# DISCRIMINATIVE EVIDENCE MINING AND RECONSTRUCTION FOR EFFICIENT CAMOUFLAGED OBJECT DETECTION

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Camouflaged Object Detection (COD) aims to detect objects that closely resemble their background in complex scenes. Existing lightweight COD methods typically reduce computation by compressing feature dimensions and simplifying feature interactions. However, such approaches tend to weaken the subtle discriminative evidence required to distinguish camouflaged objects from their backgrounds, including semantic cues and fine-grained structural details. Therefore, we propose an efficient Discriminative Evidence Mining and Reconstruction Network (DMRNet), which employs the Discriminative Evidence Mining (DEM) strategy and the Discriminative Evidence Reconstruction (DER) strategy. Through the DEM strategy, DMRNet exploits cross-layer feature interactions to mine discriminative evidence with foreground-background competition guiding evidence selection and background interference suppression reducing misleading responses. Through the DER strategy, DMRNet uses the obtained discriminative evidence to guide hierarchical feature reconstruction, thereby enhancing semantic representations at higher levels and fine-grained structural details at the shallow level. DMRNet achieves competitive detection performance on four benchmark datasets while maintaining a low parameter count and computational cost, thereby achieving an effective balance between accuracy and efficiency. The source code will be released after paper acceptance.

## 1 INTRODUCTION

Camouflaged Object Detection (COD) aims to accurately locate and segment objects that are difficult to detect in complex scenes (Han et al., 2026; Chen et al., 2025). Unlike visual tasks such as salient object detection (Kupyn et al., 2026; He et al., 2025a), the objects of interest in COD do not exhibit obvious visual salience, which demands stronger perceptual and discriminative capabilities from the model (Hao et al., 2025). Accurate detection of camouflaged objects has significant practical value, with applications in wildlife monitoring (Feng et al., 2025), agricultural vision (Wang et al., 2024), and medical image segmentation (Lei et al., 2025).

The main challenge in COD lies in the lack of distinct discriminative evidence between the object and the background. Camouflaged objects often closely resemble their surroundings in color, texture, and structure, and are affected by blurred boundaries, varying scales, and complex background interference, making the model prone to missed detections and false background activations. Meanwhile, features at different levels exhibit distinct representational characteristics. Shallow-level features preserve rich spatial details but are susceptible to interference from background textures, whereas deep-level features contain richer semantic information but may lose fine-grained structures during successive down-sampling (Ge et al., 2026). Therefore, the key to accurate COD lies in how to identify and preserve discriminative evidence from multi-level representations while suppressing misleading background interference (Song et al., 2025b).

Accordingly, existing deep COD methods enhance object representation through multi-scale feature interaction (Chen et al., 2026; He et al., 2025b), boundary and texture enhancement (Ye et al., 2025; Yue et al., 2025), context modeling (Zhou et al., 2025; Yang et al., 2026), and attention mechanisms (Sun et al., 2024; Yin & Yue, 2025), achieving significant performance improvements. However, to



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

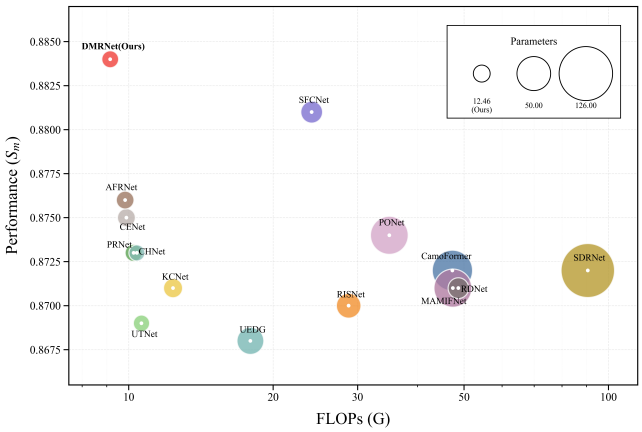



Figure 1: Performance and efficiency comparison of different models on the CAMO dataset. The bubble size represents the number of parameters.

fully capture subtle object cues, these methods typically require complex multi-branch structures, frequent cross-layer interactions, or progressive prediction refinement (Sun et al., 2025b). While intensive high-dimensional feature processing improves detection performance, it also results in a large number of parameters and high computational cost, making it difficult to balance detection accuracy and inference efficiency. Therefore, focusing solely on detection accuracy is insufficient, and research on efficient COD models is also crucial (Liang et al., 2026).

To improve efficiency, lightweight COD methods reduce model complexity by reducing feature channel dimensions and simplifying feature interaction structures, thereby significantly reducing the number of parameters and computational cost while improving inference speed (Hu et al., 2024; Wang et al., 2025a). However, lightweight models limit the network's ability to represent features, which may further weaken the subtle discriminative evidence between the object and the background. Moreover, although directly fusing features from different levels can supplement semantic and detailed information (Wu et al., 2025b), it may propagate both valid object responses and object-like background interference (Yao et al., 2024). As shown in Fig. 1, existing methods still require a better balance among the number of model parameters, computational cost, and detection accuracy. Thus, the key to efficient COD lies not only in reducing computational redundancy (Wang et al., 2026b), but also in fully exploiting effective discriminative evidence under limited representation capacity and suppressing misleading information generated by complex backgrounds.

Therefore, we propose a Discriminative Evidence Mining and Reconstruction Network (DMRNet) for efficient camouflaged object detection, which aims to preserve and enhance discriminative evidence under limited representation capacity while achieving a balance between accuracy and efficiency. Specifically, through the Discriminative Evidence Mining (DEM) strategy, DMRNet exploits cross-layer interactions among adjacent hierarchical features to mine latent discriminative evidence, with foreground-background competition guiding evidence selection and background interference suppression reducing irrelevant responses under complex backgrounds. Through the Discriminative Evidence Reconstruction (DER) strategy, DMRNet uses the reliable evidence obtained by DEM to progressively reconstruct hierarchical features, thereby enhancing semantic representations at higher levels and fine-grained structural details at the shallow level. Through the collaboration between DEM and DER, DMRNet avoids complex high-dimensional feature interactions while enhancing the capability of lightweight features to discriminate camouflaged objects from their backgrounds at a low computational cost. Experiments on four datasets demonstrate that DMRNet achieves competitive detection performance with a low parameter count and computational cost. Ablation experiments further validate the effectiveness of the proposed components. Our main contributions are as follows:

(1) We propose DMRNet, an efficient camouflaged object detection framework that addresses the degradation of discriminative evidence in lightweight networks by mining and reconstructing reliable evidence, achieving a favorable accuracy-efficiency balance.

(2) We design a Discriminative Evidence Mining (DEM) strategy that exploits cross-layer feature interactions to mine latent discriminative evidence from hierarchical representations, with foreground-background competition and background interference suppression enhancing object-related evidence while reducing misleading background interference.

(3) We design a Discriminative Evidence Reconstruction (DER) strategy that uses reliable discriminative evidence to guide hierarchical feature reconstruction, progressively enhancing semantic representations in high-level features and fine-grained structural details in shallow features to improve object representation with low computational cost.

## 2 Related Work

**Camouflaged object detection (COD)** methods primarily aim to distinguish camouflaged objects from their backgrounds by exploiting complementary visual cues and task-specific feature modeling. FGSA-Net (Zhang et al., 2025) and CamoFocus (Khan et al., 2024) enhance foreground-background discrimination through frequency guidance and contextual modeling, respectively. Structural cues, such as boundaries and textures, also play an important role in COD. BTDGNet (Song et al., 2025a), ESCNet (Ye et al., 2025), and GBNet (Wang et al., 2026c) exploit these cues to improve object localization and detail recovery. In addition, HUNTNet (Ji et al., 2025) models multi-scale spatial-semantic relationships, CODIB (Li et al., 2026) suppresses redundant information through an information bottleneck, and ICEG (He et al., 2024) improves representation learning by generating challenging camouflaged objects. Although these methods significantly improve COD performance, many high-performance models rely on complex attention mechanisms, multi-branch feature processing, and progressive refinement strategies to capture subtle structural and semantic differences between camouflaged objects and their backgrounds. Such complex designs typically incur additional parameters and computational cost. Therefore, reducing model complexity while maintaining strong detection performance has gradually become an important research direction in COD.

**Lightweight COD** methods seek to balance detection accuracy and efficiency through efficient feature modeling and interaction. PRNet (Hu et al., 2024) combines global localization with lightweight top-down refinement, while KCNet (Wu et al., 2025a) exploits edge and semantic cues for efficient localization and detail enhancement. CSFIN (Li et al., 2025) improves feature representations through cross-stage feature interaction, and ESNet (Ren et al., 2025) preserves object topology via lightweight skeleton modeling. UTNet (Sun et al., 2025a) achieves efficient detection through region-aware processing and progressive decoding. More recently, AFRNet (Xiao et al., 2026) enhances foreground-background separability from a frequency perspective, while CENet (Ye et al., 2026) suppresses information noise through efficient context aggregation, cross-layer interaction, and edge refinement. Although these methods substantially reduce model complexity, the limited representation capacity of lightweight networks may weaken subtle foreground-background discrimination, while direct cross-layer fusion can propagate background noise together with useful object cues. Therefore, lightweight COD requires both low computational cost and effective preservation of discriminative evidence. To this end, we propose DMRNet, which mines discriminative evidence through cross-layer feature interactions, guides evidence selection via foreground-background competition, suppresses background interference, and reconstructs hierarchical features using the reliable evidence with low computational cost.

