Salient Object Detection Based on Shadowed Sets and Illumination-Guided Network

Miao Li, *Graduate Student Member, IEEE*, Hongyun Zhang, Witold Pedrycz, *Life Fellow, IEEE*, Zhihua Wei, and Duoqian Miao

***Abstract*—Salient object detection (SOD) aims to distinguish salient regions from nonsalient ones in an image. In real-world scenarios, factors, such as depth variation and surface reflection, can interfere with the model's judgment, while illumination uncertainty further intensifies this interference. As a result, the uncertainty in salient boundary detection increases, leading to false or missed detections. To cope with uncertainty inherent to the problem, we introduce the concept of shadowed set, which is an effective method to process the uncertainty problem. In this article, we have designed an illumination-aware feature integration network by conducting dual-input feature integration under the implicit assistance of illumination maps. First, we devised a determination of pixel-level salient area module, which extracts illumination maps based on Retinex theory and obtains the main area of salient object based on shadowed set as the implicit feature of illumination. Next, we constructed a dual-modal compression module to solve the problem of feature alignment, which can use the dual-stream structure to process RGB and auxiliary inputs. Finally, multistage contextual complementary module can effectively recover fine object edges, and we use the outputs from the last three stages to supervise the training of the entire model. The state originality came from our previous work on illumination maps and shadowed sets, and we creatively combined them with the SOD to process uncertainty of salient area. The experiments demonstrate that our method exhibits excellent performance on multiple RGB-based datasets, at the same time, it also demonstrates unique performance on underwater and challenging scenes.**

***Index Terms*—Illumination map, Retinex theory, salient object detection (SOD), shadowed set, uncertainty analysis.**

I. Introduction

In recent years, deep learning has greatly improved the performance of salient object detection (SOD). However, in real-world applications, illumination variations, such as local shadows, reflections, highlights, and low-light regions, introduce uncertainty, leading the foreground and background to share similar features or show large differences in brightness distribution. These factors significantly affect the saliency discrimination process of conventional deep learning models. In particular, uncertain illumination conditions may degrade image quality and introduce redundant information, which further reduces the accuracy of salient edge localization and eventually results in false or missed detections.

To reduce the impact of illumination uncertainty on detection performance, most RGB-based deep learning methods focus on exploring deep features within a single image. Researchers have attempted to improve the model's ability to extract salient features under conditions, such as highlights, reflections, and local shadows by using stronger backbone networks and attention mechanisms [1], [2], [3], multiscale or cross-layer feature fusion [4], [5], and contrastive learning or data augmentation [6], [7], [8]. However, although RGB images contain rich information, their quality is greatly affected by lighting, shadows, and background variations, making it difficult for their features to reliably represent salient objects in real-world scenarios.

To address this issue, recent studies have explored the fusion of multiple auxiliary modalities, such as depth maps, infrared features, and edge information, to compensate for the degradation of RGB features under uncertain illumination conditions. These methods usually employ depth features to provide geometric and structural cues, infrared features to enhance object visibility in low-light scenes, and edge information to strengthen structural and texture priors, thereby improving model robustness. Most of these approaches rely on depth-guided fusion [9], [10], illumination-related hyperparameter guidance for RGB–infrared feature fusion [11], [12], [13], or edge enhancement mechanisms to refine the boundary representation of salient objects [14], [15], [16].

However, existing methods often fuse auxiliary information directly, resulting in a large amount of background redundancy that is unrelated to saliency. Under illumination variations, local shadows, and highlight conditions, such redundancy is often amplified by the network, leading to false detections, such as boundary adhesion, holes, and fragmentation. As shown in Fig. 1, each auxiliary modality has its own limitations: edge-based methods are easily affected by nontarget edges, depth maps may fail to accurately describe object structures when there are significant depth differences, and thermal imaging may cause confusion

Received 26 May 2025; revised 31 October 2025; accepted 18 November 2025. Date of publication 24 November 2025; date of current version 5 February 2026. This work was supported in part by the National Natural Science Foundation of China under Grant 62576247, Grant 62576251, Grant 62376198, and Grant 62376199 and in part by the National Key Research and Development Program of China under Grant 2022YFB3104700. Recommended by Associate Editor A. Albahri. *(Miao Li and Hongyun Zhang are the co-first authors.) (Corresponding author: Hongyun Zhang.)*

Miao Li, Hongyun Zhang, Zhihua Wei, and Duoqian Miao are with the School of Computer Science and Technology, Tongji University, Shanghai 201804, China (e-mail: lmiao@tongji.edu.cn; zhanghongyun@tongji.edu.cn; zhihua_wei@tongji.edu.cn; dqmiao@tongji.edu.cn).

Witold Pedrycz is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada, and also with the Systems Research Institute, Polish Academy of Sciences, 00-901 Warsaw, Poland (e-mail: wpedrycz@ualberta.ca).

The code is available online at https://github.com/limiaoair/Shadowed-set-SOD.

Digital Object Identifier 10.1109/TFUZZ.2025.3636203

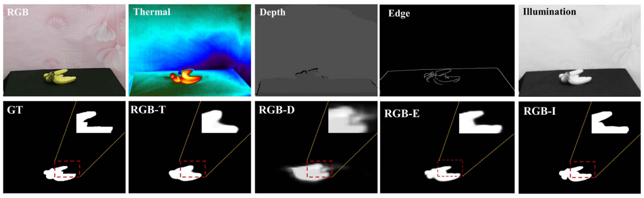

Fig. 1. RGB images with different auxiliary inputs and labels. GT, Thermal, Edge, Depth, and illumination map denote the label and different auxiliary input. The second line shows the results obtained using proposed model. We found that the edges of the results based on infrared and depth mapping were blurred. Through the magnified details in the results, the segmentation of the edge-based method are not as accurate as those based on RGB-I.

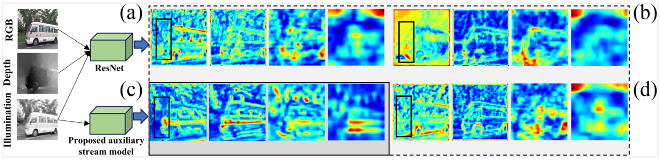

Fig. 2. Visualization of intermediate features extracted by the ResNet for different inputs. (a)–(d) Visualizations of intermediate features obtained by the ResNet module. (c) Visualizations of illumination maps using our auxiliary stream model.

between foreground and background due to heat crossover. To alleviate these issues, Li et al. [17] introduced illumination components and integrated them with RGB features, which effectively improved the model's sensitivity to illumination variations and mitigated the interference of background redundancy caused by uncertain lighting in saliency detection.

In addition, the aforementioned methods often model uncertainty implicitly and lack interpretable prior constraints, making them prone to overfitting or performance drift in complex illumination and low-contrast scenes. For example, some methods [1], [11], [18] treat the "edge uncertainty" in model outputs as a posterior feature to guide the segmentation of uncertain edges, rather than distinguishing salient and nonsalient regions at the input or intermediate feature levels. This causes the network to be easily disturbed by redundant background information during training and reduces its interpretability.

The main issue lies in the common assumption that "multimodality inherently brings improvement," without a mechanism to decide "where to fuse and where to suppress." In particular, under shadows, highlights, or under/overexposed regions, instability within auxiliary modalities propagates along the fusion path, amplifying prediction errors and resulting in blurred boundaries, missed detections, and false alarms. Furthermore, the absence of explicit modeling of illumination uncertainty makes it difficult to control its propagation near segmentation boundaries. Illumination uncertainty tends to appear first in the foreground–background transition zones, where the saliency membership becomes ambiguous. Without prior isolation of such "shadow regions," the network consumes excessive capacity adapting to global redundancy, which further amplifies noise interference. Therefore, constraining uncertainty propagation and controlling the fusion range in an interpretable manner is a key challenge in mitigating the impact of illumination uncertainty on SOD.

