Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2026.075593

ARTICLE

Abel-Net: Aggregate Bilateral Edge Localization Network for Multi-Task Binary Segmentation

Zhengyu Wu[1], **Kejun Kang**[2], **Yixiu Liu**[3,*] **and Chenpu Li**[3]

[1]School of Automation, Hangzhou Dianzi University, Hangzhou, 310018, China
[2]Zhuoyue Honors College, Hangzhou Dianzi University, Hangzhou, 310018, China
[3]School of Cyberspace, Hangzhou Dianzi University, Hangzhou, 310018, China

*Corresponding Author: Yixiu Liu. Email: liuyixiu@hdu.edu.cn

Received: 04 November 2025; Accepted: 22 December 2025; Published: 12 March 2026

ABSTRACT: Binary segmentation tasks in computer vision exhibit diverse appearance distributions and complex boundary characteristics. To address the limited generalization and adaptability of existing models across heterogeneous tasks, we propose Abel-Net, an Aggregated Bilateral Edge Localization Network designed as a universal framework for multi-task binary segmentation. Abel-Net integrates global and local contextual cues to enhance feature learning and edge precision. Specifically, a multi-scale feature pyramid fusion strategy is implemented via an Aggregated Skip Connection (ASC) module to strengthen feature adaptability, while the Edge Dual Localization (EDL) mechanism performs coarse-to-fine refinement through edge-aware supervision. Additionally, Edge Attention (EA) and Edge Fusion Attention (EFA) modules prioritize edge-critical regions and facilitate accurate boundary alignment. Extensive experiments on nine diverse binary segmentation tasks demonstrate that Abel-Net performs comparably to or surpasses state-of-the-art task-specific networks, exhibiting strong adaptability to a wide range of visual perception challenges.

KEYWORDS: Computer vision; binary segmentation; edge dual localization; attention mechanism

1 Introduction

As a fundamental component of modern computer vision, image segmentation [1] provides the mathematical foundation for high-level visual perception [2] through pixel-wise classification and structural understanding. Among various segmentation paradigms, binary segmentation [3] plays a crucial role by simplifying complex visual interpretation tasks into foreground–background separation, thereby achieving an effective balance between accuracy and computational efficiency. This property has made binary segmentation a key technique not only in industrial inspection [4] but also in a wide range of vision-based applications such as medical imaging, remote sensing, and autonomous perception systems.

With the rapid development of deep learning (DL) and multimodal perception, visual understanding tasks are evolving toward increasingly diverse and complex environments. This trend has posed significant challenges for binary segmentation models in adapting to heterogeneous data domains and maintaining precise boundary localization. For instance, existing frameworks for camouflaged object detection (COD) often struggle to distinguish concealed targets from visually similar backgrounds, leading to suboptimal detection accuracy. Similarly, salient object detection (SOD) [5] remains challenging under conditions of multiple overlapping targets or strong background interference. In remote sensing image salient object detection (RSISOD), performance degradation frequently occurs due to complex environmental noise and multidirectional variations of targets.

In other specialized domains, such as optical imaging and medical analysis, similar challenges persist. Shadows often introduce pseudo-edge artifacts that hinder precise measurement, emphasizing the need for robust shadow detection algorithms [6]. Defocus blur detection remains an open problem where accurate delineation of blurred regions is difficult to achieve. Likewise, detecting transparent and specular materials introduces substantial complexity due to their reflective and refractive properties, which distort spatial and photometric information. In medical image segmentation, although notable progress has been made, more intelligent and fine-grained models are still required to accurately delineate anatomical structures and pathological regions. Collectively, these challenges highlight the limitations of existing binary segmentation frameworks and underscore the urgent demand for unified, generalizable, and edge-aware architectures to improve precision and cross-domain adaptability across diverse visual tasks.

Current research indicates that despite the remarkable progress of DL in specific segmentation tasks, existing methods generally suffer from inadequate cross-domain generalization capabilities. In this paper, we propose a universal network for multiple binary segmentation tasks, as illustrated in Fig. 1, which effectively handles a binary segmentation task set containing nine typical scenarios.

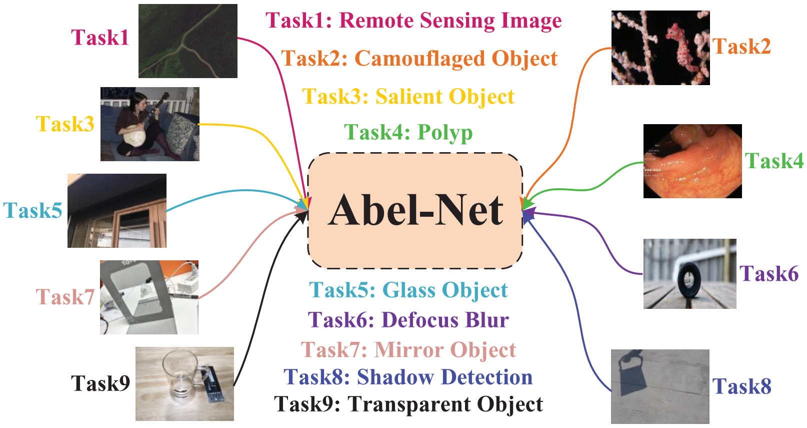


Figure 1: All binary segmentation tasks processed by Abel-Net in this paper

Firstly, following common practice, we adopt the U-Net [7] architecture as the baseline framework to leverage its feature pyramid structure, constructing aggregated skip connections. However, distinct from traditional encoder-decoder architectures like U-Net that directly concatenate shallow features through skip connections, our work introduces a multi-scale feature pyramid fusion strategy, which is materialized via the proposed Aggregated Skip Connection (ASC). This approach extracts multi-depth features through four convolutional layers with different parameters while dynamically aggregating features from the current decoder layer, upper decoder layers, and pyramid outputs. Compared with the DeepLab [8] series, which relies on dilated convolutions to expand receptive fields or multi-scale pooling in ASPP modules, our method enhances cross-domain feature representation robustness while reducing information loss through complementary inter-layer feature fusion.

Secondly, we design an Edge Dual Localization (EDL) module. Specifically, CBR^{+} and CBR^{-}, with identical structures but independent parameters, generate dilated and eroded edge probability maps, respectively. Their subtraction yields a more accurate edge probability map, effectively suppressing background noise and highlighting object boundaries. This design enables robust, fine-grained edge localization, thereby enhancing the segmentation performance across diverse tasks. We introduce the Edge Dual Localization loss to effectively supervise predicted edges at each stage, progressively aligning predicted edges with ground truth image boundaries.

Finally, we propose Edge Attention (EA) and Edge Fusion Attention (EFA) modules. Unlike traditional attention mechanisms that focus on channel or spatial-wise feature recalibration without explicit edge prior integration, our EA innovatively multiplies predicted edge maps with attention key-value matrices in an element-wise manner. The EFA module further achieves dynamic fusion between edge features and contextual semantics through edge-guided local attention focusing. Compared with self-attention in Transformers, EFA enhances the capability of capturing subtle edge structures while reducing computational complexity.

To validate the generalizability of our network, we conduct evaluations across 22 datasets covering mainstream tasks (salient and camouflaged object detection) and specialized applications (transparent, glass, and specular object detection). Experimental results demonstrate that our universal network outperforms most single-task methods in diverse binary segmentation scenarios.

The principal contributions of this work are summarized as follows:

•We propose an aggregated bilateral edge localization network (Abel-Net) that demonstrates effective performance across a wide range of binary segmentation tasks, addressing the common limitation of conventional binary segmentation networks' lack of generalizability.

•We propose a multi-scale feature pyramid fusion strategy implemented via a novel ASC. This module aggregates features from different depths of the image, allowing the decoder to access richer feature representations.

•To achieve precise edge localization, we introduce the EDL module and EDL loss. The EDL module effectively enhances edge identification through dual localization, while the EDL loss ensures that the predicted edges progressively align with the ground truth edges.

•We embed the predicted edge information in the attention mechanism through our proposed EA and EFA modules, allowing the network to focus more on the edges of the image and achieve edge exposure.

•Extensive experiments on 22 datasets demonstrate the quantitative and qualitative superiority of our method over state-of-the-art approaches in nine binary segmentation tasks.

The remainder of this paper is organized as follows: Section 2 reviews related work in binary segmentation and multi-task learning. Section 3 details the proposed Abel-Net architecture, including the aggregated skip connection, edge dual localization, and attention mechanisms. Section 4 presents the experimental setup, quantitative and qualitative comparisons with state-of-the-art methods, and ablation studies. Finally, Section 5 concludes the paper, discussing the limitations and future scope of this work.

2 Related Work

Recent research has explored a variety of binary segmentation tasks under a unified DL framework. Multi-task binary segmentation approaches, such as GateNet [9] and EVP [10], aim to share representations across multiple domains, balancing multi-level features and achieving efficient adaptation without architectural redesign. These studies inspired our overall design, motivating multi-scale feature fusion and lightweight attention mechanisms (EA, EFA) to improve generalization while maintaining model compactness.

Salient object detection (SOD) and its variants, including camouflaged object detection (COD), shadow detection, and transparent or glass object detection, all focus on segmenting visually challenging regions that differ from the background in distinct ways. Modern methods leverage feature aggregation modules and attention-based decoders to enhance localization accuracy. For instance, ICON [5] and EDN [11] improve feature refinement, while COD-related works such as DGNet [12] and HitNet [13] emphasize multi-scale feedback and texture-context decoupling.

Polyp segmentation and remote sensing salient object detection (RSISOD) extend binary segmentation to medical and aerial imagery domains, respectively. Polyp segmentation methods such as PraNet [14] and MSNet [15] focus on hierarchical multi-scale feature extraction, while RSISOD models like MCCNet [16] and SeaNet [17] address cross-sensor feature fusion and complex scene understanding through attention-based modules.

Finally, defocus blur detection [18,19] targets blurred region segmentation, where high variability in focal distances increases task difficulty. Recent methods integrate context- and edge-aware representations to enhance blur localization and image restoration performance.

In summary, while the aforementioned works have established strong baselines, clear distinctions exist between them and our proposed Abel-Net. First, compared to task-specific models like PraNet [14] or SINet [20], which are tailored to specific visual characteristics, Abel-Net is designed as a universal framework capable of generalizing across nine heterogeneous tasks without architectural redesign. Second, unlike existing multi-task approaches such as GateNet [9] and EVP [10] that primarily rely on implicit feature sharing or visual prompting, Abel-Net introduces an explicit Edge Dual Localization (EDL) mechanism. This allows for coarse-to-fine boundary refinement, directly addressing the challenge of edge ambiguity common in binary segmentation. Finally, distinct from standard multi-scale fusion methods used in general segmentation, our Aggregated Skip Connection (ASC) implements a specialized pyramid fusion strategy that preserves rich semantic details across depths, ensuring robust performance even in complex scenarios like transparent or shadow object detection.