## 3 Method

### 3.1 Overview

As shown in Fig. 2, DMRNet consists of the SMT-T (Lin et al., 2023) backbone, Discriminative Evidence Mining (DEM), and Discriminative Evidence Reconstruction (DER) strategies. The input image is processed by SMT-T to extract hierarchical features $\{F_1, F_2, F_3, F_4\}$ at four levels, where $F_4$ is first projected into $X_4$ to generate the prediction $P_4$, followed by sequential evidence mining and reconstruction at each level. Through the DEM strategy, DMRNet employs the Cross-Layer Discriminative Evidence Mining (CDEM) module to extract discriminative evidence, while the Background Interference Suppression (BIS) module further suppresses background interference at the $F_3$ and $F_2$ levels. Through the DER strategy, DMRNet enhances semantic representations

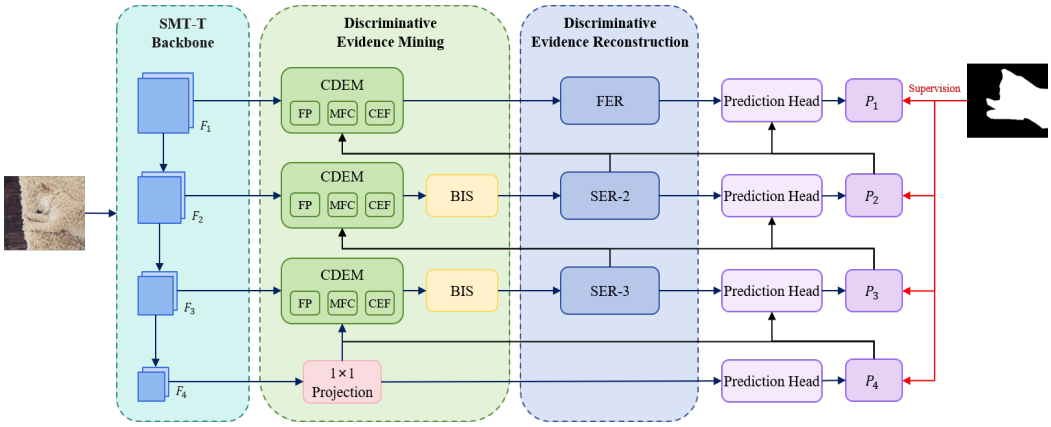



Figure 2: Overview of the proposed DMRNet. It consists of the SMT-T backbone, Discriminative Evidence Mining (DEM), and Discriminative Evidence Reconstruction (DER) strategies.

at higher levels and fine-grained structural representations at the shallow level using Semantic Evidence Reconstruction (SER) and Fine-Grained Evidence Reconstruction (FER), respectively. Finally, predictions $P_4$, $P_3$, $P_2$, and $P_1$ are progressively generated at their corresponding levels.

### 3.2 Discriminative Evidence Mining

#### 3.2.1 Cross-Layer Discriminative Evidence Mining

To fully exploit the subtle discriminative cues between camouflaged objects and their backgrounds, we design Cross-Layer Discriminative Evidence Mining (CDEM). For the $i$-th level, where $i \in \{1, 2, 3\}$, we denote the current-level feature as $F_i$, the output feature from the adjacent higher level as $X_{i+1}$, and the corresponding prediction as $P_{i+1}$.

**Feature Projection (FP).** Since features at different levels differ in channel dimension and spatial resolution, they are first projected for alignment:

$$L_i = \phi_i(F_i), H_i = \mathcal{U}(\psi_i(X_{i+1}), size(L_i)), \tag{1}$$

where $\phi_i(\cdot)$ and $\psi_i(\cdot)$ denote feature projection operations consisting of a $1 \times 1$ convolution, batch normalization, and a GELU activation function, and $\mathcal{U}(\cdot)$ denotes bilinear interpolation.

**Mask-Guided Foreground-Background Competition (MFC).** We further utilize the adjacent higher-level prediction $P_{i+1}$ to construct foreground-background competition. The prediction is then resized to the current-level feature resolution and converted into a probability mask $M_i$:

$$M_i = \sigma(\mathcal{U}(P_{i+1}, size(L_i))), \tag{2}$$

where $\sigma(\cdot)$ denotes the Sigmoid function. To reduce interference from unreliable prediction regions when constructing reference information, we further define the prediction confidence $C_i$ and obtain the foreground weight $W_i^f$ and background weight $W_i^b$:

$$C_i = |2M_i - 1|^2, \tag{3}$$

$$W_i^f = M_i \odot C_i, W_i^b = (1 - M_i) \odot C_i, \tag{4}$$

where $\odot$ denotes element-wise multiplication. Then, we use the high-level semantic features $H_i$ to construct foreground prototype $Z_i^f$ and background prototype $Z_i^b$:

$$Z_i^f = \frac{\sum_x W_i^f(x) H_i(x)}{\sum_x W_i^f(x) + \epsilon}, Z_i^b = \frac{\sum_x W_i^b(x) H_i(x)}{\sum_x W_i^b(x) + \epsilon}, \tag{5}$$

where $x$ represents spatial position, and $\epsilon$ is a constant introduced to avoid numerical instability. Based on the two reference prototypes, we calculate the similarity responses $S_i^f$ and $S_i^b$ between the current feature $L_i$ and the foreground/background prototypes.

$$S_i^f = sim(L_i, Z_i^f), S_i^b = sim(L_i, Z_i^b), \tag{6}$$

where $\mathrm{sim}(\cdot,\cdot)$ denotes cosine similarity. We then construct a foreground-background competitive response $A_i$:

$$A_i = \tanh(\tau(S_i^f - S_i^b)), \tag{7}$$

where $\tau$ is the temperature coefficient. Since the ambiguous regions in coarse predictions contain object evidence that is more difficult to distinguish, we also calculate the prediction uncertainty map $U_i$:

$$U_i = 4M_i(1 - M_i). \tag{8}$$

Finally, the foreground-background competition relationship and uncertainty are combined to generate routing weight $\Omega_i$:

$$\Omega_i = Clip(1 + \alpha_i A_i + \beta_i U_i, r_{\min}, r_{\max}), \tag{9}$$

where $\alpha_i$ and $\beta_i$ are learnable adjustment coefficients, and $r_{\min}$ and $r_{\max}$ are used to limit the routing magnitude. In this way, MFC provides adaptive routing guidance for subsequent evidence fusion.

**Competition-Guided Evidence Fusion (CEF).** After obtaining the routing weight $\Omega_i$, we use it to perform competition-guided adaptive modulation on high-level semantic information:

$$\widetilde{H}_i = \gamma_i \Omega_i \odot H_i, \tag{10}$$

where $\gamma_i$ is a learnable scaling parameter. Then, the projected current-level features, the higher-level semantic features, and the competitively modulated features are fused together:

$$E_i = \mathcal{F}_i([L_i, H_i, \widetilde{H}_i]), \tag{11}$$

where $[\cdot]$ denotes channel-wise concatenation, and $\mathcal{F}_i(\cdot)$ comprises a $1 \times 1$ projection followed by a residual local refinement using a $3 \times 3$ depthwise convolution and a $1 \times 1$ pointwise convolution. $E_i$ denotes the discriminative evidence feature produced by CDEM. This process provides more reliable feature representations for subsequent modules.

#### 3.2.2 Background Interference Suppression

Since the features at the $F_3$ and $F_2$ levels are further propagated to subsequent levels, residual background interference may accumulate during progressive decoding. Therefore, we introduce Background Interference Suppression (BIS) at these two levels. For $i \in \{2, 3\}$, we first obtain the background suppression map $Q_i$ based on $E_i$ from the CDEM output:

$$Q_i = \sigma(\mathcal{P}^i_{1\times1}(\mathcal{D}^i_{3\times3}(E_i))), \tag{12}$$

where $\mathcal{D}^i_{3\times3}(\cdot)$ denotes a depthwise convolution, and $\mathcal{P}^i_{1\times1}(\cdot)$ denotes a pointwise convolution. $Q_i$ indicates potential background interference at different spatial locations and channels. Then, we use a residual approach to suppress background responses:

$$\widetilde{E}_i = E_i - \lambda_i Q_i \odot E_i, \tag{13}$$

where $\lambda_i$ is a learnable scaling parameter. To further integrate the effective responses after suppression, we use a lightweight local refinement to recode $\widetilde{E}_i$ and obtain the background-suppressed feature $B_i$:

$$B_i = \widetilde{E}_i + \mathcal{P}^i_{1\times1}(\mathcal{D}^i_{3\times3}(\widetilde{E}_i)). \tag{14}$$

Through the above process, BIS further reduces the residual background noise in $F_3$ and $F_2$ with minimal computational cost, providing a more reliable feature representation for discriminative evidence reconstruction.