To mitigate the interference of redundant background information on the model during training and to enhance network interpretability, we introduce the shadowed-set theory to handle illumination maps based on Retinex theory in the input stage. Illumination maps not only provide information about how images reflect light, but also capture the intensity features of light, which can help the model more effectively identify salient objects. However, the redundant information of the illumination maps leads to the uncertainty of results in SOD. Locating the salient areas can effectively reduce the uncertainty of salient detection, and the shadowed set theory [19] is one of the important methods to solve the uncertainty aspect. In the proposed model, we obtain the salient areas by combining the illumination maps and the shadowed set theory, thereby assisting the model to detect the salient objects. In Fig. 1, RGB-I shows the result of using the illumination map as the auxiliary input and the proposed model can obtain more accurate salient edges.

We also design an expert network to learn from the processed illumination features. Most methods use a common pretrained backbone to extract the features. However, the performance of the backbone with fixed parameters cannot be guaranteed when facing different types of inputs. Taking ResNet as an example, for shallow-level information, the extracted features are not distinct [the first feature in Fig. 2(b) and (d)], and for deep-level information, ResNet tends to overextract features [the fourth/last feature in Fig. 2(b) and (d)].

Different from the previous methods, we consider a new solution to above challenges. First, we proposed the determination of salient area (DSA) module to automatically obtain the salient area by introducing the shadowed set theory. Second, we still use a pretrained module to extract RGB features, but we design a specially illumination branch module to extract illumination features. Then, we design dual-modal compression (DMC) module to integrate the features of the two modals, which can effectively learn illumination and RGB information. Finally, our decoding structure is a specially designed multistage contextual complementary module (CCM), where each stage contains a mini-U-Net. This design is inspired by the great success of U-Net in image segmentation, and we use this way to recover the salient maps.

Overall, the DMC is used to align and fuse RGB and illumination features by constructing a two-stream feature extraction backbone to obtain RGB and illumination features separately, while the CCM is primarily used to recover the edges of salient objects by a multistage process consisting of five modules to obtain the final result, while the output of the last three stages is used to oversee the training model. Our contributions can be summarized as follows.

1) We introduce the illumination maps and designed a novel DSA module to automatically obtain the salient area by introducing the shadowed set, which assists the model effectively detect salient objects and solves the uncertainty problem of salient detection.
2) We build a new auxiliary stream and DMC module to extract and fuse illumination features by specially designed

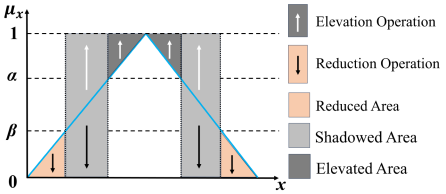

Fig. 3. Pedrycz's shadowed set. μ_s be the membership and x denotes the instance.

attention structures, and they can make full use of context information to align features at different levels.

3) We propose a multistage CCM that can fuses high-level and low-level features. In addition, the CCM combines a multistage loss function to recover the spatial locations of salient objects and reduce the introduction of noise.
4) Based on these modules, we propose an SOD method, which creatively combined the illumination maps and shadowed set. Experiments show that our method performs well in traditional SOD scenes and outperforms SOTA methods in challenging lighting conditions, such as underwater, low-light, and uneven light.

To this end, we have introduced an SOD algorithm that combines the illumination maps, inspired by the shadowed set theory. We have constructed an illumination extraction and matching module specifically designed for handling the illumination maps. To better recover the edges of the objects, we designed a multistage CCM. Furthermore, we have analyzed and discussed the feasibility of the model in complex scenes.

The rest of this article is organized as follows. In Section II, some prerequisites are presented. In Section III, the proposed model is extensively described. The experimental results are showcased in Section IV, which includes the analysis and discussion. Finally, Section V concludes this article.

II. Related Works

A. Identification of Salient Area Based on Shadowed Sets

Shadowed sets, proposed by Pedrycz [19], are constructed based on fuzzy sets, and develop a concise representation of fuzzy sets by fuzzy-rough transformation. As shown in Definition 1, X is a fuzzy set, and the shadowed set produce a new representation of the fuzzy set: $\mu_s : X \rightarrow \{0, [0, 1], 1\}$. 0 represents the instance that does not belong to X and 1 means that the element belongs to X. $[0, 1]$ quantifies the uncertainty of membership.

Definition 1: Let $S_S = (U', f)$ be a fuzzy information system, and $f : X \rightarrow [0, 1]$ be a fuzzy set of the domain U'. Given two parameters α and β, the elements whose membership values are greater than α in f are mapped to 1, the elements while membership smaller than β are mapped to 0, and the elements in the interval $[\alpha, \beta]$ are mapped to [0, 1]. The new mapping μ_s is referred to shadowed set, which can be represented as

$$\mu_s : X \overset{\mu_S(x)}{\rightarrow} \{0, 1, [0, 1]\}. \tag{1}$$

In Fig. 3, α and β determine the specific construction of the shadowed set. In essence, the fuzzy set has been transformed to a set with some clearly marked vagueness zones or, put it more descriptively, shadows. Pedrycz provided an objective function in the form of (2) to address a balance of vagueness, which can automatically construct the shadowed set representation of the fuzzy set. The optimal parameters α and β are calculated by minimizing $V_{\alpha,\beta}$

$$V_{\alpha,\beta}(\mu_s) = \Big| \int_{\mu_s(x)\geq\alpha} (1-\mu_s(x))\, \mathrm{d}x + \int_{\mu_s(x)\leq\beta} \mu_s(x)\mathrm{d}x - \int_{\beta<\mu_s(x)<\alpha} \mathrm{d}x \Big|. \tag{2}$$

In recent years, many methods based on traditional fuzzy sets and probabilities have been applied to SOD. Zhou et al. [20] introduced superpixels to divide the image into multiple blocks, and used the fuzzy set to measure the membership degree of the image blocks belonging to the salient areas. It distinguished the salient areas based on the membership degree using a determined threshold. Kapoor et al. [21] applied the fuzzy set to SOD to measure the membership degree of salient areas in image regions according to the characteristics of image brightness and color. Furthermore, Lu et al. [22] introduced a new probabilistic-based method, which first obtains the two membership values of sparse and dense superpixel areas. Then, they obtained the salient membership degree (SMD) of the image region by introducing Bayesian algorithm to integrate errors of two membership degree. Lastly, they use SMD to distinguish salient regions. The above methods combine the traditional uncertainty and deep learning methods, and use the obtained uncertainty membership degree values to obtain the rigid division about the image areas. Although traditional uncertainty-based methods can accomplish the task of SOD, they are the binary divisions with the precise value, which are sensitive to noise.

Compared with other methods based on conventional uncertainty, the shadowed set theory introduces the concept of "shadowed area," which reduces noise or uncertainty to enhance robustness. Inspired by the shadowed set, we can automatically determine the salient area in the illumination maps to guide the RGB image to locate the salient object in complex scenes by enhancing the features of salient areas. In this study, we proposed a special construction to obtain the salient area, which divides the illumination map into high-value, low-value, and salient area according to the value of pixel as given as follows:

$$U' : \begin{cases} x_{i,j} < \alpha, & \text{low} - \text{value} \quad \text{area} \\ \alpha \leq x_{i,j} \leq \beta, & \text{salient} \quad \text{area} \\ x_{i,j} > \beta, & \text{high} - \text{value} \quad \text{area} \end{cases} \tag{3}$$

where $U' = \{x_{i,j}; 0 < i \leq \text{width}, 0 < j \leq \text{height}\}$ be an illumination map, and width and height represent the values of width and height. $x_{i,j}$ be the value of pixel in illumination, which is normalized to [0,1]. i and j are the position of this pixel.

Based on the analysis of the dataset, we find that the parts in the image where the pixel value is higher than β are often the bright areas, such as the sky background. We call them high-value areas. The parts in an image where the pixels are lower than α

are low-light objects with low visibility. We call them low-value areas. In the middle part between α and β is the salient area that we really need. By locating the salient areas, it can effectively enhance the features of salient objects in RGB and improve the detection performance of our model.