3 Methodology

3.1 Overview

In this section, we provide a detailed introduction to the aggregate bilateral edge localization network (Abel-Net), including ASC, EDL, EA, and EFA. As shown in Fig. 2, in the Abel-Net, we adopt the U-Net architecture to build the network because it provides a symmetric encoder–decoder structure with skip connections, which enables the direct transfer of spatially detailed features from the encoder to the decoder. Moreover, U-Net's hierarchical feature fusion is highly compatible with multi-scale backbone networks such as the second version of the Pyramid Vision Transformer (PVTv2) [21], allowing semantic-rich deep features to be seamlessly integrated with high-resolution shallow features. Its modular design also facilitates the incorporation of advanced components, such as the ASC, EDL, and attention mechanisms, making it a flexible and robust foundation for multi-task binary segmentation. We utilize the PVTv2 as the encoder backbone because its pyramid structure enables efficient multi-scale feature extraction with progressively reduced resolution, which is crucial for handling objects of varying sizes across different tasks. Additionally, PVTv2 introduces a spatial-reduction attention mechanism, which reduces computational complexity while maintaining the ability to capture long-range dependencies, making it suitable for high-resolution inputs common in binary segmentation tasks. And we choose the Swin Transformer block [22] as the decoder backbone due to its shifted window attention mechanism, which balances local feature refinement and global context aggregation. This design not only allows the decoder to recover fine details such as object boundaries but also preserves semantic consistency across scales. Compared with traditional CNN-based decoders, Swin Transformer blocks offer superior adaptability to diverse binary segmentation scenarios by better modeling complex spatial relationships and reducing overfitting when generalizing to unseen tasks. In addition to enhancing deep features, the encoder also utilizes a feature pyramid network (FPN). The collaborative use of edge attention and edge fusion attention enables effective exposure of objects, and EDL serves as input to the decoder to improve the accuracy of edge localization.

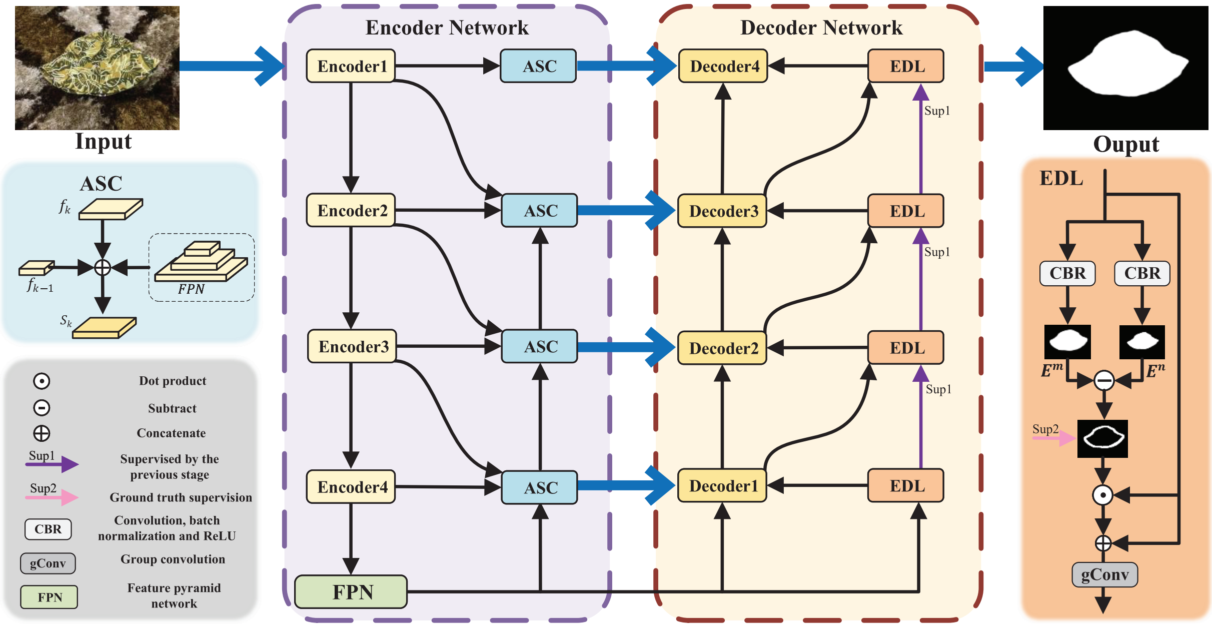


Figure 2: Overall architecture of the Abel-Net. The aggregation skip connection (ASC) aggregates features from multiple depths of the encoder and the FPN, and then delivers multi-scale contextual information to the decoder. The EDL module ensures that the predicted edges become more refined with each iteration

3.2 Aggregation Skip Connection

Aggregation skip connection based on AIMs and the Swin Transformer, combined with the integration of FPN, has significantly enhanced Abel-Net's ability to perform segmentation tasks.

FPN incorporates multiple levels of detailed information, allowing the network to effectively handle features of varying depths. Furthermore, this cross-level feature fusion facilitates the extraction of richer and more semantically representative features, thereby improving the accuracy and robustness of object detection. Given that the backbone follows a 4-level encoder-decoder structure, the number of levels in this paper is also set to 4.

Assuming the current encoder output is f^k, the module composed of the FPN is set as the fifth level, denoted as $f_5 = \text{FPN}(f_2, f_3, f_4)$, where f_2, f_3, and f_4 represent the output of the second, third, and fourth encoder levels, respectively. The output of the first encoder level is excluded from the feature pyramid network due to its high memory consumption. In the feature extraction network, the output of each stage has its own advantages. Therefore, we fuse the outputs from the current and preceding encoder layers, along with the output from the FPN, to form the output of a given stage, defined as:

$$S_k = \begin{cases} f_k \oplus f_{k-1} \oplus f_5 & (k = 2, 3, 4) \\ f_1 & (k = 1) \end{cases}, \tag{1}$$

where $\oplus$ denotes the concatenate operation, S_k represents the output of the encoder at the k-th stage. By combining the outputs of the higher-level encoders and the FPN, the data flow input into the decoder retains fine details from the upper layers while addressing issues of scale invariance and feature information loss during the encoding process. In our proposed ASC, each decoder layer performs a self-interaction operation, where the output from the previous stage's decoder is used as part of the input for the current stage's decoder, defined as:

$$D_k = \begin{cases} S_{5-k} \cdot \mathrm{gConv}[\mathrm{EA}(D_{k-1}), \mathrm{EFA}(D_{k-1}, E)] + S_{5-k} & \text{if } k = 2, 3, 4 \\ S_{5-k} \cdot \mathrm{gConv}[\mathrm{EA}(f_5), \mathrm{EFA}(f_5, E)] + S_{5-k} & \text{if } k = 1 \end{cases}. \tag{2}$$

Here, D_k represents the output of the current decoder, $gConv$ denotes group convolution, and E is the predicted edge feature map, which is the output of the EDL discussed in Section 3.3. EA and EFA are the predicted edge attention modules, which will be described in detail in Section 3.4. Through the self-interaction module, the decoder obtains stage outputs from the encoder and then connects the output features of EA and EFA using group convolution.

Finally, the output feature D_k of the current decoder layer is obtained through element-wise multiplication and addition with the outputs of EA and EFA and S_{5-k}. In this process, each decoder can perform multi-level interactions with the previous decoder layer, which helps capture long-range dependencies and enhances the representational capacity of the network. Additionally, the input to the decoder can include more semantic information, thereby improving the generalization ability of the network. The ASC module supplies the decoder with cross-layer, multi-scale features and integrates rich contextual information to enhance feature representation. Combining local gradient variations with global semantic cues helps reduce false edge generation. Moreover, when the fused features from the ASC module are fed into the EA and EFA modules, they further strengthen edge attention, suppress background interference, and provide more stable inputs for subsequent dual edge localization.

3.3 Edge Dual Localization

It is challenging to accurately detect edges with a single localization. Therefore, we perform two convolutions to capture more edge information and obtain finer edges. The first convolution produces dilated image edges, and the second convolution yields eroded image edges. By subtracting the results, we effectively enhance the feature details and eliminate background noise, resulting in more precise edge detection.

In the EDL module, CBR^{+} and CBR^{-} are two convolutional blocks with the same structure (Convolution, Batch Normalization, and ReLU) but with independent parameters. This design allows them to learn distinct transformations: CBR^{+} focuses on generating the dilated edge probability map, while CBR^{-} produces the eroded edge probability map. By subtracting the two results, the module effectively highlights edge details and suppresses background noise. In one stage, through CBR^{+} and CBR^{-}, we obtain the dilated edge probability map E^{m} and the eroded edge probability map E^{n}. We subtract E^{n} from E^{m} to obtain the edge probability map E^{p} for each stage. The edge probability map represents the approximate edges of objects within the image. Using E^{p}, we can further derive the image's edge mask e. The process of obtaining the edge mask is described in Eq. (3):

$$e = (E^{m} - E^{n}) \odot d_k, \tag{3}$$

where d_k represents the output of the decoder from the previous stage, and $\odot$ denotes the element-wise multiplication. The edge mask effectively extracts edge information from the image, highlighting the edge features of objects. With the edge mask, edge localization becomes more precise. By concatenating the edge mask with the decoder output from the previous stage and then applying group convolutions, we can obtain the predicted edges E^f for the current stage. The detailed operation is outlined in Eq. (4):

$$E^f = \mathrm{Conv}[\mathrm{gConv}(e \oplus d_k \oplus (C_1(d_k) - C_2(d_k)))], \tag{4}$$

where C_1 and C_2 represent two different convolutions. The predicted edges are fed into the current stage's decoder as input for the predicted edge attention mechanism.

The fine-grained edge mask generated by the EDL module is not only utilized for optimizing the edge branch but is also fused with the multi-scale features output by the ASC module during the decoding stage. Specifically, the edge mask serves as a guidance signal that, together with the cross-layer features from ASC, jointly contributes to the EA and EFA modules, thereby enhancing the decoder's sensitivity to object boundaries. Throughout the multi-stage processing, the high-level semantics provided by ASC help suppress edge noise, while the fine-grained edges from EDL compensate for ASC's limitations in capturing detailed structures. This complementary interaction between the two modules improves overall segmentation accuracy and enhances cross-task generalization capability.