### 3.3 Discriminative Evidence Reconstruction

#### 3.3.1 Semantic Evidence Reconstruction

After DEM processing, the features at the $F_3$ and $F_2$ levels contain relatively reliable object-related discriminative evidence. However, these responses remain locally distributed within the current representation and are not explicitly organized into more effective semantic representations. Therefore,

Semantic Evidence Reconstruction (SER) further recodes and reorganizes the retained evidence for $i \in \{2, 3\}$. The BIS output features $B_i$ are first refined through a lightweight local refinement to further integrate the valid responses retained after suppressing background noise:

$$Z_i^{(k)} = Z_i^{(k-1)} + \mathcal{P}_{1\times1}^{i,k}(\mathcal{D}_{3\times3}^{i,k}(Z_i^{(k-1)})), k = 1, ..., N_i, \tag{15}$$

where $Z_i^{(0)} = B_i$. Considering their different representation characteristics, we set $N_3 = 2$ and $N_2 = 1$, respectively. We further use grouped convolutions to recode the refined features and enhance discriminative interactions within channel groups:

$$R_i = \mathcal{G}_{3\times3}^{i}(Z_i^{(N_i)}), i \in \{2, 3\}, \tag{16}$$

where $\mathcal{G}_{3\times3}^{i}(\cdot)$ denotes a $3 \times 3$ grouped convolution with 2 groups. We then reconstruct the recoded features and feed them back into the main branch using a learnable residual structure:

$$X_i = Z_i^{(N_i)} + \eta_i R_i, i \in \{2, 3\}, \tag{17}$$

where $\eta_i$ is a learnable scaling parameter. Thus, each reconstructed feature $X_i, i \in \{2, 3\}$, is used to generate the prediction at the current level and serves as the higher-level input for discriminative evidence mining at the next lower level.

### 3.3.2 Fine-Grained Evidence Reconstruction

Unlike $F_3$ and $F_2$, $F_1$ has the highest spatial resolution and contains richer local texture and structural information. Therefore, we employ Fine-Grained Evidence Reconstruction (FER) at the $F_1$ level. First, we use lightweight local refinement to initially refine the object-related responses in $E_1$:

$$Z_1 = E_1 + \mathcal{P}_{1\times1}^{1}(\mathcal{D}_{3\times3}^{1}(E_1)). \tag{18}$$

To further enhance the boundary and fine-scale structural representation of the object, we perform two $3 \times 3$ grouped convolutions in succession for fine-grained local reconstruction:

$$R_1^{(1)} = \mathcal{G}_{3\times3}^{1,1}(Z_1), R_1^{(2)} = \mathcal{G}_{3\times3}^{1,2}(R_1^{(1)}), \tag{19}$$

where the number of groups in the grouped convolution is set to 4. Since grouped convolution primarily exchanges information within a group, we further use $1 \times 1$ convolutions to fuse evidence across different channel groups and perform the final reconstruction using a residual structure:

$$X_1 = Z_1 + \eta_1 \mathcal{F}_{1\times1}^{1}(R_1^{(2)}), \tag{20}$$

where $\mathcal{F}_{1\times1}^{1}(\cdot)$ denotes $1 \times 1$ cross-channel fusion. Compared to the SER applied to $F_3$ and $F_2$, FER focuses on reconstructing fine-grained structural representations in $F_1$ through more thorough local reconstruction and cross-group channel interactions, thereby providing a more complete representation of the object's structure for accurate final detection.

## 3.4 Prediction and Loss Function

**Prediction.** To generate prediction maps for camouflaged object localization, a prediction head is attached to each decoding level. Given the deepest feature $F_4$, the prediction map $P_4$ is generated through a $1 \times 1$ convolution. For the subsequent decoding levels, the prediction head estimates a residual prediction $\Delta P_i$ from the current feature $X_i$, which is then added to the up-sampled prediction map from the previous level to progressively refine the prediction:

$$P_i = \mathcal{U}(P_{i+1}) + \Delta P_i, i \in \{1, 2, 3\}, \tag{21}$$

where $\Delta P_i$ is obtained by applying a $1 \times 1$ convolution to $X_i$. The generated prediction maps provide high-level guidance for the subsequent cross-layer discriminative evidence mining.

**Loss Function.** To supervise the predictions at different levels, we employ both Binary Cross Entropy (BCE) loss and Intersection-over-Union (IoU) loss for joint supervision on $P_1$, $P_2$, $P_3$, and $P_4$. The overall optimization objective is formulated as:

$$\mathcal{L}_{total} = \sum_{i=1}^{4} (\mathcal{L}_{BCE}^{i}(P_i, GT) + \mathcal{L}_{IoU}^{i}(P_i, GT)), \tag{22}$$

where $\mathcal{L}_{BCE}^{i}$ and $\mathcal{L}_{IoU}^{i}$ represent the BCE loss and the IoU loss for the predictions at the i-th level, respectively.

Table 1: Performance and efficiency comparison of DMRNet and other COD models on four benchmark datasets. The best and second-best results are highlighted in red and blue, respectively. F denotes FLOPs, S denotes Speed, and P denotes the number of parameters.

| Methods | F | S | P | CAMO (250) | | | | CHAMELEON (76) | | | | COD10K (2026) | | | | NC4K (4121) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (G) | (FPS) | (M) | $S_m \uparrow$ | $F_\beta^w \uparrow$ | $M \downarrow$ | $E_m \uparrow$ | $S_m \uparrow$ | $F_\beta^w \uparrow$ | $M \downarrow$ | $E_m \uparrow$ | $S_m \uparrow$ | $F_\beta^w \uparrow$ | $M \downarrow$ | $E_m \uparrow$ | $S_m \uparrow$ | $F_\beta^w \uparrow$ | $M \downarrow$ | $E_m \uparrow$ |
| *High-complexity models* | | | | | | | | | | | | | | | | | | | |
| SDRNet | 90.54 | 21.02 | 126.0 | 0.872 | 0.826 | 0.049 | 0.924 | 0.914 | 0.868 | 0.023 | 0.950 | 0.871 | 0.785 | 0.023 | 0.924 | 0.889 | 0.842 | 0.032 | 0.934 |
| CamoFormer | 47.27 | 24.15 | 71.40 | 0.872 | 0.831 | 0.046 | 0.929 | 0.910 | 0.865 | 0.022 | 0.957 | 0.869 | 0.786 | 0.023 | 0.932 | 0.892 | 0.846 | 0.031 | 0.937 |
| MAMIFNet | 47.38 | 10.02 | 62.82 | 0.871 | 0.834 | 0.045 | 0.929 | 0.914 | 0.874 | 0.021 | 0.957 | 0.867 | 0.787 | 0.023 | 0.933 | 0.889 | 0.847 | 0.031 | 0.938 |
| PONet | 34.93 | - | 62.68 | 0.874 | 0.833 | 0.045 | 0.929 | 0.911 | 0.858 | 0.023 | 0.954 | 0.874 | 0.792 | 0.022 | 0.934 | 0.892 | 0.848 | 0.030 | 0.938 |
| *Moderate-complexity models* | | | | | | | | | | | | | | | | | | | |
| UEDG | 17.93 | 37.53 | 31.68 | 0.868 | 0.819 | 0.048 | 0.922 | 0.911 | 0.866 | 0.022 | 0.960 | 0.858 | 0.766 | 0.025 | 0.924 | 0.881 | 0.829 | 0.035 | 0.928 |
| RISNet | 28.73 | 31.24 | 26.58 | 0.870 | 0.827 | 0.050 | 0.922 | 0.908 | 0.863 | 0.024 | 0.951 | 0.873 | 0.799 | 0.025 | 0.931 | 0.882 | 0.834 | 0.037 | 0.925 |
| SFCNet | 24.06 | 37.80 | 20.87 | 0.881 | 0.844 | 0.045 | 0.932 | 0.911 | 0.868 | 0.023 | 0.962 | 0.871 | 0.793 | 0.023 | 0.932 | 0.893 | 0.851 | 0.030 | 0.938 |
| RDNet | 48.64 | 26.94 | 18.57 | 0.871 | 0.826 | 0.050 | 0.920 | 0.911 | 0.871 | 0.021 | 0.959 | 0.868 | 0.788 | 0.023 | 0.931 | 0.889 | 0.843 | 0.032 | 0.934 |
| *Lightweight models* | | | | | | | | | | | | | | | | | | | |
| KCNet | 12.37 | 46.90 | 15.96 | 0.871 | 0.822 | 0.048 | 0.923 | 0.910 | 0.858 | 0.022 | 0.957 | 0.862 | 0.775 | 0.025 | 0.924 | 0.886 | 0.836 | 0.034 | 0.930 |
| PRNet | 10.24 | 51.58 | 12.74 | 0.873 | 0.829 | 0.047 | 0.923 | 0.910 | 0.865 | 0.022 | 0.960 | 0.871 | 0.792 | 0.023 | 0.933 | 0.891 | 0.848 | 0.031 | 0.936 |
| UTNet | 10.63 | 36.57 | 11.63 | 0.869 | 0.827 | 0.049 | 0.924 | 0.913 | 0.873 | 0.020 | 0.959 | 0.868 | 0.790 | 0.023 | 0.933 | 0.889 | 0.845 | 0.032 | 0.936 |
| CHNet | 10.40 | 41.57 | 11.20 | 0.873 | 0.828 | 0.048 | 0.923 | 0.912 | 0.868 | 0.022 | 0.961 | 0.870 | 0.788 | 0.023 | 0.932 | 0.891 | 0.844 | 0.032 | 0.936 |
| AFRNet | 9.83 | - | 13.72 | 0.876 | 0.826 | 0.047 | 0.928 | - | - | - | - | 0.868 | 0.775 | 0.024 | 0.928 | 0.892 | 0.840 | 0.032 | 0.936 |
| CENet | 9.89 | 52.53 | 14.05 | 0.875 | 0.832 | 0.046 | 0.928 | 0.914 | 0.870 | 0.021 | 0.962 | 0.868 | 0.789 | 0.023 | 0.933 | 0.889 | 0.844 | 0.031 | 0.935 |
| **DMRNet** | 9.15 | 53.74 | 12.46 | 0.884 | 0.846 | 0.044 | 0.933 | 0.918 | 0.875 | 0.019 | 0.962 | 0.873 | 0.796 | 0.022 | 0.936 | 0.893 | 0.849 | 0.031 | 0.938 |