Judging whether a pixel belongs to salient area is a typical uncertainty analysis problem. In this study, we propose an adaptive algorithm based on shadowed sets theory to obtain α and β. In this way, one can determine whether a pixel belongs to high-value, low-value, or salient area.

B. *Salient Object Detection*

SOD is one of the fundamental tasks in computer vision. Early SOD tasks often focused on manual feature extraction. Liu et al. [23] defined SOD as a pixel-level binary segmentation problem. Cheng et al. [24] proposed a model based on global contrast.

With the advancement of deep neural networks, an increasing number of deep learning methods have been proposed. Wei et al. [25] proposed a model capable of aggregating features from multiple layers, which adaptively selects complementary information for fusion, effectively avoiding information redundancy. Liu et al. [26], using a U-shaped architecture, introduced a significant improvement in SOD performance. Zhou et al. [27] designed an interactive two-stream decoder for salient features. In recent years, Liu et al. [28] analyzed the limitations of traditional model outputs and designed a new model that combines side-output features. Zhuge et al. [2] introduced the method of integrity learning, which can fully learn the position of all salient objects in the image, and it also has excellent effects for some special scenes, such as multiobjects and small objects. Han et al. [1] proposed an uncertainty-guided Transformer network, which effectively improved the robustness of the model in remote sensing scenarios. Yuan et al. [18] emphasized the importance of uncertainty modeling for edges and undersaturated regions, and achieved more accurate edge learning by using existing segmentation labels to identify uncertain regions in the prediction results. These methods are based on a single RGB image; however, RGB-based methods are often not enough to solve special scenes.

For this problem, more and more researchers are using multimodal auxiliary inputs to help detect salient objects, such depth and thermal maps. Tu et al. [29] designed a multi-interaction decoder for RGB-T saliency detection. Cong et al. [30] used dynamic parameters to guide the fusion of RGB and infrared features. Song et al. [11] used illumination information to guide the fusion of RGB and infrared features, but without uncertainty analysis and suppression of the auxiliary modalities, the model is still susceptible to interference in complex scenes. Although these methods use multimodal data and produce good performance, accurate depth or infrared maps still require additional equipment. Therefore, in this article, to better enhance the model's sensitivity to illumination, we use illumination components based on Retinex theory as auxiliary inputs to mitigate the segmentation edge uncertainty caused by illumination uncertainty.

III. Methodology

A. *Overall Pipeline*

We use an innovative illumination-aware approach inspired by researches on low-light image enhancement and shadowed set. The introduction of illumination map is used to address the difficulty in obtaining auxiliary input, and the introduction of shadowed set is mainly used to determine the salient area.

Fig. 4 shows the overall structure of proposed method. First, we extract illumination maps from images, and then utilize a DSA to locate the salient area. Then, using a dual-branch module for feature extraction and a DMC module for fusion. Subsequently, employing a multistage CCM method, we obtain the final prediction maps.

B. *Determination of Salient Area*

In a sense, each point on illumination map also approximatively represents the degree of membership of that element belonging to the object, the absorption and reflection abilities of objects with the same property to light are consistent.

Our aim is to use shadowed set to determine the salient area in the illumination map by calculating the values of α and β in (3). An overview is shown in Fig. 5(a). First, we utilize the RetinexNet [31] to acquire the illumination maps [see Fig. 5(b)]. Subsequently, each illumination map is fed into the DSA to obtain the salient area features, and DSA primarily consists of three main steps: Acquisition of α-cut, calculation of salient area, and feature concatenation.

Corollary 1: Let $U' = \{x_{m,n}\}$ be an illumination map, there is a parameter α such that when $x_{m,n}$ is greater than α, $x_{m,n}$ is set to zero and others are mapped from [0, α] to [0,1]. Then, the mapping μ'_s can be represent in (4). We refer to μ'_s as α-cut (U'_α) of the U'. α be boundary parameter of U'_α

$$\mu'_s : X \rightarrow \{0, [0, 1]\} . \quad (4)$$

Acquisition of α-cut (U'_α): As shown in Corollary 1, U'_α is obtained by dividing an illumination map with a fixed $\alpha \in [0, 1]$. Specifically, in Fig. 5(c), we use both histogram and image representations to introduce our methods. Let α be the fixed threshold and U^α be a matrix consist of α, where U^α have the same shape as U'. We compare the value of corresponding pixels between U' and U^α, if the pixels in U' are less than α, then this pixel is set to zero. Then, we get a α-cut (U'_α) of U^α with respect to threshold α.

In Section IV, we convert the value 0 into a very small parameter(0.01) to avoid gradient vanishing. The first step involves pixelwise processing of the illumination map to obtain the U'_α. ($\alpha \in \{\alpha_1, \ldots, \alpha_k\}$, where k is the number of threshold.)

Calculation of salient area: After getting a series of U'_α, we need to calculate two suitable thresholds to get salient areas between the two thresholds. In the shadowed set, Pedrycz proposed an optimization function to ensure a balance of uncertainty and obtain two reasonable thresholds: α and β. In this study, however, we calculate α_i and α_j to replace α and β in (3). Different from the balance of uncertainty, based on the proportion balance of salient area in the image, we propose a measurement strategy called the minimum difference algorithm (MDA) to

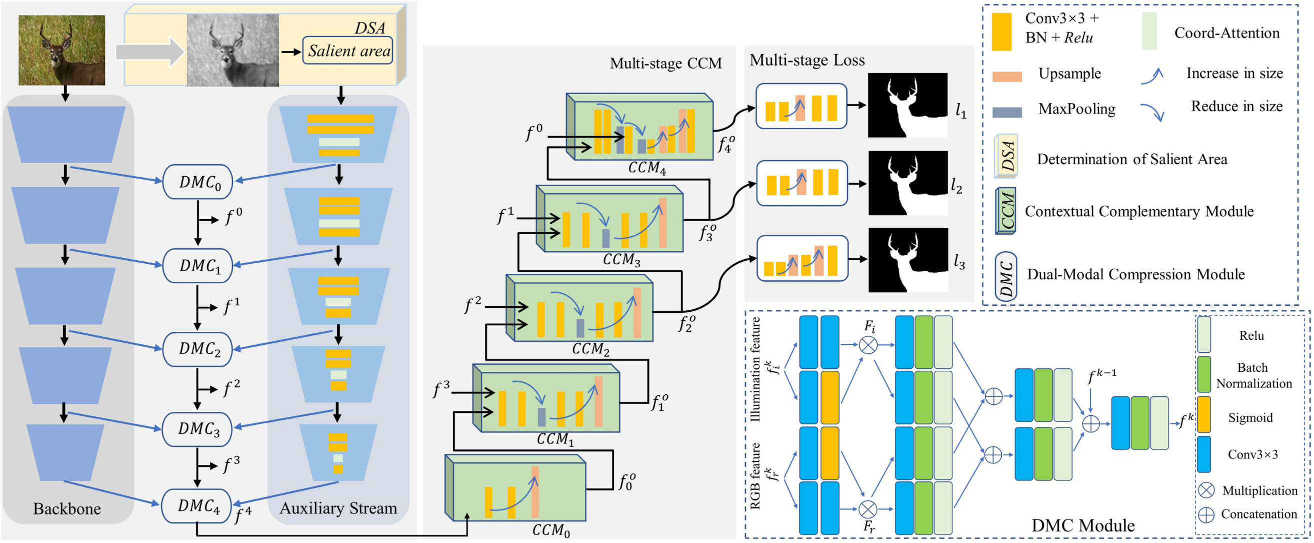

Fig. 4. Overall structure of proposed model. We obtain and process the illumination map through the DSA, and then feed the feature into designed auxiliary stream.