3.4 Predicted Edge Attention

To enhance the network's attention to image edges and thereby improve edge recognition, we propose predicted edge attention based on the self-attention mechanism, which consists of edge attention (EA) and edge fusion attention (EFA). As shown in Fig. 3, EA takes the output $D_{(k-1)}$ of the previous decoder as input X, and transforms it through weight matrices W_Q, W_K, W_V, according to Eq. (5) to obtain the query matrix Q_1, the key matrix K_1, and the value matrix V_1.

$$[Q_1, K_1, V_1] = X[W_Q, W_K, W_V]. \tag{5}$$

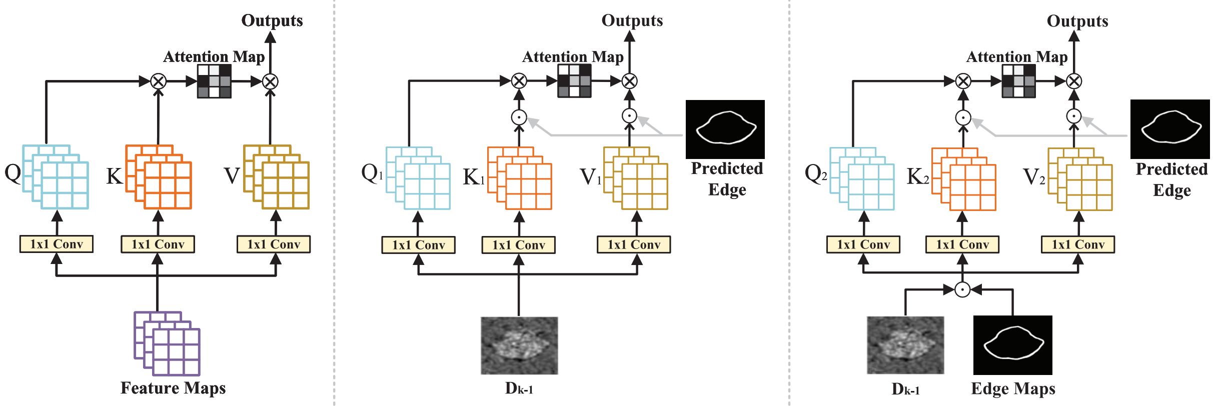


Figure 3: Comparison of our network's attention modules with a traditional attention module. Left: traditional self-attention module. Middle: edge attention (EA). Right: edge fusion attention (EFA)

By element-wise multiplying the predicted edge E^f with the key matrix K_1 and the value matrix V_1, the edge information is incorporated into the attention module, thereby enhancing the edge features. Then, the attention weights are obtained by normalizing the similarity between the query and the key, which can be represented as follows:

$$\mathrm{EA} = \mathrm{Softmax}\left(\frac{Q_1(K_1 \cdot E^f)^T}{\sqrt{d_m}}\right) V_1 \cdot E^f, \tag{6}$$

where $\frac{1}{\sqrt{d_m}}$ is a scaling factor used to alleviate the vanishing gradient problem caused by excessively large dot product values. Similarly, EFA is a module designed based on cross-attention. Unlike EA, EFA takes the output features of the previous decoder and the predicted edge feature map E as input. By integrating the edge feature map E, the attention module can more accurately expose the edges. The details are as follows:

$$[Q_2, K_2, V_2] = (X \cdot E)[W_Q, W_K, W_V]. \tag{7}$$

Finally, the attention weights can be represented as:

$$\mathrm{EFA} = \mathrm{Softmax}\left(\frac{Q_2(K_2 \cdot E^f)^T}{\sqrt{d_m}}\right)V_2 \cdot E^f. \tag{8}$$

Unlike conventional self-attention modules that compute the attention weights purely within the feature space of the input, the proposed edge fusion attention introduces an explicit integration of predicted edge maps into the attention process. Specifically, EFA differs from standard self-attention in two key aspects:

(1) Cross-attention with explicit edge priors: Instead of computing the query, key, and value solely from the same feature representation, EFA employs the predicted edge map as an additional prior, ensuring that edge cues are explicitly embedded into the attention weights. This cross-modality design allows EFA to guide the network towards more edge-sensitive feature refinement.

(2) Enhanced edge localization: While conventional self-attention emphasizes global contextual dependencies, EFA emphasizes spatial alignment between decoder features and edge maps. This design leads to more accurate boundary localization, which is particularly beneficial for tasks such as camouflaged object detection, where edges are often weak and difficult to capture. Therefore, the distinctive role of EFA is to combine the advantages of self-attention with edge-aware priors, resulting in improved edge representation and sharper object boundaries.

Through EA and EFA, our network can effectively improve the accuracy of identifying the edges of camouflaged objects.

3.5 Loss Function

The loss function includes L_{Edl} and L_{Tr}. The loss of edge dual localization L_{Edl} is supervised by the ground truth edges and composed of two parts. The first part of the loss, denoted L_s, ensures that the edges in each stage become progressively finer compared to the previous stage.

$$L_s = \sum_{k=2}^{4}\sum_{i=1}^{N}\left(\frac{[0, (E^p_{(k-1,N)} - E^p_{(k,N)})]^2}{w \times h} + L_{\mathrm{PPA}}(E^p_{(i,N)}, y^p_k)\right), \tag{9}$$

where N represents the total number of pixels in the image, $E^p_{(k,N)}$ denotes the pixel values in the edge probability map at stage k, and $[\cdot]$ indicates the minimum operation. y^p_k is the ground truth edge map after scaling, and PPA stands for Pixel Position-Aware loss [23]. The first part of L_s penalizes the portions of the current stage's edges that exceed those of the previous stage, ensuring that each stage's edge probability map is progressively finer than the previous one. To ensure that the original shape is preserved while shrinking the edges, we introduce the Pixel Position-Aware loss, which better captures the structural information contained within the features and helps the network focus more on detail-rich regions. To enhance the accuracy of PPA calculations, we manually adjust the range of the ground truth edges, allowing for varying degrees of edge refinement at different stages. In other words, we can determine the extent to which each stage's edge probability map is scaled.

The second part of the loss ensures that the edge probability map includes the ground truth edges, which we define as L_e.

$$L_e = \sum_{k=2}^{4} \sum_{i=1}^{N} \left(\frac{[0, (E^p_{(k,N)} - y^p)]^2}{w \times h} + L_{\text{BCE}}(E^f, y^p) \right), \tag{10}$$

where y^p is the ground truth edge map, $L_{BCE}(\cdot)$ represents the Binary Cross Entropy Loss, which is calculated as follows:

$$L_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^{N} \left[y^p \log(E^f) + (1 - y^p) \log(1 - E^f) \right], \tag{11}$$

where N represents the total number of pixels in the image. L_{BCE} is commonly used to measure the difference between the predicted probability distribution of the model and the ground truth. In the first part of L_e, by subtracting the ground truth edges from the edge probability map, we ensure that the shrinkage of the edge probability map does not fall below that of the ground truth map. L_{BCE} is then applied to further enhance the precision of edge localization. Finally, L_{Edl} is defined as:

$$L_{\text{Edl}} = L_s + \lambda L_e, \tag{12}$$

where λ is the balancing parameter.

L_{Tr} applies PPA to the decoder's output and the ground truth image, as follows:

$$L_{\text{Tr}} = L_{\text{PPA}}(y, D_k), \tag{13}$$

where y represents the ground truth image, and D_k represents the output of the decoder at this stage. Finally, the total loss is defined as:

$$L_{\text{total}} = L_{\text{Tr}} + L_{\text{Edl}}. \tag{14}$$

4 Experiments

4.1 Segmentation Tasks and Experimental Details

4.1.1 Segmentation Tasks

We evaluate our model on multiple datasets across nine different tasks. Specifically, for camouflaged object detection, we adopt CAMO, COD10K, and NC4K, which are widely used benchmarks covering diverse scenarios of camouflaged targets. For salient object detection, we employ OMRON, DUTS-TE, HKU-IS, PASCAL-S, and ECSSD, which provide large-scale and diverse images with complex backgrounds for robust evaluation. For defocus blur detection, we use CUHK and DUT, two standard datasets that cover typical defocus blur patterns. For remote sensing image salient object detection, we select EORSSD, a high-quality dataset capturing remote sensing scenarios. For transparent object detection, we adopt Transparent-easy and Transparent-hard, which represent different levels of difficulty in transparent object recognition. For glass object detection, we use GDD, the most widely adopted benchmark for this task. For mirror object detection, we use MSD, a challenging dataset containing mirror objects in complex environments. For polyp segmentation, we adopt CVC-300, CVC-ClinicDB, CVC-ColonDB, ETIS, and Kvasir, which together provide sufficient diversity in medical imaging conditions and polyp appearances. For shadow detection, we employ SBU and ISTD, two commonly used benchmarks with rich variations in shadow scale and intensity. These datasets are selected because they are widely recognized, diverse, and challenging benchmarks in their respective domains, ensuring both comprehensive evaluation and fair comparison with existing methods.

4.1.2 Experimental Details

We use the PyTorch framework to implement our models on 2 Nvidia RTX 3090 GPUs for 100 epochs and the batch size is 20. The input resolutions of images are resized to 352 × 352, and we employ a general multi-scale training strategy, as most methods do. For the optimizer, we use Adam [24]. The range of the learning rate is between 5×10^{-5} and 5×10^{-6}, and the learning rate varies with the cycle. During the training, the balance parameter λ is set to 8.0. The total number of parameters in the Abel-Net is 471.9 M, with a GFLOP of 232.1, and the FPS under the current training conditions is 4.24. To ensure a fair comparison, the results of the competing methods were primarily obtained by running their official pre-trained models or publicly available code on the same test sets used in this work. For methods where pre-trained weights were unavailable for specific datasets, we retrained them using their official implementations with default settings. All models were evaluated under consistent data preprocessing and evaluation protocols to guarantee the reliability of the comparison.

4.2 Evaluation Metrics

We mainly used five evaluation metrics to evaluate indicators: structure measure [25] (S_α), max F-measure [26] (F_β^m), weighted F-measure [27] (F_β^w), max enhanced-alignment measure [28] (E_ϕ), mean absolute error [29] (MAE) and precision-recall curve (PR curve).

4.3 Comparison with State-of-the-Art Methods

Tables 1–7 show the results of our method compared to other advanced methods on different datasets for nine different tasks, with the optimal method corresponding to each dataset marked in bold. We can see that in most tasks across the majority of datasets, our method can achieve good results. As shown in Figs. 4–6, we conduct a visual comparison with different state-of-the-art models. Through visual display, the superiority of our method can be more intuitively demonstrated: In nine different tasks, Abel-Net consistently achieves good results, indicating that our method has strong generalization capabilities.