# 4 Experiment and Analysis

## 4.1 Experimental Setup

**Datasets.** We used four benchmark datasets for our experiments. The CAMO (Le et al., 2019) dataset covers a variety of natural and artificial camouflage scenarios, including 1,000 training images and 250 test images. The CHAMELEON (Skurowski et al., 2018) dataset focuses primarily on camouflaged animals in natural environments and includes 76 test images. The COD10K (Fan et al., 2020) dataset covers a wide range of camouflage categories and complex scenarios, including 3,040 training images and 2,026 test images. NC4K (Lv et al., 2021) focuses on evaluating the model's generalization performance across different camouflage scenarios and includes 4,121 test images.

**Implementation details.** DMRNet is implemented using PyTorch. We select SMT-T as the backbone, and all input images are uniformly resized to $384 \times 384$. All experiments, including FPS measurements, are conducted on a single NVIDIA GeForce RTX 3090, with FPS measured using a batch size of 1. The network is trained for 300 epochs with a batch size of 8. The Adam optimizer is used. The initial learning rate is fixed at $5 \times 10^{-5}$ and decays by a factor of 0.1 every 100 epochs. The temperature coefficient $\tau$ is set to 4.0, the clipping bounds $r_{\min}$ and $r_{\max}$ are set to 0.25 and 2.0, respectively, and $\epsilon$ is set to $1 \times 10^{-6}$ for numerical stability.

**Evaluation metrics.** We use four metrics to evaluate model performance. The S-measure ($S_m$) evaluates structural consistency. The weighted F-measure ($F_\beta^w$) balances precision and recall through weighting to evaluate segmentation performance. The MAE ($M$) calculates the mean absolute error to assess overall accuracy. The mean E-measure ($E_m$) evaluates the consistency of pixel distribution in the segmentation results.

## 4.2 Comparison with State-of-the-Art

We select CamoFormer (Yin et al., 2024), UEDG (Lyu et al., 2024), RISNet (Wang et al., 2024), SDRNet (Guan et al., 2024), PRNet (Hu et al., 2024), UTNet (Sun et al., 2025a), KCNet (Wu et al., 2025a), CHNet (Wang et al., 2025a), MAMIFNet (Wang et al., 2025b), SFCNet (Zhao et al., 2025), AFRNet (Xiao et al., 2026), CENet (Ye et al., 2026), PONet (Zhang et al., 2026), and RDNet (Wang et al., 2026a) as the compared models.

**Quantitative Evaluation.** Tab. 1 presents the comparison results between DMRNet and existing camouflaged object detection methods on four benchmark datasets. DMRNet achieves competitive

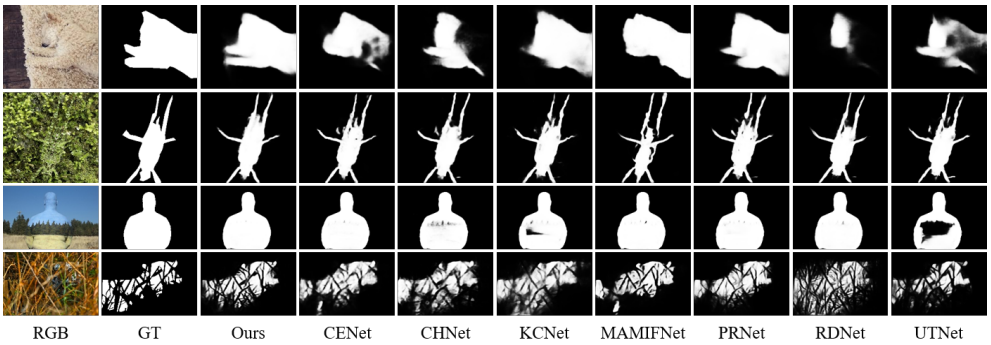



Figure 3: A qualitative comparison of DMRNet with other methods in typical complex scenarios.

Table 2: Ablation study of different components in DMRNet. The best results are highlighted in red.

| Settings | | | | CAMO | | | | CHAMELEON | | | | COD10K | | | | NC4K | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Backbone | CDEM | BIS | DER | $S_m \uparrow$ | $F_\beta^w \uparrow$ | $M \downarrow$ | $E_m \uparrow$ | $S_m \uparrow$ | $F_\beta^w \uparrow$ | $M \downarrow$ | $E_m \uparrow$ | $S_m \uparrow$ | $F_\beta^w \uparrow$ | $M \downarrow$ | $E_m \uparrow$ | $S_m \uparrow$ | $F_\beta^w \uparrow$ | $M \downarrow$ | $E_m \uparrow$ |
| ✓ | | | | 0.551 | 0.277 | 0.235 | 0.523 | 0.581 | 0.258 | 0.189 | 0.534 | 0.578 | 0.218 | 0.149 | 0.561 | 0.602 | 0.305 | 0.189 | 0.575 |
| ✓ | ✓ | | | 0.874 | 0.832 | 0.048 | 0.925 | 0.909 | 0.862 | 0.022 | 0.955 | 0.869 | 0.789 | 0.023 | 0.932 | 0.890 | 0.844 | 0.031 | 0.934 |
| ✓ | ✓ | ✓ | | 0.878 | 0.836 | 0.047 | 0.930 | 0.912 | 0.866 | 0.022 | 0.958 | 0.870 | 0.791 | 0.022 | 0.933 | 0.891 | 0.846 | 0.031 | 0.936 |
| ✓ | ✓ | ✓ | ✓ | 0.884 | 0.846 | 0.044 | 0.933 | 0.918 | 0.875 | 0.019 | 0.962 | 0.873 | 0.796 | 0.022 | 0.936 | 0.893 | 0.849 | 0.031 | 0.938 |

detection performance while maintaining low computational cost. It obtains the best or second-best results on multiple datasets and evaluation metrics, demonstrating its effectiveness in complex camouflaged scenarios. Compared with existing COD methods with different complexity levels, DMRNet achieves comparable or superior detection performance to some models with larger computational complexity while requiring only 12.46 million parameters and 9.15 G FLOPs. In addition, DMRNet achieves an inference speed of 53.74 FPS, significantly outperforming most compared methods and further demonstrating its advantage in inference efficiency. For a fair comparison of inference efficiency, the FPS of all methods is measured with a batch size of 1. Overall, DMRNet achieves an effective balance among detection performance, model complexity, and inference efficiency, demonstrating strong potential for efficient camouflaged object detection.

**Qualitative Evaluation.** As shown in Fig. 3, DMRNet demonstrates more stable segmentation quality across different types of challenging cases. In the first group, most methods exhibit significant gaps or uneven responses in the area to the right of the object, while DMRNet preserves a relatively complete object region. In the second group, DMRNet shows more continuous predictions for the object's main body and fine structures, while some methods exhibit local discontinuities. In the third group, some methods exhibit more noticeable prediction errors and missed regions within the object boundaries, while DMRNet effectively maintains the overall integrity of the object region. The fourth group contains highly complex background structures. While some methods produce noticeable false responses in the background or fragmented object regions, DMRNet generates predictions that more closely match the ground truth in terms of overall shape and region coverage. Overall, the advantages of DMRNet are mainly reflected in its ability to reduce local missing regions, preserve object integrity, and suppress false responses in complex backgrounds.