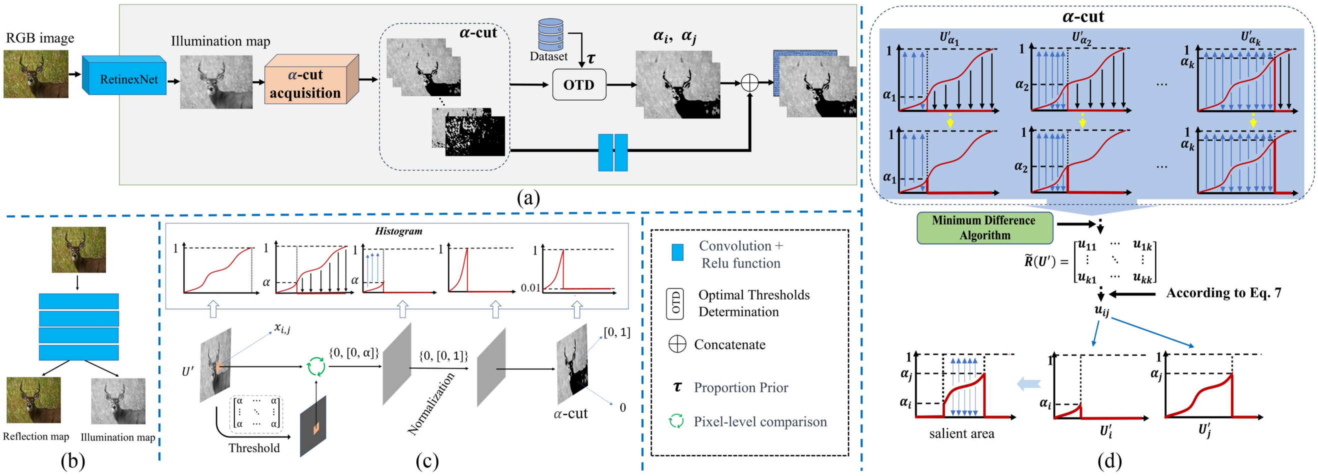

Fig. 5. Extraction and processing of illumination maps for DSA. (a) Pipeline of DSA. (b) RetinexNet, which is used to extract illumination maps. (c) α-cut acquisition, the top row denotes the pixel histogram corresponding to feature, the vertical axis indicating normalized pixel-values and the horizontal axis represents all pixels of the image. (d) Calculation the salient area, which sets multiple fixed thresholds and uses the MDA to select the optimal two thresholds.

obtain the optimal two thresholds α_i and α_j by comparing the interrelationships between different U'_{α}. First, we calculate value $u_{i,j}$ of MDA between U'_{α_i} and $U_{\alpha_j}{}'$ by the following equation:

$$u_{i,j}=\sqrt{\left(\left\|U'_{\alpha_i}\neq U'_{\alpha_j}\right\|/(w\times h)-\tau\right)^2}. \tag{5}$$

$\|U'_{\alpha_i}\neq U'_{\alpha_j}\|$ be used to count the number of different pixels between U'_{α_i} and $U_{\alpha_j}{}'$. The subscript i and j of $u_{i,j}$ corresponds subscript i and j of α_i and α_j. w and h be the width and height of U'_{α}. τ, as the prior knowledge based on the training dataset, is used to calculate $u_{i,j}$. First, we have counted all the images in DUTS-TR dataset [32] (about 11.7 k images) and found that the average proportion of salient objects in the images is 0.66, so we set $\tau=0.66$. In addition, we use the relation matrix $\widetilde{R}_{\mathbb{U}}$ defined in Corollary 2 to find the appropriate $u_{i,j}$, which is the smallest element in $\widetilde{R}_{\mathbb{U}}$. The smallest $u_{i,j}$ corresponds to the optimal α_i and α_j.

Corollary 2: Given a set $\mathbb{U}=\{U'_{\alpha_1},\ldots,U'_{\alpha_k}\}$, $u_{i,j}$ denotes the correlation between U'_{α_i} and $U_{\alpha_j}{}'$. The binary relation $\widetilde{R}$ for $\mathbb{U}$ can be expressed in (6). $\widetilde{R}_{\mathbb{U}}$ be a rivalry matrix, where each element is the correlation between two α-cuts

$$\widetilde{R}_{\mathbb{U}}=\begin{bmatrix}u_{11} & \cdots & u_{1k}\\ \vdots & \ddots & \vdots\\ u_{k1} & \cdots & u_{kk}\end{bmatrix}. \tag{6}$$

$\widetilde{R}_{\mathbb{U}}$ is a symmetric matrix that signifies the pairwise relationships between all elements on $\mathbb{U}$. Inspired by the equation of Pedrycz (2), we also explore the method for the two thresholds α_i and α_j. The subscripts (i,j) of the two thresholds can be

Algorithm 1: DSA Based on Improved Shadowed Set.

```
Input: RGB image $x$, $N$, $\epsilon$, $\tau$
Output: The output $f$ of DSA
Normalization $x$, $N \leftarrow 10$, $\varepsilon \leftarrow 0.01$
Compute illumination maps $x' = \{x_{m,n}\}$ of $x$ according to RetinexNet [31] /* (m,n) are the coordinates */
for $k = 1; k \leq N; k++$ do
    $\alpha_k = k/N$ /* The interval can be manually defined */
    for $x_{m,n}$ in $x'$ do
        if $x_{m,n} > \alpha_k$ then
            $x_{m,n} \leftarrow \varepsilon$ /* Avoid gradient disappearance */
        end if
        $U'_{\alpha_k} \leftarrow x_{m,n}$ /* Get the $\alpha$-cut */
    end for
    Normalization $U'_{\alpha_k}$
end for
$\mathbb{U} = \{U'_{\alpha_1}, U'_{\alpha_2}, \ldots, U'_{\alpha_{10}}\}$ /*take $k$=10 for example*/
Calculate $\widetilde{R}_{\mathbb{U}}$ according to Corollary. 2 and (5)
for $u_{i,j}$ in $\widetilde{R}_{\mathbb{U}}$ do
    if $u_{i,j} =$ min $\widetilde{R}_{\mathbb{U}}$ then
        Obtain $U_{\alpha_i}$ and $U_{\alpha_j}$ according to (7)
    end if
end for
$f_1 \leftarrow U_{\alpha_i}$, $f_2 \leftarrow U_{\alpha_j}$, $U'' = U' - \{U_{\alpha_i}, U_{\alpha_j}\}$
$f_3 \leftarrow U_{\alpha_1}$
for $U_{\alpha_k}$ in $U''$ do
    $f_3 \leftarrow f_3 \oplus U_{\alpha_k}$ /*Concatenate $U_{\alpha_k}$ and $f_3$ */
end for
Calculate $f_3$ by the two Convolution layers
$f \leftarrow f_1 \oplus f_2 \oplus f_3$
return $f$
```

determined in the following way:

$$\arg\min_{i,j} \mu_{i,j}. \tag{7}$$

In summary, Pedrycz used two to-be-determined thresholds to convert some elements to certainty while extending the uncertainty to balance the overall uncertainty by (2). As shown in Fig. 5(d), being different from Definition 1, our approach involves a differential analysis of a series of U'_α to obtain the image-adapted thresholds, and consider the differential portion between these two maps as the salient area.

Feature concatenation: The salient area is the part defined by the difference between two selected α-cut, while the remaining U'_{α_k} also contain valuable information. Owing to the powerful feature extraction capabilities of the convolution neural network, we concatenate all the U'_{α_k} and transform them into a single-channel feature map u_f through a two CNN layers, the processing of this step is shown in Fig. 5(a). By three steps, we obtain a three-channel feature as a output of *DSA:* Output $= \text{Concat}([u_i, u_j, u_f])$. We also present the pseudocode for the DSA module, as shown in Algorithm. 1.

C. Backbone and Auxiliary Stream

In previous methods, when dealing with RGB and auxiliary inputs, it was common to use the same backbone for feature extraction. However, for different auxiliary inputs, independent information extraction modules are necessary. In this article, we follow the approach of previous methods by using the ResNet-50 [33] to extract RGB features. Simultaneously, we have designed a unique auxiliary stream for our illumination information. The role of our auxiliary stream (*A-stream*) is to encode the positional information of the target. Therefore, we introduce an attention module proposed by Li et al. [7] to learn contextual information from the illumination data.