Table 1: Quantitative comparison of different camouflaged object methods. The best result will be marked in **bold**

Methods	**CAMO-Test(250)**					**COD10K-Test(2026)**					**NC4K(4121)**				
	$S_\alpha \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$E_\phi \uparrow$	$F_\beta^m \uparrow$	$S_\alpha \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$E_\phi \uparrow$	$F_\beta^m \uparrow$	$S_\alpha \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$E_\phi \uparrow$	$F_\beta^m \uparrow$
FAPNet [30]	0.815	0.734	0.076	0.880	0.792	0.822	0.694	0.036	0.902	0.758	0.851	0.775	0.047	0.910	0.826
BGNet [31]	0.812	0.749	0.073	0.882	0.799	0.831	0.722	0.033	0.911	0.774	0.851	0.788	0.044	0.916	0.833
SegMar [32]	0.815	0.753	0.071	0.884	0.803	**0.883**	0.724	0.034	0.906	0.774	0.841	0.781	0.046	0.907	0.826
BSA-Net [33]	0.794	0.717	0.079	0.867	0.770	0.818	0.699	0.034	0.901	0.753	0.841	0.771	0.048	0.907	0.817
SINetV2 [20]	0.820	0.743	0.070	0.895	0.801	0.815	0.680	0.037	0.906	0.752	0.847	0.770	0.048	0.914	0.823
ZoomNet [34]	0.820	0.752	0.066	0.892	0.794	0.838	0.729	0.029	0.911	0.766	0.853	0.784	0.043	0.912	0.818
FEDER [35]	0.802	0.738	0.071	0.873	0.789	0.822	0.716	0.032	0.905	0.768	0.847	0.789	0.044	0.915	0.833
DGNet [12]	0.839	0.769	0.057	0.915	0.822	0.822	0.693	0.033	0.911	0.759	0.857	0.784	0.042	0.922	0.833
MFFN [36]	0.808	0.737	0.076	0.870	0.791	0.846	0.745	0.028	0.917	0.782	0.856	0.791	0.042	0.915	0.827
FSPNet [37]	0.856	0.799	0.050	0.899	0.830	0.851	0.735	0.026	0.930	0.769	0.879	0.816	0.035	0.915	0.843
HitNet [13]	0.844	0.801	0.057	0.902	0.863	0.868	**0.798**	**0.024**	0.932	**0.838**	0.870	0.825	0.039	0.921	0.838
Abel-Net	**0.883**	**0.840**	**0.044**	**0.942**	**0.880**	0.866	0.775	0.025	**0.939**	0.826	**0.891**	**0.840**	**0.033**	**0.943**	**0.878**

Table 2: Quantitative comparison of different polyp segmentation methods. The best result will be marked in **bold**

Methods	CVC-300				CVC-ClinicDB				CVC-ColonDB				ETIS				Kvasir			
	$S_\alpha \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$E_\phi \uparrow$	$S_\alpha \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$E_\phi \uparrow$	$S_\alpha \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$E_\phi \uparrow$	$S_\alpha \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$E_\phi \uparrow$	$S_\alpha \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$E_\phi \uparrow$
U-Net [7]	0.843	0.684	0.022	0.876	0.889	0.811	0.019	0.954	0.712	0.498	0.061	0.776	0.684	0.366	0.036	0.740	0.858	0.794	0.055	0.893
U-Net++ [38]	0.839	0.687	0.018	0.898	0.873	0.785	0.022	0.931	0.691	0.467	0.064	0.760	0.683	0.390	0.035	0.776	0.862	0.808	0.048	0.910
SFA [39]	0.640	0.341	0.065	0.817	0.793	0.647	0.042	0.885	0.634	0.379	0.094	0.765	0.557	0.231	0.109	0.633	0.782	0.670	0.075	0.849
PraNet [14]	0.925	0.843	0.010	**0.972**	0.936	0.896	0.009	0.979	0.819	0.696	0.045	0.869	0.794	0.600	0.031	0.841	0.915	0.885	0.030	0.948
SANet [40]	0.928	0.859	0.008	**0.972**	0.939	0.909	0.012	0.976	0.837	0.726	0.043	0.878	0.849	0.685	0.015	0.897	0.915	0.892	0.028	0.953
MSNet [15]	0.920	0.849	0.010	0.943	0.941	0.914	0.008	0.972	0.836	0.737	0.041	0.883	0.840	0.678	0.020	0.830	0.922	0.893	0.028	0.944
PEFNet [41]	–	–	0.010	–	–	–	0.010	–	–	–	0.036	–	–	–	0.019	–	–	–	0.029	–
M2UNet [42]	–	–	**0.007**	–	–	–	0.008	–	–	–	0.036	–	–	–	0.024	–	–	–	0.025	–
MEGANet (ResNet) [43]	0.924	0.863	0.009	0.959	**0.950**	0.931	0.008	0.980	0.845	0.766	0.038	**0.899**	0.866	0.753	0.015	0.915	0.916	0.904	0.026	0.954
MEGANet (Res2Net) [43]	0.882	0.834	**0.007**	0.969	**0.950**	**0.940**	**0.006**	**0.986**	0.854	**0.779**	0.040	0.895	0.836	0.702	0.037	0.858	0.918	0.907	0.025	0.959
DuAT [44]	–	–	–	–	–	–	**0.006**	–	–	–	**0.026**	–	–	–	**0.013**	–	–	–	0.023	–
Abel-Net	**0.938**	**0.874**	0.009	**0.972**	0.931	0.896	0.010	0.968	**0.859**	0.759	0.039	0.883	**0.880**	**0.758**	0.016	**0.919**	**0.933**	**0.916**	**0.021**	**0.970**

Table 3: Quantitative comparison of different remote sensing image salient object detection methods. The best result will be marked in **bold**

Method	EORSSD			
	$S_\alpha \uparrow$	MAE $\downarrow$	$E_\phi \uparrow$	$F_\beta^m \uparrow$
LVNet [45]	0.863	0.146	0.925	0.779
DAFNet [46]	0.917	0.006	**0.986**	0.861
CorrNet [47]	0.929	0.008	0.970	0.878
MJRBM [48]	0.920	0.010	0.965	0.867
SARNet [49]	0.924	0.010	0.965	0.872
SeaNet [17]	0.921	0.007	0.971	0.865
HVPNet [50]	0.873	0.010	0.948	0.804
SAMNet [51]	0.862	0.013	0.942	0.781
MCCNet [16]	0.933	0.007	0.976	0.890
Abel-Net	**0.942**	**0.005**	0.984	**0.897**

Table 4: Quantitative comparison of different defocus blur detection. The best result will be marked in **bold**

Methods	CUHK		DUT	
	$F_\beta^w \uparrow$	MAE $\downarrow$	$F_\beta^w \uparrow$	MAE $\downarrow$
IS2Net [52]	**0.964**	0.049	0.868	0.142
DeFusionNet [53]	0.818	0.117	0.823	0.118
BTBNet [19]	0.889	0.082	0.827	0.138
CENet [54]	0.906	0.059	0.817	0.135
DAD [55]	0.884	0.079	0.794	0.153
EFENet [56]	0.914	0.053	0.854	0.094
EVPv1(Seg) [10]	0.928	0.045	**0.890**	0.068
EVPv2(Seg) [10]	0.932	0.042	0.887	0.070

(Continued)

Table 4 (continued)

Methods	CUHK		DUT	
	F_β^w ↑	MAE ↓	F_β^w ↑	MAE ↓
Abel-Net	0.904	**0.039**	0.885	**0.057**

Table 5: Quantitative comparison of different salient object detection. The best result will be marked in **bold**

Methods	OMRON			DUTS-TE			HKU-IS			PASCAL-S			ECSSD		
	S_α ↑	M ↓	F_β^m ↑	S_α ↑	M ↓	F_β^m ↑	S_α ↑	M ↓	F_β^m ↑	S_α ↑	M ↓	F_β^m ↑	S_α ↑	M ↓	F_β^m ↑
AFNet (VGG16) [57]	0.826	0.057	0.797	0.867	0.045	0.862	0.906	0.036	0.923	0.845	0.070	0.863	0.914	0.042	0.935
EGNet (VGG16) [58]	0.836	0.056	0.809	0.879	0.043	0.876	0.910	0.035	0.927	0.848	0.078	0.856	0.919	0.041	0.943
EGNet (ResNet50) [58]	0.841	0.053	0.816	0.887	0.039	0.888	0.918	0.031	0.935	0.852	0.074	0.865	0.925	0.037	0.948
CPD (VGG16) [59]	0.818	0.057	0.794	0.867	0.043	0.864	0.904	0.033	0.924	0.845	0.072	0.861	0.910	0.040	0.936
CPD (ResNet50) [59]	0.825	0.056	0.797	0.869	0.043	0.865	0.906	0.034	0.925	0.848	0.071	0.860	0.918	0.037	0.939
BASNet [60]	0.836	0.056	0.805	0.866	0.047	0.859	0.909	0.032	0.928	0.838	0.076	0.854	0.916	0.037	0.943
AADFNet [61]	0.839	0.049	0.814	0.891	0.031	0.899	0.919	0.026	0.942	0.866	0.055	0.880	0.930	0.028	0.954
GateNet (VGG16) [62]	0.821	0.061	0.794	0.871	0.045	0.870	0.910	0.036	0.929	0.857	0.069	0.870	0.917	0.028	0.954
GateNet (ResNet50) [62]	0.838	0.055	0.818	0.885	0.040	0.887	0.915	0.034	0.934	0.858	0.068	0.869	0.920	0.041	0.945
GateNet (ResNet101) [62]	0.845	0.055	0.821	0.891	0.038	0.892	0.920	0.032	0.938	0.862	0.067	0.870	0.930	0.036	0.951
U2Net [63]	0.847	0.054	0.823	0.874	0.044	0.872	0.916	0.031	0.935	0.844	0.074	0.859	0.928	0.033	0.951
MINet (VGG16) [64]	0.822	0.057	0.794	0.875	0.040	0.876	0.912	0.032	0.930	0.854	0.065	0.865	0.919	0.037	0.944
MINet (ResNet50) [64]	0.833	0.056	0.810	0.884	0.037	0.883	0.919	0.029	0.935	0.856	0.064	0.867	0.925	0.034	0.948
LDF [65]	0.839	0.052	0.820	0.892	0.034	0.897	0.920	0.028	0.939	0.863	0.060	0.874	0.924	0.034	0.950
SAC [66]	0.849	0.052	0.829	0.896	0.034	0.894	0.925	0.026	0.942	0.866	0.062	0.877	0.931	0.031	0.951
CANet [67]	0.837	0.058	0.810	0.878	0.044	0.876	0.910	0.037	0.930	0.855	0.073	0.866	0.915	0.044	0.938
SGL-KRN [68]	0.848	0.049	0.796	0.893	0.034	0.883	0.921	0.028	0.930	0.856	0.068	0.850	0.923	0.036	0.937
PA-KRN [68]	0.853	0.050	0.810	0.901	0.033	0.895	0.924	0.027	0.935	0.858	0.067	0.853	0.928	0.032	0.943
ICON [5]	0.845	0.057	0.825	0.889	0.037	0.892	0.920	0.029	0.940	0.861	0.064	0.856	0.929	0.032	0.950
EDN (VGG16) [11]	0.838	0.057	0.782	0.883	0.041	0.864	0.921	0.029	0.929	0.861	0.065	0.856	0.928	0.034	0.941
EDN (ResNet50) [11]	0.850	0.049	0.799	0.892	0.035	0.878	0.924	0.026	0.933	0.865	0.062	0.860	0.927	0.032	0.941
MENet [69]	0.850	0.045	**0.834**	0.905	0.028	0.912	0.927	0.023	0.948	0.872	0.054	**0.890**	0.928	0.031	0.955
Abel-Net	**0.873**	**0.043**	0.833	**0.922**	**0.024**	**0.919**	**0.936**	**0.021**	**0.949**	**0.884**	**0.049**	0.885	**0.942**	**0.023**	**0.956**

