settings and input resolution for fair comparison. Baseline model (B) only adopts backbone and a LAM followed by a 1×1 convolution to generate the prediction. Starting from this baseline, we progressively add WEPM (B+W), shared LAM (B+W+2L), SBAM (B+W+2L+S), then we also evaluate variants with only mask refinement (only mask), only edge refinement (only edge) to analyze the bidirectional design of DRM in Table 3. Relative to the baseline, full LBANet increases mean F_β^w by 10.88% and reduces mean MAE by 26.01%, with parameters increasing from 2.00M to 2.05M. Fig. 7 presents the visualization of the ablation study.

4.4.1 Effectiveness of WEPM

As shown in Table 3, adding WEPM (No.2) increased the average S_α, E_ϕ^{ad} and F_β^w by 1.57%, 4.25% and 8.10%, respectively, and reduced average MAE by 19.73%. The qualitative results in Fig. 7 also shows that WEPM reduces unstable responses around weak boundaries. No.7 in Table 3 shows that removing Haar high-frequency information from the full model reduced the average S_α by approximately 0.41%, and

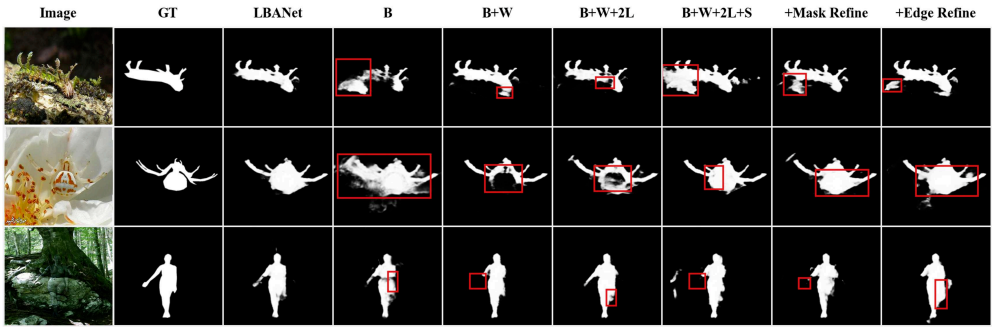

Fig. 7 Visualization for LBANet ablation Study.

increased MAE by approximately 4.24%. These results demonstrate the effectiveness of incorporating wavelet-based structural cues into semantic boundary modeling.

4.4.2 Effectiveness of SBAM

To distinguish SBAM's contribution from repeated feature aggregation, we compare variants with and without a second shared LAM pass. No.3 vs. No.2 in Table 3shows marginal changes in four metrics. This result suggests that the performance gain is not simply caused by repeating the decoder aggregation. In contrast, adding SBAM under the same aggregation structure (No.4 vs. No.3) improves the average S_α, E_ϕ^{ad}, and F_β^w by 0.50%, 0.61% and 1.42%, respectively, while reducing MAE by 7.18%. This comparison demonstrates that boundary-region activation provides additional benefits beyond repeated feature aggregation.

We further examine the contribution of individual cues in CEM, including the high-frequency feature H_f, edge prior E_{raw}, uncertainty map, and mask-gradient map. As reported in Table 4, removing each cue decreases the performance on average. In particular, removing H_f decreases the average F_β^w decreasing by 1.54% and MAE increasing by 6.67%, while removing the mask-gradient cue decreases F_β^w by 1.22% and increases MAE by 6.67%. These results highlight the complementary contributions of local structural responses and coarse-mask transitions.

Finally, we compare STEGate with a continuous Sigmoid Gate to assess the effectiveness of binary spatial activation. As shown in Table 4, STEGate improves the average S_α, F_β^w by approximately 0.53%, 0.68%, and reduces the average MAE by 2.37%. Figure 8 visualizes the corresponding activation patterns. STEGate generates hard binary responses and activates spatial regions extending beyond thin boundary contours. The statistics in Appendix Table A1 show that STEGate produced a larger activated region ratio (38.61% vs. 8.90%) and a stronger mean response within the ground-truth boundary band (0.841 vs. 0.700).

4.4.3 Effectiveness of DRM

DRM couples mask and edge refinement through a bidirectional interaction mechanism. Specifically, the edge prior guides mask refinement, while the boundary information extracted from the refined mask is used to update edge predictions. To

Table 4 Ablation results of SBAM cues and gate mechanisms. Best results are highlighted in bold.