D. DMC Module

To extract diverse information of RGB and illumination map, we have designed two different branch, distinguishing ourselves from previous methods: the main backbone and the auxiliary stream module. As shown in Fig. 4, pretrained ResNet-50 is used for the backbone to obtain the five RGB features, auxiliary stream is built to generate five corresponding illumination features. We employ the coord-attention to focus on the position of objects of illumination features. Finally, these five sets of features are fed into DMC.

Fig. 4-DMC shows the main process of DMC. Let f_r^k and f_i^k denote the RGB and illumination feature derived by backbone or auxiliary stream, where k is the number of backbone. Specifically, the preliminary feature $F_{r,i}$ can be determined as follows:

$$F_{r,i} = \text{Conv}^2\left(f_{r,i}^k\right) \otimes \text{Sig}\left(\text{Conv}^1\left(f_{r,i}^k\right)\right) \tag{8}$$

where Conv^i denotes convolution with i layers, $\otimes$ denotes element-by-element multiplication, and Sig is the Sigmoid function. Each $F_{r,i}$ will receive two transitional features by two branches. The four transitional features can be expressed as $F_{r,i}^m\{m = 1, 2\} = \Gamma(F_{r,i})$, where $\Gamma(*)$ denotes the convolution with batch normalization and Relu function. The final output f_k is obtained by cross-fusing $F_{r,i}^m$ and f_{k-1}

$$f_k = \Gamma\left(\Gamma\left(F_r^1 \oplus F_i^2\right) \oplus \Gamma\left(F_r^2 \oplus F_i^1\right) \oplus f_{k-1}\right) \tag{9}$$

where $\oplus$ denote concatenation. The above describes the feature fusion method for the k output in DMC. When $k = 0$, we only fuse the RGB and illumination maps.

E. Multistage CCM

In Fig. 4, let $f^k (k \in 0, 1, 2, 3, 4)$ denote the final output obtained by DMC. The multistage CCM consists of five CCMs (CCM_k). Except for CCM_0, each CCM_k has two inputs and one output. The output of each CCM_k, except for CCM_4, is fed into the next CCM_{k+1}. In detail, let the output of CCM_k is expressed as f_k^o. CCM_0 consists of two convolution layers with batch normalization and Relu. CCM_1, CCM_2, and CCM_3 are relatively more complex, these modules involve two upsampling and one downsampling operations, and contextual features need to be integrated. The formula can be shown as

$$f_{4-k}^o = \text{Up}\left(\Gamma^2\left(\text{Do}\left(\Gamma^2\left(f^k \oplus f_{3-k}^o\right)\right)\right)\right) \tag{10}$$

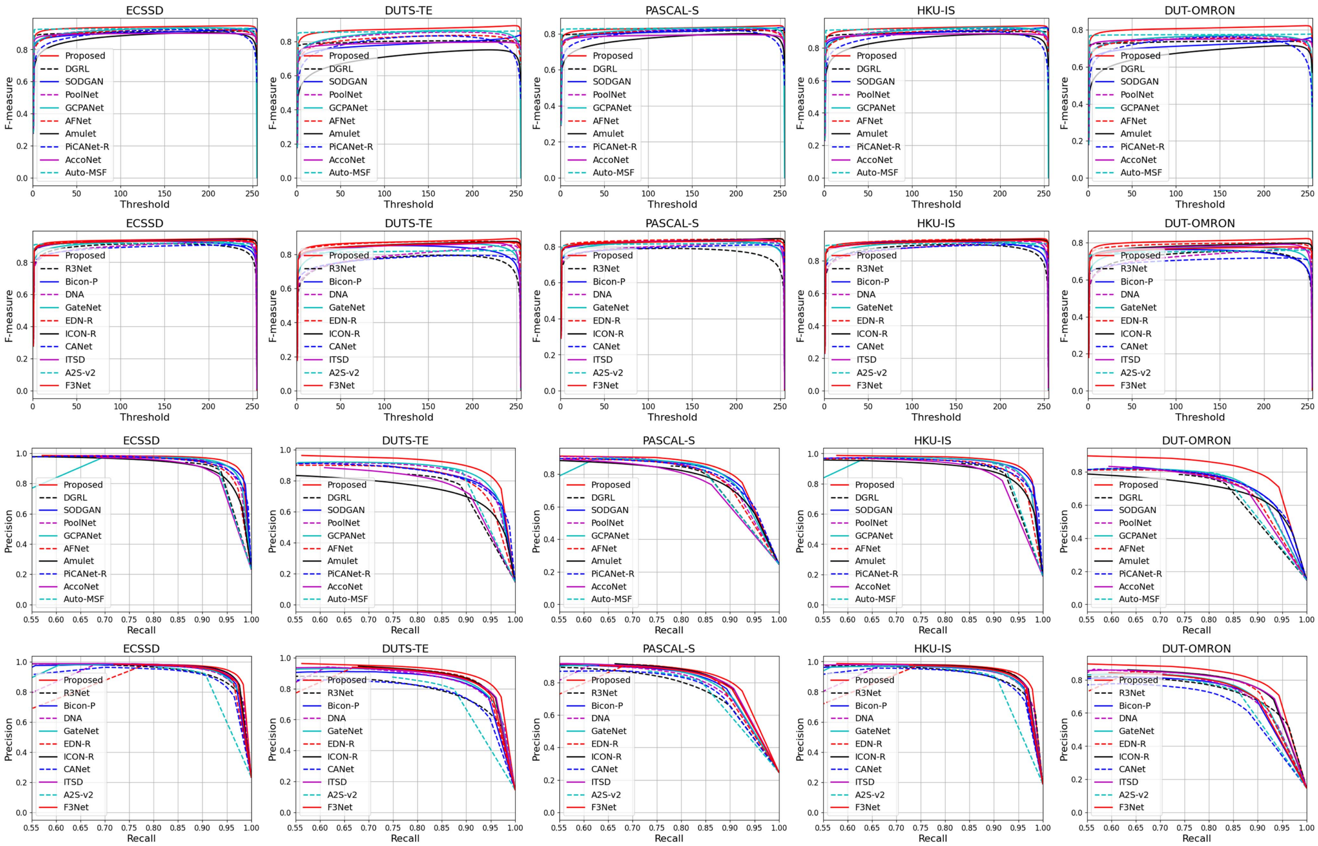

Fig. 6. PR and Fm curves compared proposed method with others on five datasets. To provide better clarity, we divided the methods into two groups.

where $k \in 0, 1, 2, 3$ and Γ^2 represents a two-layer convolution with batch normalization and ReLU function, $\mathrm{Up}(*)$ and $\mathrm{Do}(*)$ denotes the upsampling and downsampling. Specially, the structure of CCM_4 is composed of two upsampling and downsampling, the process can be expressed as

$$f_4^o = \Gamma\left(\mathrm{Up}\Gamma^2\left(P\Gamma\left(\Gamma^2\left(f_3^o \oplus f^0\right)\right)\right)\right) \tag{11}$$

where $P\Gamma(*)$ represents Γ followed by a pooling layer, $\mathrm{Up}\Gamma$ denotes Γ followed by an upsampling layer, and $\mathrm{Up}\Gamma^2$ denote two layer $\mathrm{Up}\Gamma$. Finally, the output is three features: f_2^o, f_3^o, and f_4^o, whose size are $\frac{H}{4} \times \frac{W}{4}$, $\frac{H}{2} \times \frac{W}{2}$, and $\frac{H}{2} \times \frac{W}{2}$, respectively. These three results need to be reconstructed as segmentation maps, the minimum size of the output is $\frac{H}{4} \times \frac{W}{4}$ rather than a smaller size to reduce subsequent upsampling operations.

F. Multistage Loss Function

We designed a multistage supervised network to avoid the introduction of noise. First, three upsampling modules are used to recover the previously mentioned f_2^o, f_3^o, and f_4^o, resulting in corresponding prediction maps (pre1, pre2, and pre3) from three scales. Proposed method employs three loss functions with equal weights to supervise the results. Our loss function can be expressed in the following form:

$$\mathrm{Loss}_{\mathrm{total}} = L_{\mathrm{pre1}} + L_{\mathrm{pre2}} + L_{\mathrm{pre3}} \tag{12}$$

where L can be expressed as $L_* = L_{\text{w-IoU}} + L_{\text{w-BCE}}$. Inspired by Wei et al. [25], the weight binary cross entropy (w-BCE) and weighted intersection over union loss (w-IoU) are used for this study. IoU and BCE have already been widely used in image segmentation, and these metrics are also extensively employed in SOD tasks [25].