### 4.3 Ablation Study

**Different Components.** Tab. 2 presents the ablation results of different components in DMRNet. Starting from the SMT-T baseline and gradually introducing the DEM and DER strategies, the performance progressively improves. First, after introducing cross-layer discriminative evidence mining (CDEM), the model's detection performance significantly improves on all four datasets, validating the effectiveness of CDEM in enhancing feature representation capabilities. Then, after incorporating Background Interference Suppression (BIS) based on CDEM, the model's performance is further improved, indicating that BIS can provide additional performance gains for the DEM strategy. Finally, after incorporating the DER strategy, DMRNet achieves the best detec-

Table 3: Ablation study of different components in CDEM. The best results are highlighted in red.

| Settings | CAMO | | | | CHAMELEON | | | | COD10K | | | | NC4K | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $S_m \uparrow$ | $F_\beta^w \uparrow$ | $M \downarrow$ | $E_m \uparrow$ | $S_m \uparrow$ | $F_\beta^w \uparrow$ | $M \downarrow$ | $E_m \uparrow$ | $S_m \uparrow$ | $F_\beta^w \uparrow$ | $M \downarrow$ | $E_m \uparrow$ | $S_m \uparrow$ | $F_\beta^w \uparrow$ | $M \downarrow$ | $E_m \uparrow$ |
| w/o FP | 0.880 | 0.837 | 0.046 | 0.930 | 0.917 | 0.871 | 0.021 | 0.956 | 0.870 | 0.791 | 0.023 | 0.933 | 0.891 | 0.845 | 0.032 | 0.935 |
| w/o MFC | 0.880 | 0.840 | 0.046 | 0.928 | 0.911 | 0.867 | 0.022 | 0.958 | 0.871 | 0.792 | 0.023 | 0.932 | 0.892 | 0.848 | 0.031 | 0.937 |
| w/o CEF | 0.871 | 0.824 | 0.049 | 0.923 | 0.911 | 0.865 | 0.021 | 0.962 | 0.867 | 0.783 | 0.023 | 0.932 | 0.888 | 0.840 | 0.033 | 0.933 |
| **DMRNet** | 0.884 | 0.846 | 0.044 | 0.933 | 0.918 | 0.875 | 0.019 | 0.962 | 0.873 | 0.796 | 0.022 | 0.936 | 0.893 | 0.849 | 0.031 | 0.938 |

Table 4: Ablation study of different components in DER. The best results are highlighted in red.

| Settings | CAMO | | | | CHAMELEON | | | | COD10K | | | | NC4K | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $S_m \uparrow$ | $F_\beta^w \uparrow$ | $M \downarrow$ | $E_m \uparrow$ | $S_m \uparrow$ | $F_\beta^w \uparrow$ | $M \downarrow$ | $E_m \uparrow$ | $S_m \uparrow$ | $F_\beta^w \uparrow$ | $M \downarrow$ | $E_m \uparrow$ | $S_m \uparrow$ | $F_\beta^w \uparrow$ | $M \downarrow$ | $E_m \uparrow$ |
| w/o SER-3 | 0.877 | 0.837 | 0.046 | 0.929 | 0.913 | 0.868 | 0.021 | 0.955 | 0.871 | 0.792 | 0.023 | 0.933 | 0.892 | 0.848 | 0.031 | 0.937 |
| w/o SER-2 | 0.877 | 0.835 | 0.047 | 0.927 | 0.914 | 0.871 | 0.021 | 0.961 | 0.871 | 0.793 | 0.023 | 0.934 | 0.892 | 0.847 | 0.031 | 0.937 |
| w/o FER | 0.879 | 0.837 | 0.045 | 0.930 | 0.912 | 0.867 | 0.021 | 0.961 | 0.870 | 0.788 | 0.023 | 0.933 | 0.890 | 0.844 | 0.032 | 0.935 |
| **DMRNet** | 0.884 | 0.846 | 0.044 | 0.933 | 0.918 | 0.875 | 0.019 | 0.962 | 0.873 | 0.796 | 0.022 | 0.936 | 0.893 | 0.849 | 0.031 | 0.938 |

tion results, demonstrating that DER plays a crucial role in further improving the model's detection performance. These results indicate that DEM extracts and filters object-related discriminative evidence, while DER uses it to reconstruct hierarchical features, with both strategies jointly enhancing the model's ability to detect camouflaged objects.

**CDEM.** Tab. 3 presents the ablation results of different components in CDEM. By individually removing feature projection (FP), mask-guided foreground-background competition (MFC), and competition-guided evidence fusion (CEF), we analyze their impacts on CDEM. Removing any component results in performance degradation to varying degrees, confirming the effectiveness of each component. In particular, removing FP results in a performance degradation, indicating that FP plays a crucial role in cross-layer feature interactions within CDEM. After removing MFC, the model's detection performance degrades on multiple datasets, indicating that MFC effectively enhances CDEM's evidence mining capabilities. The performance degradation is most evident after removing CEF, demonstrating its importance in integrating discriminative evidence from different sources. The above results indicate that FP, MFC, and CEF together form an effective CDEM.

**DER.** Tab. 4 presents the ablation results of evidence reconstruction components at different levels in the DER strategy. By individually removing SER-3, SER-2, and FER, we analyze the impact of evidence reconstruction at different levels on model performance. After removing any reconstruction component, the model's performance degrades to varying degrees, indicating that evidence reconstruction at different levels positively contributes to detection performance. In particular, removing SER-3 or SER-2 results in performance degradation across multiple datasets, indicating that evidence reconstruction at higher levels plays a crucial role in enhancing the model's overall representation capacity. The performance degradation after removing FER indicates that shallow evidence reconstruction further enhances the utilization of high-resolution features. The full DMRNet achieves the best performance when combined with reconstruction components at different levels, validating the effectiveness of evidence reconstruction at different levels in the DER strategy.

## 5 Conclusion

To address the challenge of balancing model complexity and discriminative power in lightweight camouflaged object detection, we propose a Discriminative Evidence Mining and Reconstruction Network (DMRNet). Through the Discriminative Evidence Mining (DEM) strategy, DMRNet mines and integrates discriminative evidence from hierarchical features. Through the Discriminative Evidence Reconstruction (DER) strategy, DMRNet reconstructs hierarchical feature representations using reliable evidence, enhancing semantic representations at higher levels and fine-grained structural details at the shallow level. Experimental results on four benchmark datasets demonstrate that DMRNet achieves competitive detection performance with only 12.46M parameters and 9.15G FLOPs at 53.74 FPS, showing a favorable balance between accuracy and efficiency. Ablation experiments further validate the effectiveness of the proposed strategies and components. Future work will explore more efficient feature interactions to improve detection in complex camouflaged scenarios.

## ETHICS STATEMENT

This work follows the ICLR Code of Ethics. We use only publicly available camouflaged object detection datasets, including CAMO, CHAMELEON, COD10K, and NC4K, without involving human subjects, animal experiments, or personally identifiable information. Our research focuses on improving the efficiency and effectiveness of camouflaged object detection, and we carefully consider the responsible use of visual recognition technologies throughout the research process.

## REPRODUCIBILITY STATEMENT

To facilitate reproducibility, we provide detailed descriptions of the proposed method, training configurations, and evaluation protocols in the paper. All experiments are conducted on four publicly available camouflaged object detection benchmarks, including CAMO, CHAMELEON, COD10K, and NC4K. Additional prediction results on these datasets are provided in the supplementary material. The source code will be released after paper acceptance to support further verification and future research.

## LLM USAGE

Large Language Models (LLMs) were used solely to assist in language polishing and improving the readability of this manuscript. Specifically, LLMs were used to refine sentence structures, improve linguistic clarity, and enhance the overall presentation of the paper. The LLMs were not involved in the development of research ideas, methodology design, experimental implementation, or result analysis. All LLM-assisted text revisions were carefully reviewed, verified, and further modified by the authors rather than directly adopted without verification. The authors take full responsibility for the accuracy, originality, and integrity of the final manuscript.

## REFERENCES

Chunyuan Chen, Weiyun Liang, Mingyang Yu, Ji Du, Shujuan Li, and Jing Xu. Rethinking multi-scale perception for camouflaged object detection. In *ICASSP*, pp. 10132–10136, 2026.

Xuehan Chen, Guangyu Ren, Tianhong Dai, Tania Stathaki, and Hengyan Liu. Enhancing prompt generation with adaptive refinement for camouflaged object detection. In *ICCV*, pp. 20672–20682, 2025.

Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. Camouflaged object detection. In *CVPR*, pp. 2777–2787, 2020.

Yu'ang Feng, Shuyong Gao, Fuzhen Yan, Yicheng Song, Lingyi Hong, Junjie Hu, and Wenqiang Zhang. Scoring, remember, and reference: Catching camouflaged objects in videos. In *ICCV*, pp. 13043–13052, 2025.

Yanliang Ge, Yuxi Zhong, Junchao Ren, Min He, Hongbo Bi, and Qiao Zhang. Camouflaged object detection based on edge screening and cross-layer fusion. *Pattern Recognit.*, 179:113604, 2026.

Juwei Guan, Xiaolin Fang, Tongxin Zhu, and Weiqi Qian. Sdrnet: Camouflaged object detection with independent reconstruction of structure and detail. *Knowl.-Based Syst.*, 299:112051, 2024.