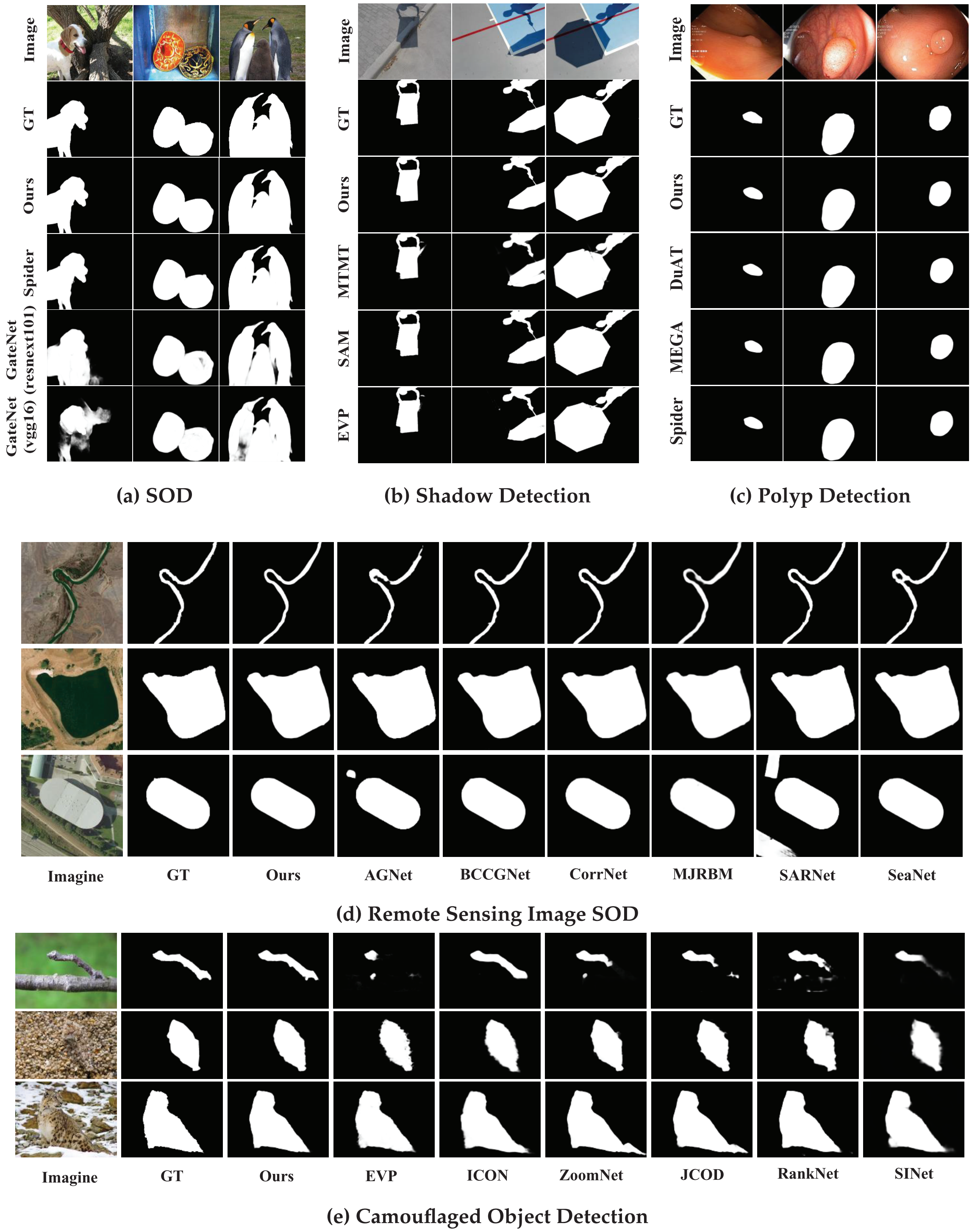
Table 6: Quantitative comparison of different transparent, glass, and mirror detection methods. The best result will be marked in **bold**

Dataset	Method	$S_\alpha \uparrow$	$F_\beta^w \uparrow$	MAE $\downarrow$
Transparent object detection				
Transparent-easy	Translab [70]	0.935	0.941	0.022
	GatedNet [9]	**0.963**	**0.974**	**0.011**
	Abel-Net	**0.963**	0.968	0.013
Transparent-hard	Translab [70]	0.798	0.783	0.087
	GatedNet [9]	0.871	**0.874**	**0.053**
	Abel-Net	**0.877**	0.872	0.054
Glass detection				
GDD	GDNet [71]	0.864	0.901	0.061
	EBLNet [72]	0.875	0.908	0.056
	GatedNet [9]	**0.892**	0.921	0.049
	Abel-Net	0.882	**0.927**	**0.046**
Mirror detection				
MSD	MirrorNet [73]	0.846	0.744	0.085
	GatedNet [9]	0.872	0.829	0.053
	Abel-Net	**0.908**	**0.896**	**0.037**

Table 7: Quantitative comparison of different shadow detection. The best result will be marked in **bold**

Methods	SBU				ISTD			
	$S_\alpha \uparrow$	$F_\beta^w \uparrow$	MAE $\downarrow$	$F_\beta^m \uparrow$	$S_\alpha \uparrow$	$F_\beta^w \uparrow$	MAE $\downarrow$	$F_\beta^m \uparrow$
DSC [6]	0.856	0.861	0.032	0.914	–	–	–	–
ADNet [74]	0.700	0.424	0.201	0.877	–	–	–	–
BDRAR [75]	0.844	0.827	0.039	0.884	0.901	0.878	0.027	0.910
DSD [76]	0.851	0.835	0.036	0.896	0.930	0.897	0.023	0.933
GatedNet [9]	**0.886**	0.889	0.025	**0.937**	0.956	0.938	0.012	**0.965**
Abel-Net	**0.886**	**0.900**	**0.023**	0.931	**0.958**	**0.949**	**0.009**	0.960

At the same time, Fig. 7 shows the PR curve of the results achieved by our method. The PR curve is a tool that is used to evaluate the performance of binary classification models. It helps analyze the classification performance of different models by plotting the relationship between precision and recall. From Fig. 7, we can further see the superiority of our method. Whether in popular classification tasks like camouflaged object detection and remote sensing image salient object detection, or the well-developed polyp segmentation task, our method performs well.



(a) SOD

(b) Shadow Detection

(c) Polyp Detection

(d) Remote Sensing Image SOD

(e) Camouflaged Object Detection

Figure 4: Comparison with other methods on some common datasets. (**a**) Spider [77] and GateNet [9] on DUTS dataset (top-left). (**b**) MTMT [78], SAM [79], and EVP [10] on ISTD dataset (top-middle). (**c**) DuAT [44], MEGA [43], and Spider [77] on Kvasir dataset (top-right). (**d**) AGNet [80], BSCGNet [81], CorrNet [47], MJRBM [48], SARNet [49], and SeaNet on EORSSD dataset (middle). (**e**) EVP [10], ICON [5], ZoomNet [34], JCOD [82], RankNet [83], and SINet [20] on CAMO dataset (bottom)

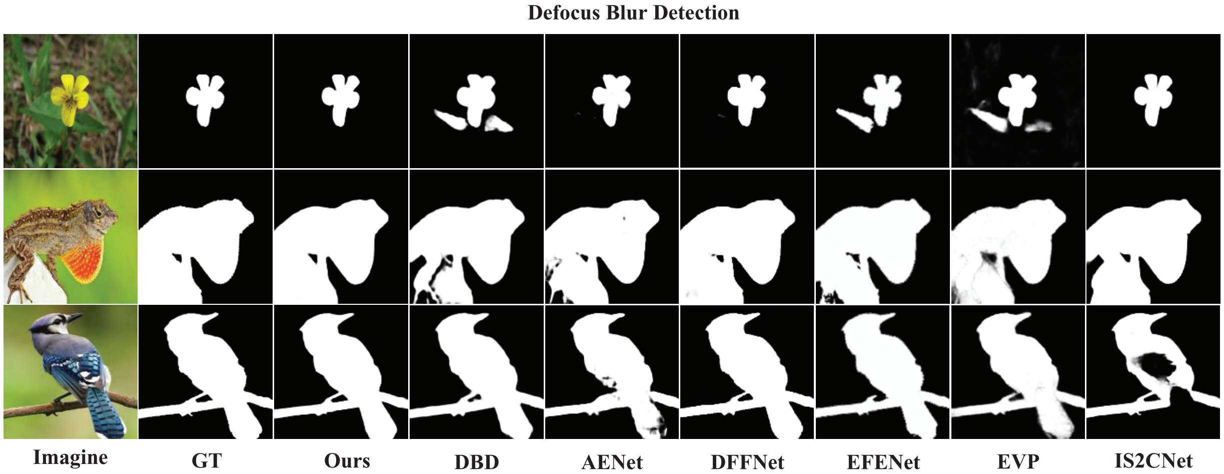


Figure 5: Our method can achieve good results on the defocus blur detection task. Here are the visualization results of our method and other advanced methods on the CUHK dataset: DBD [84], AENet [56], DFFNet [85], EFENet [56], EVP [10], IS2CNet [52]