No.	Method	Param (M)	CHAMELEON (76)				CAMO (250)				COD10K (2,026)				NC4K (4,121)			
			$S_\alpha\uparrow$	$E_\phi^{ad}\uparrow$	$F_\beta^w\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$E_\phi^{ad}\uparrow$	$F_\beta^w\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$E_\phi^{ad}\uparrow$	$F_\beta^w\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$E_\phi^{ad}\uparrow$	$F_\beta^w\uparrow$	$M\downarrow$
1	w/o HF	2.048	0.875	0.929	0.810	0.034	0.828	**0.896**	0.764	0.062	0.821	0.885	0.702	0.035	0.851	0.907	0.784	0.045
2	w/o Edge prior	2.050	0.882	0.938	0.826	0.031	0.822	0.890	0.756	0.065	0.827	**0.898**	0.715	0.033	0.853	0.912	0.789	0.044
3	w/o Uncertainty	2.050	**0.894**	**0.941**	**0.839**	**0.029**	0.821	0.889	0.758	0.064	0.827	0.894	0.715	0.033	0.852	0.910	0.787	0.044
4	w/o Mask gradient	2.050	0.877	0.926	0.816	0.032	0.822	0.890	0.760	0.065	0.823	0.890	0.709	0.034	0.850	0.910	0.785	0.045
5	Sigmoid gate	2.050	0.882	0.935	0.825	0.030	0.824	0.889	0.758	0.062	0.826	0.897	0.715	0.033	0.853	0.912	0.789	0.044
6	LBANet	2.050	0.889	**0.941**	0.831	**0.029**	**0.832**	0.892	**0.769**	**0.061**	**0.828**	0.897	**0.717**	**0.032**	**0.854**	**0.913**	**0.791**	**0.043**

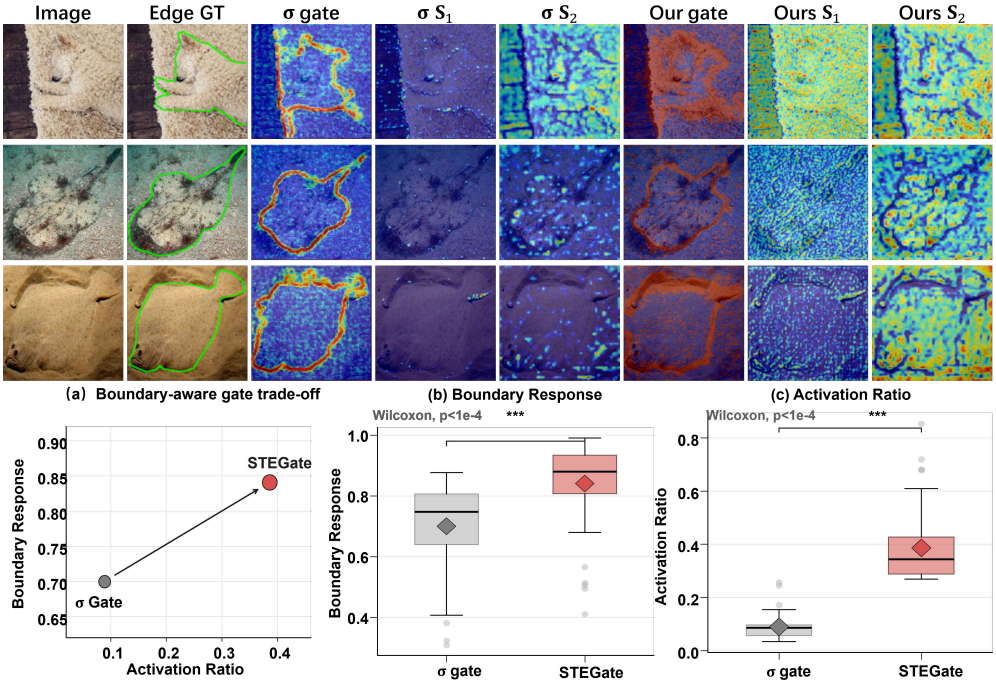

Fig. 8 Visualization of the different gate mechanisms in SBAM ablation variants. The bottom row shows the boundary response and boundary-region activation ratio.

evaluate the contribution of each refinement direction, we compare mask-only, edge-only, and bidirectional refinement variants in Table 3. Compared the variant without DRM (No.4) to LBANet, it decreases the average S_α, E_ϕ^{ad}, and F_β^w by 0.53%, 0.41%, and 1.13%, respectively. Compared with the complete LBANet, mask-only refinement decreases E_ϕ^{ad}, F_β^w by 0.52%, 0.51%, with a 3.64% increase in MAE. Similarly, edge-only refinement decreases S_α, E_ϕ^{ad}, and F_β^w by 0.76%, 0.74%, and 1.29%, while increasing MAE by 5.45%.