Jinyu Han, Changguang Wu, Fuming Sun, and Jinhui Tang. Beyond appearance: Camouflaged object detection via geometric structure. In *CVPR*, pp. 25830–25840, 2026.

Chao Hao, Zitong Yu, Xin Liu, Jun Xu, Huanjing Yue, and Jingyu Yang. A simple yet effective network based on vision transformer for camouflaged object and salient object detection. *IEEE Trans. Image Process.*, 34:608–622, 2025.

Chunming He, Kai Li, Yachao Zhang, Yulun Zhang, Zhenhua Guo, Xiu Li, Martin Danelljan, and Fisher Yu. Strategic preys make acute predators: Enhancing camouflaged object detectors by generating camouflaged objects. In *ICLR*, pp. 36657–36675, 2024.

Jiahao He, Keren Fu, Xiaohong Liu, and Qijun Zhao. Samba: A unified mamba-based framework for general salient object detection. In *CVPR*, pp. 25314–25324, 2025a.

Jinpeng He, Biyuan Liu, and Huaixin Chen. Hdpnet: Hourglass vision transformer with dual-path feature pyramid for camouflaged object detection. In *WACV*, pp. 8638–8647, 2025b.

Xihang Hu, Xiaoli Zhang, Fasheng Wang, Jing Sun, and Fuming Sun. Efficient camouflaged object detection network based on global localization perception and local guidance refinement. *IEEE Trans. Circuits Syst. Video Technol.*, 34(7):5452–5465, 2024.

Haolin Ji, Fengying Xie, Linpeng Pan, Yushan Zheng, and Zhenwei Shi. Huntnet: Homomorphic unified nexus topology for camouflaged object detection. *IEEE Trans. Image Process.*, 34:6068–6082, 2025.

Abbas Khan, Mustaqeem Khan, Wail Gueaieb, Abdulmotaleb El Saddik, Giulia De Masi, and Fakhri Karray. Camofocus: Enhancing camouflage object detection with split-feature focal modulation and context refinement. In *WACV*, pp. 1434–1443, 2024.

Orest Kupyn, Hirokatsu Kataoka, and Christian Rupprecht. S3od: Towards generalizable salient object detection with synthetic data. In *ICLR*, pp. 145157–145179, 2026.

Trung-Nghia Le, Tam V Nguyen, Zhongliang Nie, Minh-Triet Tran, and Akihiro Sugimoto. Anabranch network for camouflaged object segmentation. *Comput. Vis. Image Underst.*, 184: 45–56, 2019.

Mengqi Lei, Haochen Wu, Xinhua Lv, and Xin Wang. Condseg: A general medical image segmentation framework via contrast-driven feature enhancement. In *AAAI*, volume 39, pp. 4571–4579, 2025.

Guanyi Li, Junjie Zhang, Rui Gao, Wubang Yuan, Gloria Jin, and Dan Zeng. Learning compact representations with an information bottleneck for camouflaged object detection. *IEEE Trans. Multimedia*, 28:360–372, 2026.

Minghong Li, Yuqian Zhao, Fan Zhang, Gui Gui, Biao Luo, Chunhua Yang, Weihua Gui, and Kan Chang. Csfin: A lightweight network for camouflaged object detection via cross-stage feature interaction. *Expert Syst. Appl.*, 269:126451, 2025.

Weiyun Liang, Chunyuan Chen, Mingyang Yu, Ji Du, Shujuan Li, and Jing Xu. Multi-view frequency alignment and state space parameter fusion for lightweight camouflaged object detection. In *ICASSP*, pp. 9262–9266, 2026.

Weifeng Lin, Ziheng Wu, Jiayu Chen, Jun Huang, and Lianwen Jin. Scale-aware modulation meet transformer. In *ICCV*, pp. 6015–6026, 2023.

Yunqiu Lv, Jing Zhang, Yuchao Dai, Aixuan Li, Bowen Liu, Nick Barnes, and Deng-Ping Fan. Simultaneously localize, segment and rank the camouflaged objects. In *CVPR*, pp. 11591–11601, 2021.

Yixuan Lyu, Hong Zhang, Yan Li, Hanyang Liu, Yifan Yang, and Ding Yuan. Uedg: Uncertainty-edge dual guided camouflage object detection. *IEEE Trans. Multimedia*, 26:4050–4060, 2024.

Peng Ren, Tian Bai, and Fuming Sun. Esnet: An efficient skeleton-guided network for camouflaged object detection. *Knowl.-Based Syst.*, 311:113056, 2025.

Przemysław Skurowski, Hassan Abdulameer, Jakub Błaszczyk, Tomasz Depta, Adam Kornacki, and Przemysław Kozieł. Animal camouflage analysis: Chameleon database. *Unpublished manuscript*, 2(6):7, 2018.

Xiaogang Song, Pengfei Zhang, Xiaochang Li, Xinhong Hei, and Rongrong Liu. Btdgnet: A dual-guided camouflaged object detection network leveraging boundary and texture information. *IEEE Trans. Multimedia*, 27:9467–9477, 2025a.

Ze Song, Xudong Kang, Xiaohui Wei, Jinyang Liu, Zheng Lin, and Shutao Li. Continuous feature representation for camouflaged object detection. *IEEE Trans. Image Process.*, 34:5672–5685, 2025b.

Fuming Sun, Jinyu Han, Weiyi Wu, Jing Sun, Mengyin Wang, and Haojie Li. A unet-like transformer network for camouflaged object detection. *IEEE Trans. Multimedia*, 27:9267–9280, 2025a.

Ke Sun, Zhongxi Chen, Xianming Lin, Xiaoshuai Sun, Hong Liu, and Rongrong Ji. Conditional diffusion models for camouflaged and salient object detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 47(4):2833–2848, 2025b.

Yanguang Sun, Chunyan Xu, Jian Yang, Hanyu Xuan, and Lei Luo. Frequency-spatial entanglement learning for camouflaged object detection. In *ECCV*, pp. 343–360, 2024.

Jun Wang, Songxiao Geng, Jinjia Peng, Weibin Liu, and Shuaiqi Liu. Rdnet: Dynamic filtering guided transformer with cross-batch feature retention for camouflaged object detection. *Pattern Recognit.*, 180:114353, 2026a.

Kuan Wang, Xiuhong Li, Songlin Li, Yulong Bai, Boyuan Li, Mengge Lu, and Zhenhong Jia. Efficient camouflaged object detection network based on channel reconstruction and hybrid attention. In *ICMR*, pp. 2063–2067, 2025a.

Liqiong Wang, Jinyu Yang, Yanfu Zhang, Fangyi Wang, and Feng Zheng. Depth-aware concealed crop detection in dense agricultural scenes. In *CVPR*, pp. 17201–17211, 2024.

Tianyu Wang, Xinghua Xu, Changchong Sheng, Shaohua Qiu, Li Liu, Denghua Guo, and Xianqiang Bao. Fast diffusion-based camouflaged object detection via asynchronous denoising and linear attention. *IEEE Trans. Multimedia*, 28:6501–6516, 2026b.

Tingran Wang, Zaiyang Yu, Jianwei Fang, Jinlong Xie, Feng Yang, Huang Zhang, Liping Zhang, Minghua Du, Lusi Li, and Xin Ning. Multidimensional fusion of frequency and spatial domain information for enhanced camouflaged object detection. *Inform. Fusion*, 117:102871, 2025b.

Xiandong Wang, Fengqin Yao, Guoqiang Zhong, Qing Cai, Shengke Wang, and James Tin-Yau Kwok. Gbnet: Gated boundary-aware network for camouflaged object detection. *IEEE Trans. Image Process.*, 35:5297–5310, 2026c.

Dan Wu, Mengyin Wang, Jing Sun, and Xu Jia. Knowledge-guided and collaborative learning network for camouflaged object detection. *Eng. Appl. Artif. Intell.*, 153:110771, 2025a.

Fei Wu, Jun Yin, Xiaochuan Li, Jianfeng Wu, Da Jin, and Jiamin Yang. Conet: A consistency-oriented network for camouflaged object segmentation. *IEEE Trans. Circuits Syst. Video Technol.*, 35(1):287–299, 2025b.

He Xiao, Fugui Luo, Ziyang Liu, and Liping Deng. Afrnet: Adaptive frequency regulation network with camouflage spectral gap modeling for camouflaged object detection. *Neurocomputing*, 694: 133969, 2026.

Boran Yang, Min Zhang, Yong Wang, and Duoqian Miao. Multi-scale local–global fusion for camouflaged object detection. *IEEE Trans. Circuits Syst. Video Technol.*, 36(5):6390–6399, 2026.

Siyuan Yao, Hao Sun, Tian-Zhu Xiang, Xiao Wang, and Xiaochun Cao. Hierarchical graph interaction transformer with dynamic token clustering for camouflaged object detection. *IEEE Trans. Image Process.*, 33:5936–5948, 2024.

Qian Ye, Qingwu Li, Chang Xu, Xiao Li, and Chenkai Chang. Context-aware and edge-refinement network for camouflaged object detection. *Neurocomputing*, 701:134588, 2026.