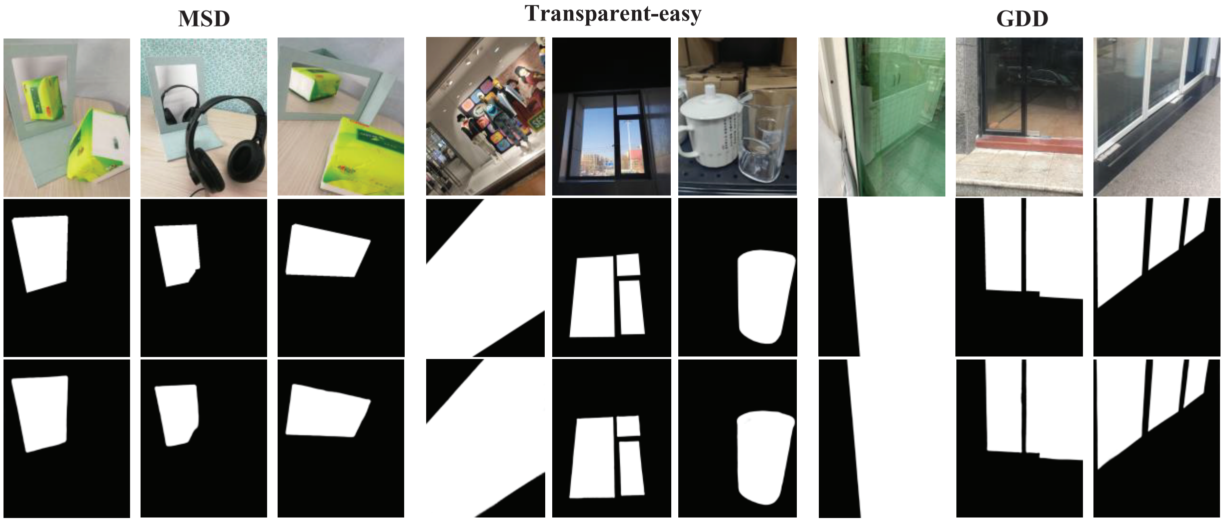
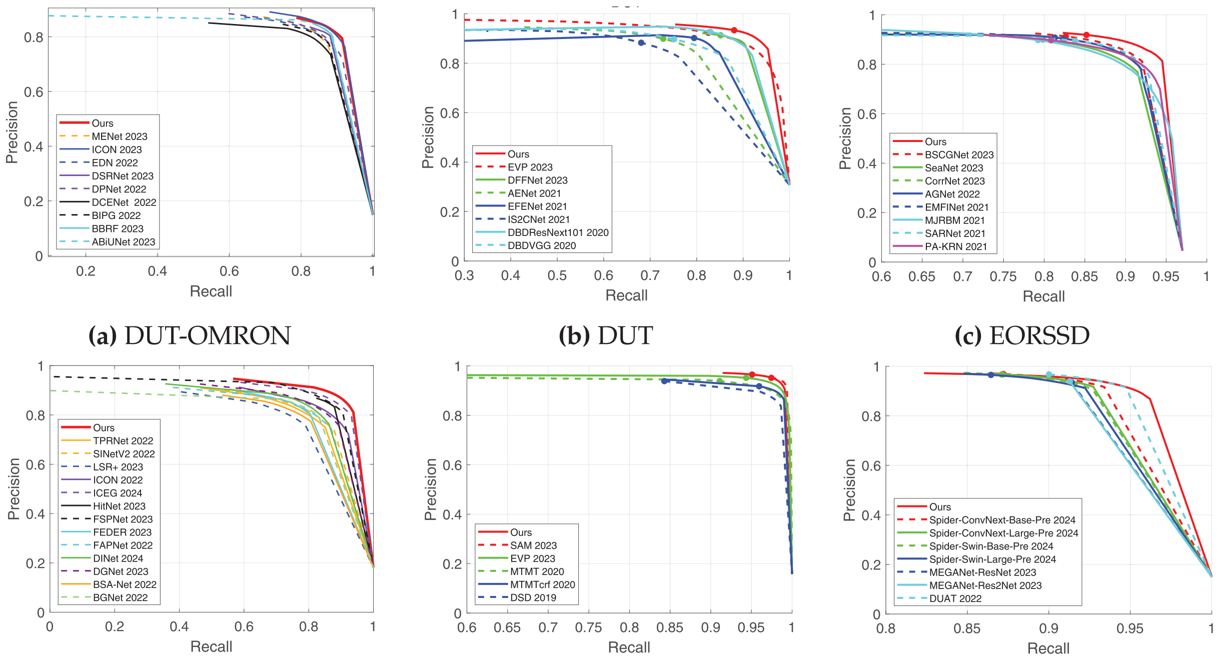


Figure 6: Results of our method on mirror detection, transparent object detection, and glass detection. The top row represents the original images, the middle row represents the GT, and the bottom row represents our results

4.4 Parameter Analysis

Compared to most networks with multi-parameter fine-tuning, we introduce only a small number of adjustable parameters, yet our performance can surpass the best networks in the vast majority of tasks. It is worth noting that the substantial total parameter count (471.9 M) primarily originates from the PVTv2 backbone employed to ensure robust feature extraction, while the computational overhead introduced by our proposed modules (ASC, EDL, EA, EFA) remains marginal. As shown in Fig. 8, we performed multiple experiments with different λ parameters on five SOD datasets. In Fig. 8, we test the impact of different parameters on the evaluation metrics S_α, MAE, and F_β^m. The experiments and the curves plotted based on the experimental results demonstrate that with lambda = 8, our network achieves the best performance. We analyze the impact of the parameter on the network: When the parameter is too small, the network fails to accurately extract edge information from the ground truth map. When the parameter is too large, the edge information at the current stage cannot be effectively transmitted to the next stage, resulting in the loss of too much valuable information during the training process.



(a) DUT-OMRON **(b)** DUT **(c)** EORSSD

(d) CAMO **(e)** ISTD **(f)** Kvasir

Figure 7: Comparison of our method's PR curve with other methods on six different task datasets

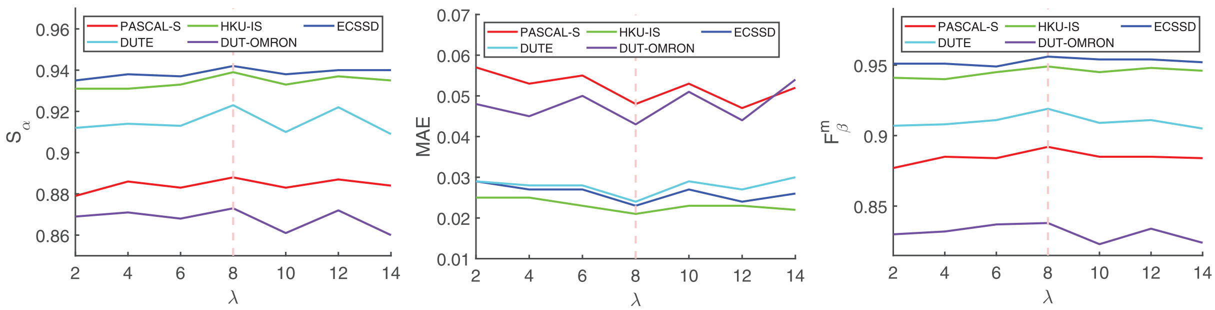


Figure 8: The impact of different parameters λ on experimental results. Left: The impact of parameters on S_α. Middle: The impact of parameters on MAE. Right: The impact of parameters on F^m_β

4.5 Ablation Experiments

To verify the impact of each module on the entire network, we conduct an ablation study on nine datasets corresponding to nine tasks using S_α and MAE as evaluation metrics. Table 8 presents four sets of data, showing the experimental results of our network after removing certain components. Without EA or EFA means replacing EA and EFA with regular attention modules. Without EDL indicates that we transform the results of the dual localization into an all-ones matrix, which is then used as input for the next step of edge prediction. Without ASC means that at each stage, the decoder only receives the feature information from the encoder of the current stage, while ignoring the feature information from the encoder of the previous stage, as well as the fused features from the FPN. The results indicate that the experimental outcomes under the baseline (with no missing modules) are significantly better compared to the aforementioned four experimental groups. This fully demonstrates the importance of each module to the network. Regardless of which module is missing, the network's performance is substantially affected, highlighting that these modules effectively improve the accuracy of edge prediction in binary image segmentation.

Table 8: Ablation experiments for nine binary segmentation tasks. The best result will be marked in **bold**

Methods	**PASCAL-S**		**COD10K**		**ISTD**		**Trans-Hard**		**GDD**		**MSD**		**Kvasir**		**EORSSD**		**CUHK**	
	$S_\alpha\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$M\downarrow$
w/o EA	0.874	0.058	0.858	0.027	0.935	0.020	0.828	0.090	0.847	0.058	0.906	0.038	0.887	0.047	0.853	0.015	0.889	0.060
w/o EFA	0.866	0.086	0.847	0.030	0.957	0.010	0.829	0.088	0.827	0.076	0.898	0.041	0.913	0.031	0.877	0.010	0.896	0.051
w/o EA&EFA	0.864	0.074	0.602	0.145	0.846	0.060	0.847	0.076	0.811	0.093	0.890	0.044	0.916	0.028	0.781	0.030	0.894	0.056
w/o EDL	0.883	0.055	0.840	0.032	0.949	0.014	0.826	0.094	0.823	0.082	0.874	0.061	0.903	0.034	0.855	0.014	0.899	0.051
w/o ASC	0.883	0.052	0.864	0.028	0.947	0.013	0.876	0.056	0.870	0.047	0.881	0.040	0.910	0.028	0.942	0.006	0.900	0.041
Baseline	**0.884**	**0.049**	**0.866**	**0.025**	**0.958**	**0.009**	**0.877**	**0.054**	**0.882**	**0.046**	**0.908**	**0.037**	**0.922**	**0.021**	**0.943**	**0.005**	**0.904**	**0.039**

5 Statements

5.1 Conclusion

In this paper, we propose Abel-Net, a universal network designed for multi-task binary image segmentation. This network aggregates multi-level feature information through a feature pyramid module and explicitly optimizes boundary identification through a two-stage edge dual localization (EDL) strategy. By integrating the Aggregated Skip Connection (ASC) and edge-aware attention mechanisms (EA/EFA), Abel-Net effectively improves the accuracy of segmentation edges. Extensive experiments demonstrate that our network performs comparably to, and in many cases surpasses, state-of-the-art task-specific networks across nine diverse binary segmentation tasks. In summary, our method outperforms most networks designed specifically for these tasks in single-task scenarios and exhibits strong adaptability to a wide range of visual perception challenges.

5.2 Limitations and Future Scope

Despite the superior performance, we acknowledge certain limitations in the current framework. First, the introduction of the dual-branch EDL module and attention mechanisms introduces extra computational burden and memory requirements. As noted in our parameter analysis, the model operates at 4.24 FPS with 471.9 M parameters (largely due to the backbone), which restricts its deployment in real-time or resource-constrained applications. Second, while the network is designed as a universal framework, the optimal balance between edge and semantic features may vary across tasks with distinct boundary characteristics, potentially requiring task-specific fine-tuning of hyperparameters.

Future work will focus on two main directions: (1) Enhancing robustness and generalization to unseen complex scenarios where extremely fine edges are present; (2) Optimizing the architecture for better efficiency, potentially by exploring lightweight backbones or knowledge distillation techniques to make the model suitable for real-time applications. We hope this work will encourage further exploratory research in the field of general binary segmentation.

Acknowledgement: The authors would like to thank their colleagues and mentors for their helpful discussions and encouragement during the course of this research.

Funding Statement: This work is supported by the National Natural Science Foundation of China under Grant 62206083.

Author Contributions: The authors confirm their contribution to the paper as follows: Zhengyu Wu conceptualized the study, conducted the experiments, collected and analyzed the data, and contributed to the method development, model evaluation, performance comparison, and manuscript writing. Kejun Kang participated in the study conceptualization and experiment implementation, contributed to data collection and performance comparison. Yixiu Liu provided overall guidance and supervision throughout the research process, and contributed to the manuscript revision. Chenpu Li conducted some ablation experiment verification and paper polishing work. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Data available on request from the authors.