IV. Experiments

A. Experiment Setup

1) Implementation Details: We implement proposed method on the Linux operation system with Python 3.7. The experiments are performed on Ubuntu 20.04 with Nvidia RTX 3090 GPU. We utilize the Adam optimizer with the parameters $\beta_1 = 0.9$ and $\beta_2 = 0.1$ and with learning rate $le = 0.0001$. The Backbone is ResNet-50 and DUTS-TR [32] as the training set, which has 10 553 pair of images. Before training, we get the illumination maps by RetinexNet (see [31]).

2) Datasets and Metrics: In this study, we employ multiple classic datasets to assess the differences between various methods, including ECSSD [45], HKU-IS [46], PASCAL [47], DUT-OMRON [48], and DUTS-TE [32]. These datasets contains 1000, 4447, 850, 5167, and 5019 images, respectively. Meanwhile, we employ four metrics to assess performance among comparative methods, including mean absolute error (MAE, the smaller the better), mean F-measure [49] (m-F), mean E-measure [50] (m-E), and S-measure [51] (Sm).

TABLE I
PERFORMANCE COMPARISON BETWEEN PROPOSED METHOD AND THE OTHERS ON FIVE DATASETS

		ECSSD				DUTS-TE				PASCAL-S				HKU-IS				DUT-OMRON			
Methods	Pub/Year	MAE	m-F	m-E	Sm	MAE	m-F	m-E	Sm	MAE	m-F	m-E	Sm	MAE	m-F	m-E	Sm	MAE	m-F	m-E	Sm
Amulet [34]	ICCV$_{17}$	.0592	.8727	.9104	.8936	.0846	.7015	.8156	.8198	.0980	.7672	.8349	.8936	.0521	.8453	.9071	.8832	.0977	.6677	.7934	.7804
DGRL [35]	CVPR$_{18}$	.0408	.9049	.9418	.9027	.0495	.7968	.8929	.8420	.0793	.8072	.8743	.8362	.0355	.8906	.9438	.8946	.0618	.7328	.8493	.8060
PiCANet-R [36]	TIP$_{18}$	.0465	.8902	.9252	.9168	.0498	.7887	.8747	.8480	.0768	.7922	.8536	.8538	.0433	.8658	.9161	.9042	.0648	.7233	.8294	.8302
R3Net [37]	IJCAI$_{18}$	.0563	.8905	.9132	.9028	.0661	.7560	.8452	.8355	.1050	.7648	.8069	.8088	.0483	.8660	.9092	.8019	.0711	.7188	.8263	.8173
AFNet [38]	CVPR$_{19}$	.0418	.9059	.9355	.9135	.0453	.8107	.8909	.8494	.0760	.8073	.8662	.9135	.0358	.8878	.9343	.9052	.0573	.7397	.8424	.8254
F3Net [25]	AAAI$_{20}$	.0332	.9246	.9478	.9242	.0453	.8107	.8909	.8885	.0685	.8216	.8779	.8598	.0280	.9096	.9515	.9171	.0526	.7637	.8611	.8382
GCPANet [39]	AAAI$_{20}$	.0356	.9124	.9420	.9267	.0378	.8358	.9067	.8901	.0678	.8148	.8763	**.8655**	.0316	.8976	.9419	.9203	.0566	.7488	.8468	.8375
GateNet [40]	ECCV$_{20}$	.0418	.9034	.9310	.9197	.0448	.8139	.8893	.8700	.0741	.8104	.8643	.8569	.0360	.8891	.9322	.9151	.0613	.7292	.8357	.8374
ITSD [27]	CVPR$_{20}$	.0346	.9208	.9470	.9249	.0408	.8390	.9125	.8851	.0728	.8194	.8774	.8614	.0307	.9035	.9472	.9173	.0608	.7672	.8640	.8404
Auto-MSF [6]	ACM MM$_{22}$	.0334	.9265	.9508	.9145	**.0344**	.8548	**.9274**	.8772	.0705	.8273	.8806	.8516	.0273	.9117	**.9559**	.9084	.0495	.7718	.8727	.8322
PoolNet [26]	TPAMI$_{22}$	.0388	.9141	.9400	.9208	.0397	.8297	.9011	.8826	.0795	.8107	.8617	.8499	.0321	.8990	.9422	.9167	.0554	.7491	.8476	.8343
Bicon-P [15]	PR$_{22}$	.0358	.9187	.9490	.9174	.0418	.8262	.8888	.8714	.0772	.8130	.8624	.8511	.0351	.8956	.9282	.9094	.0582	.7413	.8359	.8232
SODGAN [6]	ACM MM$_{22}$	.0389	.9004	.9438	.9129	.0530	.7808	.8870	.8491	.0700	.8071	.8823	.8502	.0324	.8909	.9492	.9039	.0768	.7121	.8333	.8024
CANet [5]	TCyb$_{22}$	.0492	.8844	.9257	.8979	.0556	.7659	.8703	.8398	.0862	.7894	.8612	.8293	.0401	.8707	.9296	.8954	.0705	.6995	.8298	.8398
DNA [28]	TCyb$_{22}$	.0427	.8941	.9345	.9147	.0464	.7869	.8881	.8603	.0836	.7938	.8576	.8363	.0360	.8735	.9362	.9051	.0630	.7204	.8441	.8185
EDN-R [3]	TIP$_{22}$	.0328	.9270	.9482	.9267	.0352	.8596	.9216	**.8925**	.0699	**.8286**	.8764	**.8629**	.0271	.9163	.9523	**.9242**	**.0497**	**.7846**	.8744	**.8496**
A2S-V2 [41]	CVPR$_{23}$	.0441	.9144	.9366	.8937	.0468	.8143	.9010	.8426	.0796	.8125	.8712	.8297	.0365	.9014	.9494	.8902	.0609	.7500	.8636	.8122
AccoNet [7]	TCyb$_{23}$	.0415	.9076	.9365	.8886	.0557	.7925	.8808	.8327	.0805	.8006	.8592	.8174	.0405	.8841	.9297	.8759	.0604	.7606	.8640	.8126
ICON-R [2]	TPAMI$_{23}$	**.0318**	.9279	**.9542**	**.9290**	.0370	.8529	.9229	.8888	.0706	.8261	.8806	.8621	.0289	.9121	.9534	.9202	.0569	.7790	**.8751**	.8444
HybridSOD [8]	TCSVT$_{23}$	.0517	.8908	.9236	.8851	.0501	.7919	.8893	.8366	.0820	.8005	.8641	.8274	.0386	.8821	.9354	.8873	-	-	-	-
FPSI [42]	PR$_{24}$	.036	.907	-	-	.041	.820	-	-	.069	.808	-	-	.029	.898	-	-	.054	.743	-	-
VRF [43]	TCSVT$_{24}$	.037	**.928**	-	-	.037	**.861**	-	-	.071	.859	-	-	.032	**.919**	-	-	**.033**	**.873**	-	-
DC-Net [44]	PR$_{25}$	.034	-	.945	.924	.035	-	**.927**	**.896**	**.066**	-	**.892**	.857	**.027**	-	.954	.924	.053	-	**.876**	**.849**
R-Net [14]	ESWA$_{25}$	**.0303**	**.9307**	**.9534**	**.9310**	**.0336**	**.8607**	.9234	.8920	**.0680**	**.8300**	**.8846**	.8580	**.0261**	**.9179**	**.9558**	**.9256**	.0516	.7689	.8625	.8320
Proposed	-	**.0275**	**.9345**	**.9596**	**.9345**	**.0313**	**.8685**	**.9365**	**.9004**	**.0664**	**.8270**	**.8862**	**.8655**	**.0254**	**.9186**	**.9605**	**.9322**	**.0452**	**.8047**	**.8930**	**.8634**

The bold red is the best result, blue are the next best results, and the green are the third levels.