Sheng Ye, Xin Chen, Yan Zhang, Xianming Lin, and Liujuan Cao. Escnet: Edge-semantic collaborative network for camouflaged object detection. In *ICCV*, pp. 20053–20063, 2025.

Bowen Yin, Xuying Zhang, Deng-Ping Fan, Shaohui Jiao, Ming-Ming Cheng, Luc Van Gool, and Qibin Hou. Camoformer: Masked separable attention for camouflaged object detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 46(12):10362–10374, 2024.

Xiangjun Yin and Huihui Yue. Reveal object in lensless photography via region gaze and amplification. In *ICLR*, pp. 41726–41743, 2025.

Guanghui Yue, Shangjie Wu, Tianwei Zhou, Gang Li, Jie Du, Yu Luo, and Qiuping Jiang. Progressive region-to-boundary exploration network for camouflaged object detection. *IEEE Trans. Multimedia*, 27:236–248, 2025.

Jinxia Zhang, Yin Yuan, Xuwen Zhu, Yang Hu, and Kaihua Zhang. Partitioned observation network for camouflaged object detection. *Pattern Recognit.*, 172:112637, 2026.

Shizhou Zhang, Dexuan Kong, Yinghui Xing, Yue Lu, Lingyan Ran, Guoqiang Liang, Hexu Wang, and Yanning Zhang. Frequency-guided spatial adaptation for camouflaged object detection. *IEEE Trans. Multimedia*, 27:72–83, 2025.

Rui Zhao, Mengyin Wang, Fasheng Wang, Fuming Sun, and Haojie Li. Spatial-frequency collaborative learning for camouflaged object detection. *IEEE Trans. Multimedia*, 27:7756–7768, 2025.

Zhangjun Zhou, Yiping Li, Chunlin Zhong, Jianuo Huang, Jialun Pei, Hua Li, and He Tang. Rethinking detecting salient and camouflaged objects in unconstrained scenes. In *ICCV*, pp. 22372–22382, 2025.

# A Appendix

## A.1 Motivation and Distinction of Camouflaged Object Detection

As illustrated in Fig. 4, generic, camouflaged, and salient objects exhibit substantially different visual characteristics. Generic objects usually contain recognizable appearance cues and clear foreground-background differences, while salient objects tend to attract visual attention due to prominent color, contrast, or structural cues. In contrast, camouflaged objects exhibit color, texture, and structural patterns highly similar to those of their surroundings, resulting in extremely weak foreground-background discrimination. Therefore, COD cannot be simply addressed by conventional object or salient object detection paradigms and requires dedicated modeling of subtle discriminative evidence between objects and their backgrounds. This distinction also motivates our DMRNet to focus on discriminative evidence mining and reconstruction under limited representation capacity.




Figure 4: Visual comparison among generic object detection (GOD), camouflaged object detection (COD), and salient object detection (SOD). Compared with generic and salient objects, camouflaged objects exhibit substantially weaker foreground-background differences in color, texture, and structure.

## A.2 Additional Qualitative Results on Camouflaged Object Detection

To complement the qualitative analysis in Section 4.2, we further present additional prediction results of DMRNet and existing COD methods under more challenging scenarios. As shown in Fig. 5, the selected samples cover various challenging camouflaged scenes, including highly similar appearances between objects and backgrounds, small object regions, complex boundary structures, and severe background texture interference. Compared with other methods, DMRNet can locate camouflaged objects more accurately and generate more complete prediction results.

Benefiting from the Discriminative Evidence Mining (DEM) strategy, DMRNet effectively mines object-related discriminative information from hierarchical features and reduces the interference caused by complex backgrounds, thereby maintaining stable detection performance under low-contrast conditions and strong background interference. Meanwhile, through the Discriminative Evidence Reconstruction (DER) strategy, DMRNet further utilizes reliable evidence to reconstruct hierarchical feature representations, enabling the model to better recover semantic regions and fine-grained structural information of objects, especially for small objects and those with complex structures or boundaries.

Overall, the additional qualitative results demonstrate the effectiveness of DMRNet in complex camouflaged scenarios. These results indicate that DMRNet not only maintains competitive detection accuracy but also generates more complete and accurate camouflaged object segmentation results.

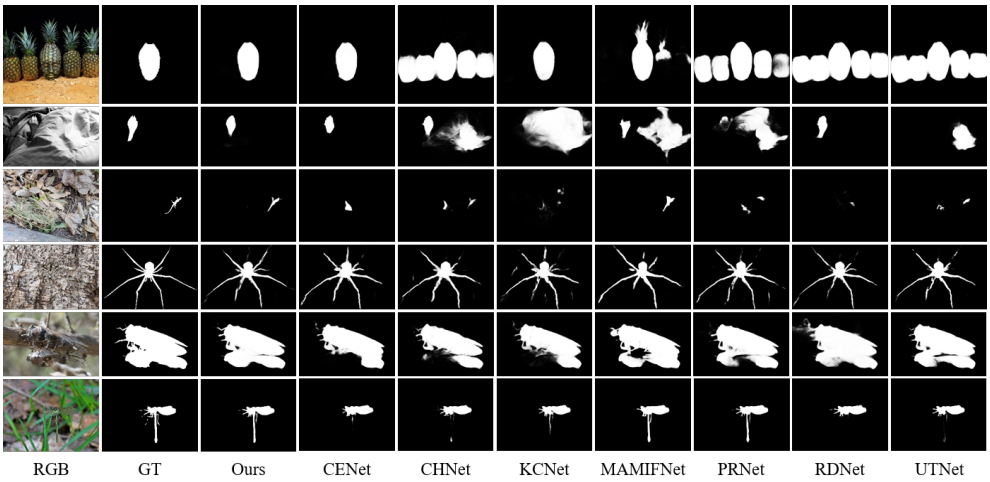



Figure 5: Additional qualitative comparison of DMRNet and COD methods on challenging scenes.

### A.3 Efficiency and Performance Analysis of Main Components

To further analyze the contribution of each main component in terms of performance improvement and computational cost, Tab. 5 reports the FLOPs, parameters, and average performance across the four benchmark datasets under different configurations. The backbone-only model achieves limited detection performance. After introducing CDEM, the FLOPs increase from 7.07 G to 7.94 G and the parameters increase from 11.13 M to 11.70 M, while the average $S_m$, $F_\beta^w$, and $E_m$ improve significantly from 0.578, 0.265, and 0.548 to 0.886, 0.832, and 0.937, respectively. Meanwhile, the average $M$ decreases substantially from 0.191 to 0.031, demonstrating that cross-layer discriminative evidence mining effectively enhances feature representation with limited additional computational cost.

By further incorporating BIS, only a small increase in complexity is introduced, with 0.16 G additional FLOPs and 0.17 M additional parameters, while the overall performance continues to improve. This indicates that background interference suppression can effectively improve the reliability of discriminative evidence by reducing misleading background responses. Finally, after integrating DER, DMRNet achieves the best performance across all four average metrics, while maintaining low complexity with 9.15 G FLOPs and 12.46 M parameters. These results demonstrate that CDEM, BIS, and DER provide complementary improvements at different stages, enabling DMRNet to achieve an effective balance between detection accuracy and computational efficiency.

Table 5: Efficiency and average performance analysis of the main components in DMRNet. The reported metrics are averaged over CAMO, CHAMELEON, COD10K, and NC4K. a$S_m$, a$F_\beta^w$, a$M$, and a$E_m$ denote the average values of the corresponding metrics. The best results are highlighted in red.

| Components | | | | Complexity | | Average Performance | | | |
|---|---|---|---|---|---|---|---|---|---|
| Backbone | CDEM | BIS | DER | FLOPs (G) | Params (M) | a$S_m$ ↑ | a$F_\beta^w$ ↑ | a$M$ ↓ | a$E_m$ ↑ |
| ✓ | | | | 7.07 | 11.13 | 0.578 | 0.265 | 0.191 | 0.548 |
| ✓ | ✓ | | | 7.94 | 11.70 | 0.886 | 0.832 | 0.031 | 0.937 |
| ✓ | ✓ | ✓ | | 8.10 | 11.87 | 0.888 | 0.835 | 0.031 | 0.939 |
| ✓ | ✓ | ✓ | ✓ | 9.15 | 12.46 | 0.892 | 0.842 | 0.029 | 0.942 |

### A.4 Qualitative Analysis of Main Component Ablations

As shown in Fig. 6, we further analyze the effects of CDEM, BIS, and DER through qualitative visualization. With CDEM alone, the model can already localize the main object regions, but incomplete predictions, blurred boundaries, and false background responses still occur in scenes with

high foreground-background similarity, multiple objects, or complex structures. In the first row, CDEM captures only part of the object region, while incorporating BIS noticeably improves prediction completeness. In the multi-object scene in the second row, CDEM produces an obvious false response between the two objects. After introducing BIS, this false activation is effectively suppressed, indicating that BIS improves the reliability of discriminative evidence by reducing background interference. For objects with elongated structures or complex local details in the third and fifth rows, progressively incorporating BIS and DER produces more continuous predictions and preserves fine-grained structures more effectively. In the low-contrast scene shown in the fourth row, the full DMRNet also generates a clearer object contour that is more consistent with the ground truth. Overall, these qualitative results are consistent with the quantitative ablation results. CDEM provides effective discriminative evidence for object localization. BIS further suppresses misleading background responses. DER improves hierarchical feature representations, leading to more complete object regions and more accurate fine-grained structures.