4.5 Failure analysis

Although LBANet improves performance on the evaluated benchmarks, several difficult cases remain unresolved, as shown in Fig. 9. When object-background contrast is extremely low or object boundaries are weak and ambiguous, the available structural evidence may be insufficient, resulting in under-segmentation or incomplete masks.

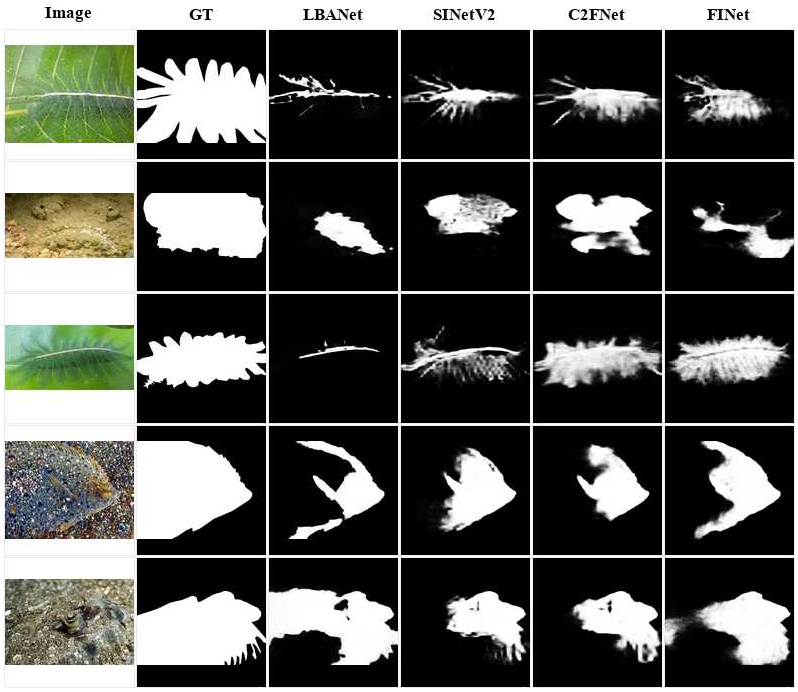

Fig. 9 Failure case illustration on challenging COD samples. The selected examples cover fragmented thin structures, low-contrast incomplete masks, weak-edge details, ambiguous boundaries, and texture-clutter-induced over-segmentation.

Thin elongated structures are also prone to fragmented predictions, since their boundary cues are discontinuous and occupy only a small pixel ratio. In cluttered scenes, background textures with object-like high-frequency responses may still trigger false positives and over-segmentation. These cases suggest that LBANet remains dependent on discernible boundary and texture cues in RGB images. Improving robustness to extreme camouflage, ambiguous contours, and fine-structure recovery is therefore an important direction for future lightweight COD models.

5 Conclusion

We presented LBANet, a lightweight COD framework that moves from thin-edge refinement to boundary-region activation. Experiments on four COD benchmarks showed competitive detection accuracy with 2.05M parameters, and ablations supported the roles of WEPM, SBAM and DRM. The results also suggest that lightweight COD benefits from treating boundary refinement as a selection problem. WEPM supplies compact structural evidence, SBAM converts multiple cues into a binary boundary-region decision, and DRM translates this decision into consistent mask and edge outputs. This interpretation also defines the boundary of the claim. LBANet uses spike-inspired binary activation, but it is not a temporal SNN and does not claim membrane dynamics, event-camera input or neuromorphic acceleration. STEGate

is optimized for segmentation quality and boundary-region selection. The efficiency of the model mainly comes from the lightweight encoder, shared aggregation and compact refinement design. WEPM uses fixed Haar decomposition, and adaptive frequency bases improve robustness in extremely cluttered textures.

Acknowledgements. This work was supported by the Key Project of "Collaborative Quality Improvement Plan for Teacher Education" of the Chinese Society of Education (Grant No. 202400002207ZXB).