Ethics Approval: The study did not include human or animal subjects.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Shu X, Zhang A, Xu Z, Zhu F, Hua W. Adaptive encoding and comprehensive attention decoding network for medical image segmentation. Appl Soft Comput. 2025;174:112990. doi:10.1016/j.asoc.2025.112990.
2. Al-Sahaf H, Mesejo P, Bi Y, Zhang M. Evolutionary deep learning for computer vision and image processing. Appl Soft Comput. 2024;151:111159. doi:10.1016/j.asoc.2023.111159.
3. Liu X, Li D. Binary segmentation based on visual attention consistency under background-change. Appl Soft Comput. 2022;121:108738. doi:10.1016/j.asoc.2022.108738.
4. Liu CL, Chung CC. Anomaly detection and segmentation in industrial images using multi-scale reverse distillation. Appl Soft Comput. 2025;168:112502. doi:10.1016/j.asoc.2024.112502.
5. Zhuge M, Fan DP, Liu N, Zhang D, Xu D, Shao L. Salient object detection via integrity learning. IEEE Trans Pattern Anal Mach Intell. 2023;45(3):3738–52. doi:10.1109/TPAMI.2022.3179526.
6. Hu X, Zhu L, Fu CW, Qin J, Heng PA. Direction-aware spatial context features for shadow detection. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 7454–62. doi:10.1109/CVPR.2018.00778.
7. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015. Cham, Switzerland: Springer International Publishing; 2015. p. 234–41. doi:10.1007/978-3-319-24574-4_28.
8. Ghosh S, Das N, Das I, Maulik U. Understanding deep learning techniques for image segmentation. ACM Comput Surv. 2020;52(4):1–35. doi:10.1145/3329784.
9. Zhao X, Pang Y, Zhang L, Lu H, Zhang L. Towards diverse binary segmentation via a simple yet general gated network. Int J Comput Vis. 2024;132(10):4157–234. doi:10.1007/s11263-024-02058-y.
10. Liu W, Shen X, Pun CM, Cun X. Explicit visual prompting for universal foreground segmentations. IEEE Trans Pattern Anal Mach Intell. 2025:1–16. doi:10.1109/tpami.2025.3619490.
11. Wu YH, Liu Y, Zhang L, Cheng MM, Ren B. EDN: salient object detection via extremely-downsampled network. IEEE Trans Image Process. 2022;31:3125–36. doi:10.1109/TIP.2022.3164550.
12. Ji GP, Fan DP, Chou YC, Dai D, Liniger A, Van Gool L. Deep gradient learning for efficient camouflaged object detection. Mach Intell Res. 2023;20(1):92–108. doi:10.1007/s11633-022-1365-9.
13. Hu X, Wang S, Qin X, Dai H, Ren W, Luo D, et al. High-resolution iterative feedback network for camouflaged object detection. Proc AAAI Conf Artif Intell. 2023;37(1):881–9. doi:10.1609/aaai.v37i1.25167.

14. Fan DP, Ji GP, Zhou T, Chen G, Fu H, Shen J, et al. PraNet: parallel reverse attention network for polyp segmentation. In: Medical image computing and computer assisted intervention–MICCAI 2020. Cham, Switzerland: Springer International Publishing; 2020. p. 263–73. doi:10.1007/978-3-030-59725-2_26.
15. Zhao X, Zhang L, Lu H. Automatic polyp segmentation via multi-scale subtraction network. In: Medical image computing and computer assisted intervention–MICCAI 2021. Cham, Switzerland: Springer International Publishing; 2021. p. 120–30. doi:10.1007/978-3-030-87193-2_12.
16. Li G, Liu Z, Lin W, Ling H. Multi-content complementation network for salient object detection in optical remote sensing images. IEEE Trans Geosci Remote Sens. 2022;60:5614513. doi:10.1109/TGRS.2021.3131221.
17. Li G, Liu Z, Zhang X, Lin W. Lightweight salient object detection in optical remote-sensing images via semantic matching and edge alignment. IEEE Trans Geosci Remote Sens. 2023;61:5601111. doi:10.1109/TGRS.2023.3235717.
18. Shi J, Xu L, Jia J. Discriminative blur detection features. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition; 2014 Jun 23–28; Columbus, OH, USA. p. 2965–72. doi:10.1109/CVPR.2014.379.
19. Zhao W, Zhao F, Wang D, Lu H. Defocus blur detection via multi-stream bottom-top-bottom network. IEEE Trans Pattern Anal Mach Intell. 2020;42(8):1884–97. doi:10.1109/TPAMI.2019.2906588.
20. Fan DP, Ji GP, Cheng MM, Shao L. Concealed object detection. IEEE Trans Pattern Anal Mach Intell. 2022;44(10):6024–42. doi:10.1109/TPAMI.2021.3085766.
21. Wang W, Xie E, Li X, Fan DP, Song K, Liang D, et al. Pyramid vision transformer: a versatile backbone for dense prediction without convolutions. In: Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV); 2021 Oct 10–17; Montreal, QC, Canada. p. 548–58. doi:10.1109/ICCV48922.2021.00061.
22. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV); 2021 Oct 10–17; Montreal, QC, Canada. p. 9992–10002. doi:10.1109/ICCV48922.2021.00986.
23. Wei J, Wang S, Huang Q. F^{3}Net: fusion, feedback and focus for salient object detection. Proc AAAI Conf Artif Intell. 2020;34(7):12321–8. doi:10.1609/aaai.v34i07.6916.
24. Kingma DP, Ba J. Adam: a method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015; 2015 May 7–9; San Diego, CA, USA. p. 1–15.
25. Cheng MM, Fan DP. Structure-measure: a new way to evaluate foreground maps. Int J Comput Vis. 2021;129(9):2622–38. doi:10.1007/s11263-021-01490-8.
26. Achanta R, Hemami S, Estrada F, Susstrunk S. Frequency-tuned salient region detection. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2009 Jun 20–25; Miami, FL, USA. p. 1597–604. doi:10.1109/CVPR.2009.5206596.
27. Margolin R, Zelnik-Manor L, Tal A. How to evaluate foreground maps. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition; 2014 Jun 23–28; Columbus, OH, USA. p. 248–55. doi:10.1109/CVPR.2014.39.
28. Fan DP, Gong C, Cao Y, Ren B, Cheng MM, Borji A. Enhanced-alignment measure for binary foreground map evaluation. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence; 2018 Jul 13–19; Stockholm, Sweden. International Joint Conferences on Artificial Intelligence Organization; 2018. p. 698–704. doi:10.24963/ijcai.2018/97.
29. Perazzi F, Krhenbhl P, Pritch Y, Hornung A. Saliency filters: contrast based filtering for salient region detection. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition; 2012 Jun 16–21; Providence, RI, USA. p. 733–40. doi:10.1109/CVPR.2012.6247743.
30. Zhou T, Zhou Y, Gong C, Yang J, Zhang Y. Feature aggregation and propagation network for camouflaged object detection. IEEE Trans Image Process. 2022;31:7036–47. doi:10.1109/TIP.2022.3217695.
31. Sun Y, Wang S, Chen C, Xiang TZ. Boundary-guided camouflaged object detection. In: Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence; 2022 Jul 23–29; Vienna, Austria; 2022. p. 1335–41. doi:10.24963/ijcai.2022/186.
32. Jia Q, Yao S, Liu Y, Fan X, Liu R, Luo Z. Segment, magnify and reiterate: detecting camouflaged objects the hard way. In: Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022 Jun 18–24; New Orleans, LA, USA. p. 4703–12. doi:10.1109/CVPR52688.2022.00467.

33. Zhu H, Li P, Xie H, Yan X, Liang D, Chen D, et al. I can find you! Boundary-guided separated attention network for camouflaged object detection. Proc AAAI Conf Artif Intell. 2022;36(3):3608–16. doi:10.1609/aaai.v36i3.20273.
34. Pang Y, Zhao X, Xiang TZ, Zhang L, Lu H. Zoom in and out: a mixed-scale triplet network for camouflaged object detection. In: Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022 Jun 18–24; New Orleans, LA, USA. p. 2150–60. doi:10.1109/CVPR52688.2022.00220.
35. He C, Li K, Zhang Y, Tang L, Zhang Y, Guo Z, et al. Camouflaged object detection with feature decomposition and edge reconstruction. In: Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2023 Jun 17–24; Vancouver, BC, Canada. p. 22046–55. doi:10.1109/CVPR52729.2023.02111.
36. Zheng D, Zheng X, Yang LT, Gao Y, Zhu C, Ruan Y. MFFN: multi-view feature fusion network for camouflaged object detection. In: Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV); 2023 Jan 2–7; Waikoloa, HI, USA. p. 6221–31. doi:10.1109/WACV56688.2023.00617.
37. Huang Z, Dai H, Xiang TZ, Wang S, Chen HX, Qin J, et al. Feature shrinkage pyramid for camouflaged object detection with transformers. In: Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2023 Jun 17–24; Vancouver, BC, Canada. p. 5557–66. doi:10.1109/CVPR52729.2023.00538.
38. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans Med Imaging. 2020;39(6):1856–67. doi:10.1109/TMI.2019.2959609.
39. Fang Y, Chen C, Yuan Y, Tong KY. Selective feature aggregation network with area-boundary constraints for polyp segmentation. In: Medical image computing and computer assisted intervention–MICCAI 2019. Cham, Switzerland: Springer International Publishing; 2019. p. 302–10. doi:10.1007/978-3-030-32239-7_34.
40. Wei J, Hu Y, Zhang R, Li Z, Zhou SK, Cui S. Shallow attention network for polyp segmentation. In: Medical image computing and computer assisted intervention–MICCAI 2021. Cham, Switzerland: Springer International Publishing; 2021. p. 699–708. doi:10.1007/978-3-030-87193-2_66.
41. Nguyen-Mau TH, Trinh QH, Bui NT, Thi PV, Nguyen MV, Cao XN, et al. PEFNet: positional embedding feature for polyp segmentation. In: MultiMedia modeling. Cham, Switzerland: Springer Nature Switzerland; 2023. p. 240–51. doi:10.1007/978-3-031-27818-1_20.
42. Trinh QH, Bui NT, Nguyen-Mau TH, Nguyen MV, Phan HM, Tran MT, et al. M2UNet: MetaFormer multi-scale upsampling network for polyp segmentation. In: Proceedings of the 2023 31st European Signal Processing Conference (EUSIPCO); 2023 Sep 4–8; Helsinki, Finland. p. 1115–9.
43. Bui NT, Hoang DH, Nguyen QT, Tran MT, Le N. MEGANet: multi-scale edge-guided attention network for weak boundary polyp segmentation. In: Proceedings of the 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV); 2024 Jan 3–8; Waikoloa, HI, USA. p. 7970–9. doi:10.1109/WACV57701.2024.00780.
44. Ho NV, Nguyen T, Diep GH, Le N, Hua BS. Point-unet: a context-aware point-based neural network for volumetric segmentation. In: Medical image computing and computer assisted intervention–MICCAI 2021. Cham, Switzerland: Springer International Publishing; 2021. p. 644–55. doi:10.1007/978-3-030-87193-2_61.
45. Li C, Cong R, Hou J, Zhang S, Qian Y, Kwong S. Nested network with two-stream pyramid for salient object detection in optical remote sensing images. IEEE Trans Geosci Remote Sens. 2019;57(11):9156–66. doi:10.1109/TGRS.2019.2925070.
46. Zhang Q, Cong R, Li C, Cheng MM, Fang Y, Cao X, et al. Dense attention fluid network for salient object detection in optical remote sensing images. IEEE Trans Image Process. 2021;30:1305–17. doi:10.1109/TIP.2020.3042084.
47. Li G, Liu Z, Bai Z, Lin W, Ling H. Lightweight salient object detection in optical remote sensing images via feature correlation. IEEE Trans Geosci Remote Sens. 2022;60:5617712. doi:10.1109/TGRS.2022.3145483.
48. Tu Z, Wang C, Li C, Fan M, Zhao H, Luo B. ORSI salient object detection via multiscale joint region and boundary model. IEEE Trans Geosci Remote Sens. 2022;60:5607913. doi:10.1109/TGRS.2021.3101359.
49. Huang Z, Chen H, Liu B, Wang Z. Semantic-guided attention refinement network for salient object detection in optical remote sensing images. Remote Sens. 2021;13(11):2163. doi:10.3390/rs13112163.
50. Liu Y, Gu YC, Zhang XY, Wang W, Cheng MM. Lightweight salient object detection via hierarchical visual perception learning. IEEE Trans Cybern. 2021;51(9):4439–49. doi:10.1109/TCYB.2020.3035613.
51. Liu Y, Zhang XY, Bian JW, Zhang L, Cheng MM. SAMNet: stereoscopically attentive multi-scale network for lightweight salient object detection. IEEE Trans Image Process. 2021;30:3804–14. doi:10.1109/TIP.2021.3065239.