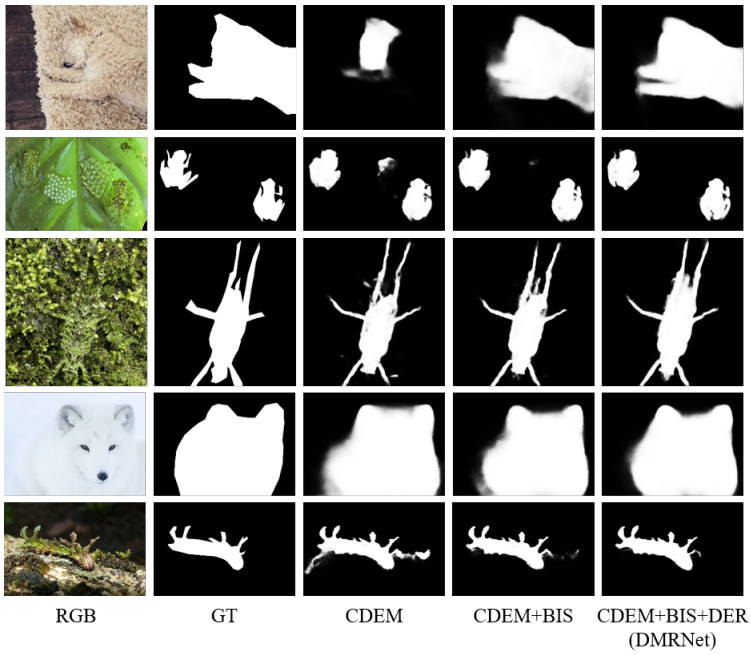



Figure 6: Qualitative visualization of the main component ablations. From left to right: RGB images, ground truths, predictions with CDEM, predictions with CDEM+BIS, and predictions with the full DMRNet (CDEM+BIS+DER).

### A.5 Sensitivity Analysis of the Temperature Coefficient

To investigate the sensitivity to the temperature coefficient $\tau$ in Eq. (7), we evaluate different settings of $\tau$ ranging from 1 to 6 and report the average performance over the four benchmark datasets. As shown in Fig. 7, the overall performance improves as $\tau$ increases from 1 to 4, and the best results are achieved at $\tau = 4$, with average $S_m$, $F_\beta^w$, $M$, and $E_m$ values of 0.892, 0.842, 0.029, and 0.942, respectively. This demonstrates that an appropriate temperature coefficient helps regulate the strength of foreground-background competition, allowing the model to better distinguish object-related evidence from background interference. When $\tau$ further increases to 5 and 6, the performance slightly decreases, indicating that an overly large temperature coefficient may produce overly strong competition signals and degrade the stability of evidence mining. Therefore, the default setting $\tau = 4$ also shows a favorable balance between discriminative capability and stability.

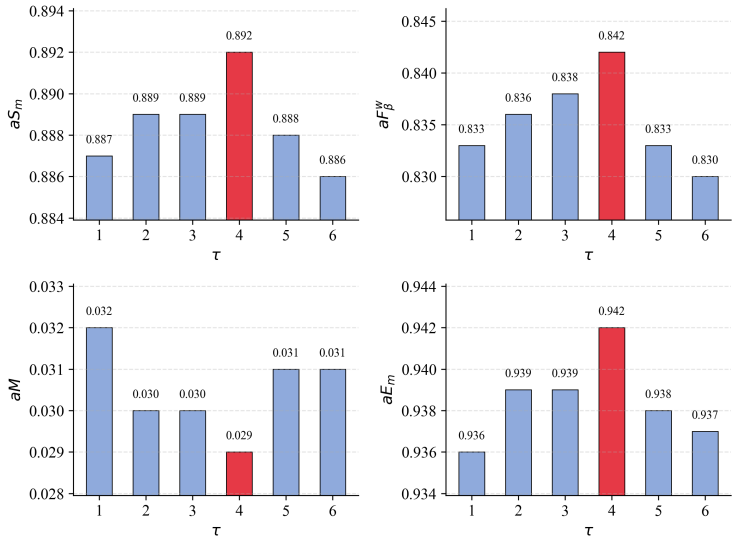



Figure 7: Sensitivity analysis of the temperature coefficient $\tau$. The average performance is calculated over four benchmark datasets. The results demonstrate the influence of different $\tau$ values on $aS_m$, $aF_\beta^w$, $aM$, and $aE_m$.

### A.6 Visualization of Discriminative Evidence Mining (DEM)

As shown in Fig. 8, we visualize the intermediate representations in DEM at the second decoding level. The coarse prediction $P_3$ already provides an approximate localization of the camouflaged objects, but its boundaries and fine structures remain relatively blurred. After CDEM, $E_2$ produces evident responses over the object regions. Compared with $E_2$, $B_2$ exhibits fewer and less prominent low-response regions and more concentrated high-response regions, as reflected by the reduced purple responses and more prominent yellow responses, indicating that BIS further enhances the effective discriminative evidence while suppressing less informative responses. Meanwhile, the predictions progressively become clearer from $P_3$ to $P_2$ and finally to $P_1$, with increasingly complete object regions and more accurate boundary and structural details. These visualizations further demonstrate the effectiveness of DEM in enhancing discriminative representations during progressive decoding.

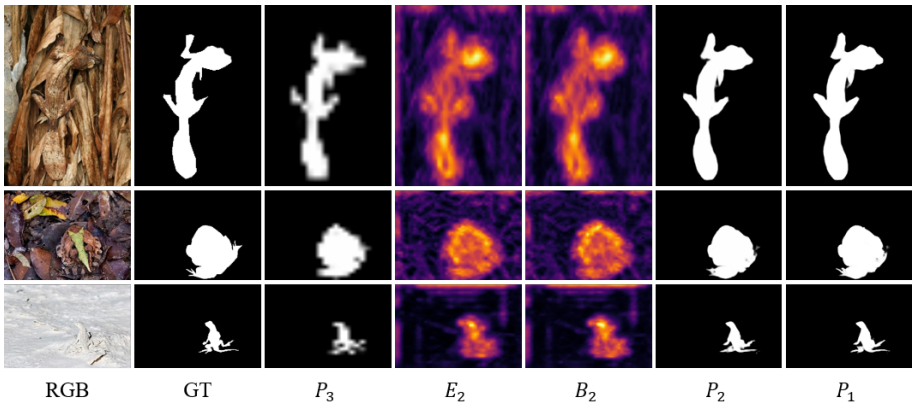



Figure 8: Visualization of feature representations from DEM and progressive predictions. From left to right: RGB images, ground truths, the coarse prediction $P_3$, CDEM output $E_2$, BIS output $B_2$, prediction $P_2$, and final prediction $P_1$. Warmer colors in $E_2$ and $B_2$ indicate stronger feature responses.

### A.7 VISUALIZATION OF DISCRIMINATIVE EVIDENCE RECONSTRUCTION (DER)

As shown in Fig. 9, $X_3$, $X_2$, and $X_1$, produced by SER-3, SER-2, and FER, respectively, consistently focus on the camouflaged object regions, indicating that the reconstructed features at different decoding levels can accurately localize the objects. Meanwhile, the hierarchical representations are progressively enhanced from $X_3$ to $X_2$ and finally to $X_1$. Specifically, $X_3$ provides reliable high-level object localization, $X_2$ further refines the object representation, and $X_1$ preserves finer spatial details and produces clearer object-aware responses. The final prediction $P_1$ consequently exhibits more complete object regions and more accurate structural details. These results demonstrate that DER progressively reconstructs hierarchical representations while maintaining accurate object localization.

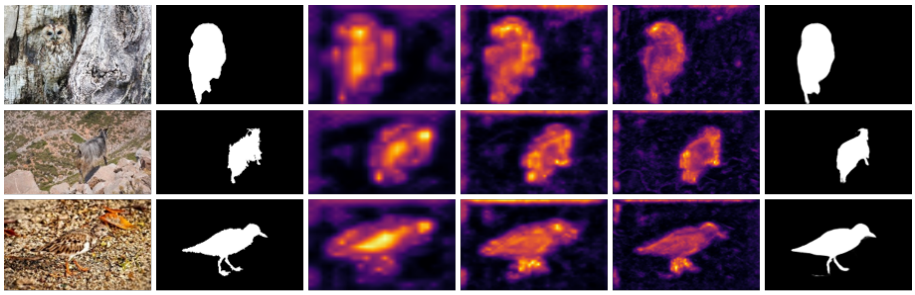



Figure 9: Visualization of feature representations from DER and the final prediction. From left to right: RGB images, ground truths, $X_3$ produced by SER-3, $X_2$ produced by SER-2, $X_1$ produced by FER, and the final prediction $P_1$. Warmer colors indicate stronger feature responses.