Appendix A Quantitative analysis of gate activation maps

Table A1 Quantitative analysis of gate activation maps. Activation ratio denotes the proportion of activated pixels in the gate map. Boundary response denotes the mean gate value within the ground-truth boundary band. Gate standard deviation measures the sharpness of the gate distribution.

Gate	Activation ratio	Boundary response	Gate std.
Sigmoid Gate	8.90%	0.700	0.200
STEGate	38.61%	0.841	0.468

References

[1] Yue, G., Li, S., Cong, R., Zhou, T., Lei, B., Wang, T.: Attention-guided pyramid context network for polyp segmentation in colonoscopy images. IEEE Transactions on Instrumentation and Measurement **72**, 1–13 (2023) https://doi.org/10.1109/TIM.2023.3244219

[2] Zhang, X., Li, Q., Li, W., Guo, Y., Zhang, J., Guo, C., Chang, K., Lovell, N.H.: Fd-net: Feature distillation network for oral squamous cell carcinoma lymph node segmentation in hyperspectral imagery. IEEE Journal of Biomedical and Health Informatics **28**(3), 1552–1563 (2024) https://doi.org/10.1109/JBHI.2024.3350245

[3] Cong, R., Yang, H., Jiang, Q., Gao, W., Li, H., Wang, C., Zhao, Y., Kwong, S.: Bcs-net: Boundary, context, and semantic for automatic covid-19 lung infection segmentation from ct images. IEEE Transactions on Instrumentation and Measurement **71**, 1–11 (2022) https://doi.org/10.1109/TIM.2022.3196430

[4] Liu, Z., Zhou, Y., Xu, Y., Wang, Z.: Simplenet: A simple network for image anomaly detection and localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20402–20411 (2023). https://doi.org/10.1109/CVPR52729.2023.01954

[5] Zhang, X., Xu, M., Zhou, X.: Realnet: A feature selection network with realistic synthetic anomaly for anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16699–16708 (2024). https://doi.org/10.1109/CVPR52733.2024.01580

[6] Wang, L., Yang, J., Zhang, Y., Wang, F., Zheng, F.: Depth-aware concealed crop detection in dense agricultural scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 17201–17211 (2024). https://doi.org/10.1109/CVPR52733.2024.01628

[7] Rustia, D.J.A., Lin, C.E., Chung, J.-Y., Zhuang, Y.-J., Hsu, J.-C., Lin, T.-T.: Application of an image and environmental sensor network for automated greenhouse insect pest monitoring. Journal of Asia-Pacific Entomology **23**(1), 17–28 (2020) https://doi.org/10.1016/j.aspen.2019.11.006

[8] Suo, X., Jiang, Y., Lin, P., Zhang, Y., Wu, M., Guo, K., Xu, L.: Neuralhumanfvv: Real-time neural volumetric human performance rendering using rgb cameras. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6226–6237 (2021). https://doi.org/10.1109/CVPR46437.2021.00616

[9] Sun, Y., Chen, G., Zhou, T., Zhang, Y., Liu, N.: Context-aware cross-level fusion network for camouflaged object detection. arXiv preprint arXiv:2105.12555 (2021) https://doi.org/10.48550/arXiv.2105.12555

[10] Song, X., Zhang, P., Lu, X., Hei, X., Liu, R.: A universal multi-view guided network for salient object and camouflaged object detection. IEEE Transactions on Circuits and Systems for Video Technology **34**(11), 11184–11197 (2024) https://doi.org/10.1109/TCSVT.2024.3417607

[11] Zhu, H., Li, P., Xie, H., Yan, X., Liang, D., Chen, D., Wei, M., Qin, J.: I can find you! boundary-guided separated attention network for camouflaged object detection. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, pp. 3608–3616 (2022). https://doi.org/10.1609/aaai.v36i3.20273

[12] Ji, G.-P., Zhu, L., Zhuge, M., Fu, K.: Fast camouflaged object detection via edge-based reversible re-calibration network. Pattern Recognition **123**, 108414 (2022) https://doi.org/10.1016/j.patcog.2021.108414

[13] Wu, J., Liang, W., Hao, F., Xu, J.: Mask-and-edge co-guided separable network for camouflaged object detection. IEEE Signal Processing Letters **30**, 748–752 (2023) https://doi.org/10.1109/LSP.2023.3286787