52. Zhao F, Lu H, Zhao W, Yao L. Image-scale-symmetric cooperative network for defocus blur detection. IEEE Trans Circuits Syst Video Technol. 2022;32(5):2719–31. doi:10.1109/TCSVT.2021.3095347.
53. Tang C, Zhu X, Liu X, Wang L, Zomaya A. DeFusionNET: defocus blur detection via recurrently fusing and refining multi-scale deep features. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15–20; Long Beach, CA, USA. p. 2695–704. doi:10.1109/CVPR.2019.00281.
54. Zhao W, Zheng B, Lin Q, Lu H. Enhancing diversity of defocus blur detectors via cross-ensemble network. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15–20; Long Beach, CA, USA. p. 8897–905. doi:10.1109/CVPR.2019.00911.
55. Zhao W, Shang C, Lu H. Self-generated defocus blur detection via dual adversarial discriminators. In: Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021 Jun 20–25; Nashville, TN, USA. p. 6929–38. doi:10.1109/cvpr46437.2021.00686.
56. Zhao W, Hou X, He Y, Lu H. Defocus blur detection via boosting diversity of deep ensemble networks. IEEE Trans Image Process. 2021;30:5426–38. doi:10.1109/tip.2021.3084101.
57. Feng M, Lu H, Ding E. Attentive feedback network for boundary-aware salient object detection. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15–20; Long Beach, CA, USA. p. 1623–32. doi:10.1109/CVPR.2019.00172.
58. Zhao J, Liu JJ, Fan DP, Cao Y, Yang J, Cheng MM. EGNet: edge guidance network for salient object detection. In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV); 2019 Oct 27–Nov 2; Seoul, Republic of Korea. p. 8778–87. doi:10.1109/iccv.2019.00887.
59. Wu Z, Su L, Huang Q. Cascaded partial decoder for fast and accurate salient object detection. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15–20; Long Beach, CA, USA. p. 3902–11. doi:10.1109/CVPR.2019.00403.
60. Qin X, Zhang Z, Huang C, Gao C, Dehghan M, Jagersand M. BASNet: boundary-aware salient object detection. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15–20; Long Beach, CA, USA. p. 7471–81. doi:10.1109/CVPR.2019.00766.
61. Zhu L, Chen J, Hu X, Fu CW, Xu X, Qin J, et al. Aggregating attentional dilated features for salient object detection. IEEE Trans Circuits Syst Video Technol. 2020;30(10):3358–71. doi:10.1109/TCSVT.2019.2941017.
62. Zhao X, Pang Y, Zhang L, Lu H, Zhang L. Suppress and balance: a simple gated network for salient object detection. In: Computer vision-ECCV 2020. Cham, Switzerland: Springer International Publishing; 2020. p. 35–51. doi:10.1007/978-3-030-58536-5_3.
63. Qin X, Zhang Z, Huang C, Dehghan M, Zaiane OR, Jagersand M. U^2-Net: going deeper with nested U-structure for salient object detection. Pattern Recognit. 2020;106:107404. doi:10.1016/j.patcog.2020.107404.
64. Zhang L, Wu J, Wang T, Borji A, Wei G, Lu H. A multistage refinement network for salient object detection. IEEE Trans Image Process. 2020;29:3534–45. doi:10.1109/TIP.2019.2962688.
65. Wei J, Wang S, Wu Z, Su C, Huang Q, Tian Q. Label decoupling framework for salient object detection. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Seattle, WA, USA. p. 13022–31. doi:10.1109/cvpr42600.2020.01304.
66. Hu X, Fu CW, Zhu L, Wang T, Heng PA. SAC-net: spatial attenuation context for salient object detection. IEEE Trans Circuits Syst Video Technol. 2021;31(3):1079–90. doi:10.1109/TCSVT.2020.2995220.
67. Ren Q, Lu S, Zhang J, Hu R. Salient object detection by fusing local and global contexts. IEEE Trans Multimed. 2021;23:1442–53. doi:10.1109/TMM.2020.2997178.
68. Xu B, Liang H, Liang R, Chen P. Locate globally, segment locally: a progressive architecture with knowledge review network for salient object detection. Proc AAAI Conf Artif Intell. 2021;35(4):3004–12. doi:10.1609/aaai.v35i4.16408.
69. Wang Y, Wang R, Fan X, Wang T, He X. Pixels, regions, and objects: multiple enhancement for salient object detection. In: Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2023 Jun 17–24; Vancouver, BC, Canada. p. 10031–40. doi:10.1109/CVPR52729.2023.00967.
70. Xie E, Wang W, Wang W, Ding M, Shen C, Luo P. Segmenting transparent objects in the wild. In: Computer vision-ECCV 2020. Cham, Switzerland: Springer International Publishing; 2020. p. 696–711. doi:10.1007/978-3-030-58601-0_41.

71. Mei H, Yang X, Wang Y, Liu Y, He S, Zhang Q, et al. Don't hit me! Glass detection in real-world scenes. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Seattle, WA, USA. p. 3684–93. doi:10.1109/cvpr42600.2020.00374.
72. He H, Li X, Cheng G, Shi J, Tong Y, Meng G, et al. Enhanced boundary learning for glass-like object segmentation. In: Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV); 2021 Oct 10–17; Montreal, QC, Canada. p. 15839–48. doi:10.1109/ICCV48922.2021.01556.
73. Yang X, Mei H, Xu K, Wei X, Yin B, Lau R. Where is my mirror? In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV); 2019 Oct 27–Nov 2; Seoul, Republic of Korea. p. 8808–17. doi:10.1109/iccv.2019.00890.
74. Le H, Vicente TFY, Nguyen V, Hoai M, Samaras D. A+D net: training a shadow detector with adversarial shadow attenuation. In: Computer vision-ECCV 2018. Cham, Switzerland: Springer International Publishing; 2018. p. 680–96. doi:10.1007/978-3-030-01216-8_41.
75. Zhu L, Deng Z, Hu X, Fu CW, Xu X, Qin J, et al. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In: Computer vision-ECCV 2018. Cham, Switzerland: Springer International Publishing; 2018. p. 122–37. doi:10.1007/978-3-030-01231-1_8.
76. Zheng Q, Qiao X, Cao Y, Lau RWH. Distraction-aware shadow detection. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15–20; Long Beach, CA, USA. p. 5162–71. doi:10.1109/CVPR.2019.00531.
77. Zhao X, Pang Y, Ji W, Sheng B, Zuo J, Zhang L, et al. Spider: a unified framework for context-dependent concept segmentation. arXiv:2405.01002. 2024.
78. Chen Z, Zhu L, Wan L, Wang S, Feng W, Heng PA. A multi-task mean teacher for semi-supervised shadow detection. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Seattle, WA, USA. p. 5610–9. doi:10.1109/cvpr42600.2020.00565.
79. Jie L, Zhang H. AdapterShadow: adapting segment anything model for shadow detection. arXiv:2311.08891. 2023.
80. Lin Y, Sun H, Liu N, Bian Y, Cen J, Zhou H. Attention guided network for salient object detection in optical remote sensing images. In: Artificial neural networks and machine learning-ICANN 2022. Cham, Switzerland: Springer International Publishing; 2022. p. 25–36. doi:10.1007/978-3-031-15919-0_3.
81. Feng D, Chen H, Liu S, Liao Z, Shen X, Xie Y, et al. Boundary-semantic collaborative guidance network with dual-stream feedback mechanism for salient object detection in optical remote sensing imagery. IEEE Trans Geosci Remote Sens. 2023;61:4706317. doi:10.1109/TGRS.2023.3332282.
82. 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 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021 Jun 20–25; Nashville, TN, USA. p. 10066–76. doi:10.1109/CVPR46437.2021.00994.
83. Lv Y, Zhang J, Dai Y, Li A, Liu B, Barnes N, et al. Simultaneously localize, segment and rank the camouflaged objects. In: Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021 Jun 20–25; Nashville, TN, USA. p. 11586–96. doi:10.1109/cvpr46437.2021.01142.
84. Cun X, Pun CM. Defocus blur detection via depth distillation. In: Computer vision-ECCV 2020. Cham, Switzerland: Springer International Publishing; 2020. p. 747–63. doi:10.1007/978-3-030-58601-0_44.
85. Jin Y, Qian M, Xiong J, Xue N, Xia GS. Depth and DOF cues make a better defocus blur detector. In: Proceedings of the 2023 IEEE International Conference on Multimedia and Expo (ICME); 2023 Jul 10–14; Brisbane, QLD, Australia. p. 882–7. doi:10.1109/ICME55011.2023.00156.