B. Comparison With the State-of-The-Art Methods

In this part, we have selected 24 state-of-the-art methods to compare the results, including Amulet [34], DGRL [35], GCPANet [39], PiCANet-R [36], PoolNet [26], R3Net [37], Bicon-P [15], SODGAN [6], DNA [28], GateNet [40], HybridSOD [8], CANet [5], ITSD [27], A2S-v2 [41], AccoNet [7], CANet [5], ICON-R [2], EDN-R [3], Auto-MSF [6], F3Net [25], FPSI [42], VRF [43], DC-Net [44], and R-Net [14]. The salient maps of these methods are provided by the public code, these models are all excellent ones in recent years.

1) Quantitative Comparison: In Table I, our method consistently performs well on the *Sm*, this is a significant improvement because, apart from MAE metric, Sm is better at measuring the structural. For the *Fm* of the PASCAL-S and the *Fm m-E* of the HKU-IS, R-Net and DC-Net show superior results, but the advantages are not as prominent (difference less than about 0.01). The main reason is that these methods focus on the positioning of salient objects within the model as our approach does. In addition, the VRF outperforms all methods on both MAE and Fm on DUT-OMRON, mainly because it uses a scribble-based approach that uses specialized branches to locate the salient objects. Although it improves performance, complex visual generation modules need to be designed. The recursive decoder is designed in R-Net, and the attention maps and prediction maps are obtained at the same time, which can also achieve better results.

Fig. 6 presents the precision-recall (PR) and F-measure curves generated by our method and 20 excellent methods on five datasets. To provide clarity, we randomly divided the 20 methods into two groups and placed them in two comparison chart. From the Fig. 6, it can be observed that our method achieves the highest precision at certain recall levels. In addition, our method also obtains excellent F-scores.

2) Visual Result Analysis: Fig. 7 present the results of salient maps generated by 20 methods for five categories of scenes, including similar scene, complex background, multiobjects scene with different depth, and multiobjects scene in low-light and high-light scenes. It can be observed that our method exhibits relatively good performance in the localization and segmentation of salient objects. In particular, our method outperforms others in multiple object detection tasks. For the "dolphin" in the third image, many methods exhibit instances of missed detection, such as SODGAN [6] and AFNet [38], while our approach accurately detects its position. For scenes with extreme lighting, our approach also shows excellent results. For the low-light environment in the fourth image of Fig. 7, our method is able to segment the left person contour in the dark scene. Methods, such as A2S-v2 and R3Net, contain blurry edges. Similarly, our method still has a good effect on high-light scenes. One of the main reasons why our method can get better results in extreme light environments is that our method is an illumination-guided model, which extracts illumination information and accommodates it into the model to increase the sensitivity of the algorithm to light, so as to avoid interference from low-light or high-light environments.

Furthermore, while various methods can locate the object, such as the second set of examples in Fig. 7, the majority of them produce results with blurry edges, especially HybridSOD [8], PiCANet-R [36] and DNA [28]. To better illustrate the results for target edges, we introduce difference value-colormaps to visualize the edges of the prediction. As shown in Fig. 8, we compare the results obtained from all methods to the ground truth, creating difference maps. Subsequently, we use the "*COOL*" color space to generate the corresponding color maps. We can observe that AFNet, HybridSOD, R3Net, and AccoNet contain more artifacts. Our method, along with ICON-R and F3Net, performs better in object edge detection.

3) Statistical Significance Analysis: To test whether there is a statistical difference in segmentation performance compared with the recent SOTA methods, we chose to compare with

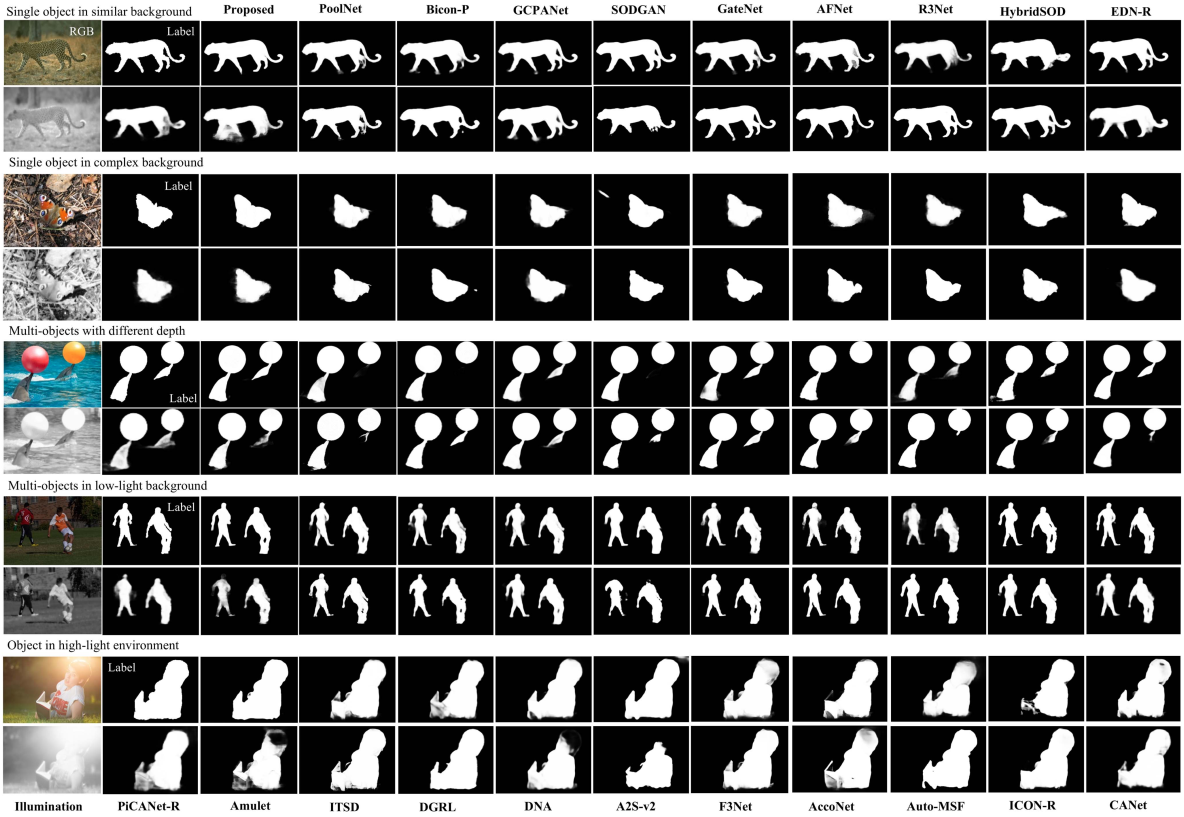
Single object in similar background
Proposed
PoolNet
Bicon-P
GCPANet
SODGAN
GateNet
AFNet
R3Net
HybridSOD
EDN-R
RGB
Label
Single object in complex background
Label
Multi-objects with different depth
Label
Multi-objects in low-light background
Label
Object in high-light environment
Label
Illumination
PiCANet-R
Amulet
ITSD
DGRL
DNA
A2S-v2
F3Net
AccoNet
Auto-MSF
ICON-R
CANet

Fig. 7. Subjective comparison of the experimental results among SOTA methods. Every two lines represent different predictions on the same test image.

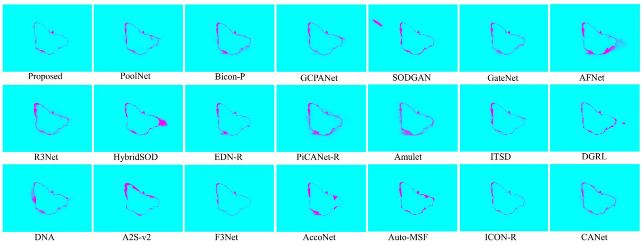

Fig. 8. Edge comparison of the results among SOTA methods. *Cool* color space is selected to show the differences between ground truth and predictions.