[14] Cong, R., Sun, M., Zhang, S., Zhou, X., Zhang, W., Zhao, Y.: Frequency perception network for camouflaged object detection. In: Proceedings of the 31st ACM International Conference on Multimedia, pp. 1179–1189 (2023). https://doi.org/10.1145/3581783.3612083

[15] Xie, C., Xia, C., Yu, T., Li, J.: Frequency representation integration for camouflaged object detection. In: Proceedings of the 31st ACM International Conference on Multimedia, pp. 1789–1797 (2023). https://doi.org/10.1145/3581783.3611773

[16] Zhong, Y., Li, B., Tang, L., Kuang, S., Wu, S., Ding, S.: Detecting camouflaged object in frequency domain. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4504–4513 (2022). https://doi.org/10.1109/CVPR52688.2022.00446

[17] Lin, J., Tan, X., Xu, K., Ma, L., Lau, R.W.: Frequency-aware camouflaged object detection. ACM Transactions on Multimedia Computing, Communications and Applications **19**(2), 1–16 (2023) https://doi.org/10.1145/3545609

[18] Zhang, S., Kong, D., Xing, Y., Lu, Y., Ran, L., Liang, G., Wang, H., Zhang, Y.: Frequency-guided spatial adaptation for camouflaged object detection. IEEE Transactions on Multimedia **27**, 72–83 (2025) https://doi.org/10.1109/TMM.2024.3521681

[19] Sun, Y., Wang, S., Chen, C., Xiang, T.-Z.: Boundary-guided camouflaged object detection. arXiv preprint arXiv:2207.00794 (2022) https://doi.org/10.48550/arXiv.2207.00794

[20] Wang, X., Yao, F., Zhong, G., Cai, Q., Wang, S., Kwok, J.T.-Y.: Gbnet: Gated boundary-aware network for camouflaged object detection. IEEE Transactions on Image Processing **35**, 5297–5310 (2026) https://doi.org/10.1109/TIP.2026.3690312

[21] Xing, H., Gao, S., Tang, H., Mok, T.Q., Kang, Y., Zhang, W.: Tinycod: Tiny and effective model for camouflaged object detection. In: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1–5 (2023). https://doi.org/10.1109/ICASSP49357.2023.10095226 . IEEE

[22] Liang, W., Wu, J., Wu, Y., Mu, X., Xu, J.: Finet: Frequency injection network for lightweight camouflaged object detection. IEEE Signal Processing Letters **31**, 526–530 (2024) https://doi.org/10.1109/LSP.2024.3356416

[23] Xiao, H., Deng, L., Chen, X., Zou, H.: Fmlnet: A lightweight camouflage object detection network via frequency-rgb multi-domain feature mutual learning. Applied Soft Computing, 115178 (2026) https://doi.org/10.1016/j.asoc.2026.115178

[24] Itti, L., Koch, C.: Computational modelling of visual attention. Nature Reviews Neuroscience **2**, 194–203 (2001) https://doi.org/10.1038/35058500

[25] Field, D.J., Hayes, A., Hess, R.F.: Contour integration by the human visual system: Evidence for a local association field. Vision Research **33**(2), 173–193 (1993) https://doi.org/10.1016/0042-6989(93)90156-Q

[26] Zhou, H., Friedman, H.S., Heydt, R.: Coding of border ownership in monkey visual cortex. Journal of Neuroscience **20**(17), 6594–6611 (2000) https://doi.org/10.1523/JNEUROSCI.20-17-06594.2000

[27] Heydt, R.: Figure–ground organization and the emergence of proto-objects in the visual cortex. Frontiers in Psychology **6**, 1695 (2015) https://doi.org/10.3389/fpsyg.2015.01695

[28] Li, M., Zhao, Y., Zhang, F., Gui, G., Luo, B., Yang, C., Gui, W., Chang, K.: CSFIN: A lightweight network for camouflaged object detection via cross-stage feature interaction. Expert Systems with Applications **269**, 126451 (2025) https://doi.org/10.1016/j.eswa.2025.126451

[29] Khan, A., Ullah, H., Munir, A.: LiteCOD: Lightweight camouflaged object detection via holistic understanding of local-global features and multi-scale fusion. AI **6**(9), 197 (2025) https://doi.org/10.3390/ai6090197

[30] Xiao, H., Liu, Z., Luo, F., Chen, X., Deng, L.: ULCOD-Net: An ultra-lightweight camouflage object detection framework with gated multi-level feature fusion and dual-constraint refinement. Complex & Intelligent Systems **12**, 76 (2026) https://doi.org/10.1007/s40747-025-02201-3 . Published online 16 December 2025; article 76

[31] Wang, X., Li, B., Zhang, M., Zhao, H., Li, Y.C., Jiang, H., Fu, Q.: DS-PRNet: Lightweight camouflaged object detection with dynamic sparse attention and feature-, attention-, and output-level distillation. Scientific Reports **16**, 30061 (2026) https://doi.org/10.1038/s41598-026-66102-2

[32] 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 (CVPR), pp. 2777–2787 (2020). https://doi.org/10.1109/CVPR42600.2020.00285

[33] Fan, D.-P., Ji, G.-P., Cheng, M.-M., Shao, L.: Concealed object detection. IEEE transactions on pattern analysis and machine intelligence **44**(10), 6024–6042 (2021) https://doi.org/10.1109/TPAMI.2021.3085766

[34] Pang, Y., Zhao, X., Xiang, T.-Z., Zhang, L., Lu, H.: Zoom in and out: A mixed-scale triplet network for camouflaged object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2160–2170 (2022). https://doi.org/10.1109/CVPR52688.2022.00220

[35] Zhang, M., Xu, S., Piao, Y., Shi, D., Lin, S., Lu, H.: Preynet: Preying on camouflaged objects. In: Proceedings of the 30th ACM International Conference on Multimedia, pp. 5323–5332 (2022). https://doi.org/10.1145/3503161.3548178

[36] Zhao, Y., Zhang, Q., Li, Y.: FLRNet: A bio-inspired three-stage network for

camouflaged object detection via filtering, localization and refinement. Neurocomputing **626**, 129523 (2025) https://doi.org/10.1016/j.neucom.2025.129523

[37] Chen, C., Liang, W., Wang, D., Wang, B., Xu, J.: Vision-inspired boundary perception network for lightweight camouflaged object detection. IEEE Signal Processing Letters **32**, 1176–1180 (2025) https://doi.org/10.1109/LSP.2025.3549698

[38] Zhang, J., Yuan, Y., Zhu, X., Hu, Y., Zhang, K.: Partitioned observation network for camouflaged object detection. Pattern Recognition **172**, 112637 (2026) https://doi.org/10.1016/j.patcog.2025.112637

[39] Sun, D., Jiang, S., Qi, L.: Edge-aware mirror network for camouflaged object detection. In: 2023 IEEE International Conference on Multimedia and Expo (ICME), pp. 2465–2470. IEEE, ??? (2023). https://doi.org/10.1109/ICME55011.2023.00420

[40] Zhuang, C., Zhang, Q., Zhang, C., Yuan, X.: Boundary-and-object collaborative learning network for camouflaged object detection. Image and Vision Computing, 105596 (2025) https://doi.org/10.1016/j.imavis.2025.105596

[41] He, C., Li, K., Zhang, Y., Tang, L., Zhang, Y., Guo, Z., Li, X.: Camouflaged object detection with feature decomposition and edge reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 22046–22055 (2023). https://doi.org/10.1109/CVPR52729.2023.02111

[42] S, A.A., Naidu, A.P., Rath, A.: LaFINet: Laplacian-based frequency injection network for camouflage object detection (student abstract). In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 40, pp. 41107–41109 (2026). https://doi.org/10.1609/aaai.v40i48.42179 . Student Abstract and Poster Program

[43] Mehta, S., Rastegari, M.: Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. arXiv preprint arXiv:2110.02178 (2021) https://doi.org/10.48550/arXiv.2110.02178

[44] Maass, W.: Networks of spiking neurons: the third generation of neural network models. Neural networks **10**(9), 1659–1671 (1997) https://doi.org/10.1016/S0893-6080(97)00011-7

[45] Wu, Y., Deng, L., Li, G., Zhu, J., Xie, Y., Shi, L.: Direct training for spiking neural networks: Faster, larger, better. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 1311–1318 (2019). https://doi.org/10.1609/aaai.v33i01.33011311

[46] Bengio, Y., Léonard, N., Courville, A.: Estimating or Propagating Gradients

Through Stochastic Neurons for Conditional Computation (2013)

[47] Milletari, F., Navab, N., Ahmadi, S.-A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571 (2016). https://doi.org/10.1109/3DV.2016.79 . IEEE

[48] Skurowski, P., Abdulameer, H., Błaszczyk, J., Depta, T., Kornacki, A., Kozieł, P.: Animal camouflage analysis: Chameleon database. Unpublished manuscript **2**(6), 7 (2018)

[49] Le, T.-N., Nguyen, T.V., Nie, Z., Tran, M.-T., Sugimoto, A.: Anabranch network for camouflaged object segmentation. Computer vision and image understanding **184**, 45–56 (2019) https://doi.org/10.1016/j.cviu.2019.04.006

[50] Lv, Y., Zhang, J., Dai, Y., Li, A., Liu, B., Barnes, N., Fan, D.-P.: Simultaneously localize, segment and rank the camouflaged objects. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11586–11596 (2021). https://doi.org/10.1109/CVPR46437.2021.01142 . IEEE

[51] Zhai, Q., Li, X., Yang, F., Chen, C., Cheng, H., Fan, D.-P.: Mutual graph learning for camouflaged object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12997–13007 (2021). https://doi.org/10.1109/CVPR46437.2021.01280

[52] Li, A., Zhang, J., Lv, Y., Liu, B., Zhang, T., Dai, Y.: Uncertainty-aware joint salient object and camouflaged object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10071–10081 (2021). https://doi.org/10.1109/CVPR46437.2021.00994

[53] Yang, F., Zhai, Q., Li, X., Huang, R., Luo, A., Cheng, H., Fan, D.-P.: Uncertainty-guided transformer reasoning for camouflaged object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 4146–4155 (2021)

[54] Mei, H., Ji, G.-P., Wei, Z., Yang, X., Wei, X., Fan, D.-P.: Camouflaged object segmentation with distraction mining. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8772–8781 (2021). https://doi.org/10.1109/CVPR52688.2022.00220

[55] Li, P., Yan, X., Zhu, H., Wei, M., Zhang, X.-P., Qin, J.: Findnet: Can you find me? boundary-and-texture enhancement network for camouflaged object detection. IEEE Transactions on Image Processing **31**, 6396–6411 (2022) https://doi.org/10.1109/TIP.2022.3189828

[56] Chen, G., Liu, S.-J., Sun, Y.-J., Ji, G.-P., Wu, Y.-F., Zhou, T.: Camouflaged

object detection via context-aware cross-level fusion. IEEE Transactions on Circuits and Systems for Video Technology **32**(10), 6981–6993 (2022) https://doi.org/10.1109/TCSVT.2022.3178173

[57] Zhai, W., Cao, Y., Xie, H., Zha, Z.-J.: Deep texton-coherence network for camouflaged object detection. IEEE Transactions on Multimedia **25**, 5155–5165 (2023) https://doi.org/10.1109/TMM.2022.3188401

[58] Zhai, Q., Li, X., Yang, F., Jiao, Z., Luo, P., Cheng, H., Liu, Z.: Mgl: Mutual graph learning for camouflaged object detection. IEEE Transactions on Image Processing **32**, 1897–1910 (2023) https://doi.org/10.1109/TIP.2022.3223216

[59] Yan, X., Sun, M., Han, Y., Wang, Z.: Camouflaged object segmentation based on matching–recognition–refinement network. IEEE Transactions on Neural Networks and Learning Systems **35**(11), 15993–16007 (2024) https://doi.org/10.1109/TNNLS.2023.3291595

[60] Zhou, X., Wu, Z., Cong, R.: Decoupling and integration network for camouflaged object detection. IEEE Transactions on Multimedia **26**, 7114–7129 (2024) https://doi.org/10.1109/TMM.2024.3360710

[61] Fang, X., Chen, J., Wang, Y., Jiang, M., Ma, J., Wang, X.: Epfdnet: Camouflaged object detection with edge perception in frequency domain. Image and Vision Computing **154**, 105358 (2025) https://doi.org/10.1016/j.imavis.2024.105358

[62] Wang, Z., Zhang, Y., Liu, Y., Qin, C., Coleman, S.A., Kerr, D.: Larnet: Towards lightweight, accurate and real-time salient object detection. IEEE Transactions on Multimedia **26**, 5207–5222 (2023) https://doi.org/10.1109/TMM.2023.3330082