TABLE II
Average Ranking and Statistical Test of Six Methods Under Five Datasets

	Proposed	R-Net	ICON-R	EDN-R	Auto-MSF	AccoNet	Prob>Chi-sq
MAE	1	2.4	4.6	3.4	3.6	6	.0007
m-F	1.6	2.4	4	2.8	4.2	6	.0035
m-E	1	3.4	3	4.6	3	5.8	.0021
Sm	1	3.2	3.4	2.6	4.8	6	.0006

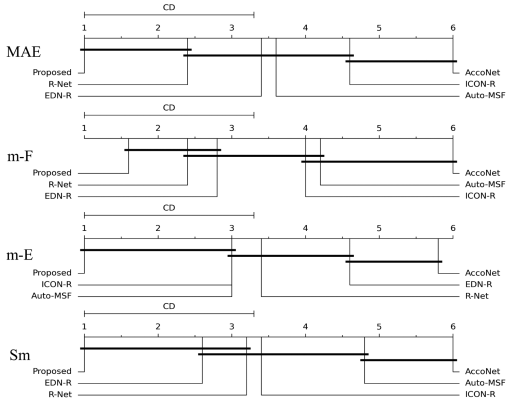

Fig. 9. CD diagrams of proposed method and 20 compared methods on four metrics across five datasets.

five SOTA methods. The Friedman test is first adopted at the significant level $P = 0.05$. Table II shows the results of Friedman test, and all the p-values are much less than 0.05. So these methods have significant differences in the four metrics.

Therefore, we set up the Nemenyi post hoc test to determine the substantial differences between any two methods. The critical distance of Nemenyi test is obtained by $\mathrm{CD} = q_{\alpha}\sqrt{\frac{k(k+1)}{6N}}$, where $q_{\alpha=0.05} = 2.850$ when $k = 6$ and $N = 5$. When the test distances of the two methods exceed $\mathrm{CD} = 3.3722$, it indicates that there is a significant difference. Fig. 9 shows our critical difference (CD) diagram. The results indicate that our method outperforms the other methods in all four metrics of the five datasets.

C. Model Ablation Analysis

To ensure the accuracy of ablation experiments, all methods are trained on DUTS-TR [32] and tested on ECSSD [45].

TABLE III
PERFORMANCE COMPARISON BETWEEN PROPOSED METHOD AND ABLATION EXPERIMENTS ON ECSSD [45]

A-stream	DSA	2-f	3-f	MAE	Fm	Em	Sm	Var(Median)
			✓	.0387	.9145	.9423	.9317	.00331(.0164)
	✓		✓	.0314	.9274	.9517	.9295	.00303(.0166)
✓			✓	.0304	.9284	.9526	.9310	.00292(.0155)
✓	✓	✓		.0304	.9279	.9543	.9319	.00261(.0154)
✓	✓		✓	**.0275**	**.9345**	**.9596**	**.9345**	**.00256(.0119)**

The bold values denote the best results.

We added the variance and median score of MAE to measure the statistical information. Lower median values correlate with decreased noise and variance, demonstrating improved result robustness. In addition, we also give the parameter number of the model, which is shown in the Appendix of the Supplementary Material.

1) Effectiveness of DSA and Auxiliary Stream: To demonstrate the effectiveness of (DSA), we trained two versions, "w DSA" and "w/o DSA." In version w/o DSA, the illumination maps are directly fed into the auxiliary stream without undergoing processing in module DSA. In the original method, as shown in Fig. 5(a), the structure involved learning two features from the α-cut with different granularity thresholds, while the remaining features were extracted using a convolution neural network into a single feature. This means that the DSA module outputted an illumination map with three channels. Now, we introduce an ablation experiment (2-f) where we only obtain features with two channels in the DSA module (all remaining thresholds are discarded). 3-f denote three channels in compared method. To assess the role of auxiliary stream (A-stream), we replaced the auxiliary stream with ResNet-50 similar to backbone and compared their effects. In addition, 2-f only makes sense when the DSA module is used. The ablation results are presented in Table III.

In Table III, the first column indicates whether the proposed auxiliary stream is used, the second column represents the DSA, and the third and fourth columns denote the number of channels in the illumination features obtained by DSA. We can observe that the DSA is beneficial for performance improvement regardless of whether our proposed auxiliary stream is used. When our model does not use the auxiliary stream, the model becomes an RGB-based method. The results in the second and third rows indicate that the performance can be effectively improved after introducing the illumination maps and the auxiliary stream.

2) Impacts of the Threshold Selection: In DSA, we designed the MDA algorithm, which utilizes the shadowed set theory to automatically calculate the appropriate thresholds. These thresholds can effectively determine the locations of salient areas and different thresholds will affect the model's location of salient areas. To better describe the robustness of our MDA, we conducted experiments with different thresholds, the results are shown in Table IV. In Table IV, we changed τ in (5) to achieve the purpose of changing the α_i and α_j. For the convenience of the experiment, we set 0.2, 0.4, 0.6, and 0.8 as comparative thresholds. In addition, we added 0.01 and 0.09 threshold experiments and we found that our choice was indeed the most suitable (It has improved by 11% compared with the second choice). The experiment verified that when the threshold changes, it can affect the performance of the model and lead to poor results.

TABLE IV
PERFORMANCE COMPARISON BETWEEN DIFFERENT τ

Metrics	τ=0.01	τ=0.2	τ=0.4	τ=0.6	τ=0.8	τ=0.99	Our choice$_{\tau=0.66}$
MAE	.0565	.0422	.0324	<u>.0309</u>	.0315	.0608	**.0275**
Fm	.7818	.8123	.8517	<u>.9041</u>	.8938	.7375	**.9345**
Em	.8601	.8942	.9022	<u>.9252</u>	.9109	.8161	**.9596**
Sm	.7921	.8456	.8889	<u>.9127</u>	.8963	.7704	**.9345**
Median	.0443	.0387	<u>.0296</u>	.0299	.0307	.0509	**.0119**
Var	.00323	.00357	.00343	<u>.00311</u>	.00368	.00409	**.00256**

The best and second results are in bold and underscored.

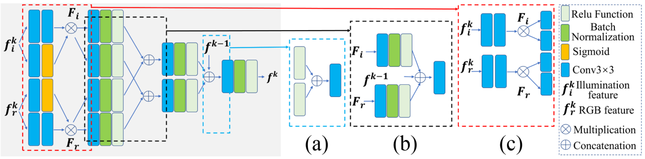

Fig. 10. Comparison of DMC module with other ablation strategies. (a) DMC, the red box indicates the part we want to verify. (b) DMC without cross-fusion. (c) DMC without feature fusion. (d) DMC without the skip connection from the last DMC.

TABLE V
PERFORMANCE COMPARISON BETWEEN PROPOSED METHOD AND ABLATION EXPERIMENTS OF DMC

Metrics	Cross	Fusion	Skip	Cross-T	Dynamic	Concat	Proposed
MAE	<u>.0296</u>	.0297	.0308	.0311	.0322	.3430	**.0275**
Fm	.9309	.9283	.9279	<u>.9321</u>	.9303	.9068	**.9345**
Em	.9556	.9557	.9539	**.9601**	.9523	.9436	<u>.9596</u>
Sm	.9324	.9340	<u>.9325</u>	.9225	.9300	.9117	**.9345**
Median	.0137	<u>.0124</u>	.0135	.0146	.0144	.0151	**.0119**
Var	<u>.00255</u>	.00262	.00269	.00274	.00269	.00303	**.00256**

Bold and underlined are optimal and suboptimal results.

3) Effectiveness of DMC: DMC is designed for fusing RGB and illumination features, whose purpose is to perform cross-fusion while introducing contextual information. Fig. 10 displays the ablation structures regarding the DMC, and we validate the structures of its three components through experiments. Meanwhile, we present the objective results of various methods in Table V. Skip, cross, and Fusion represent the structure just like Fig. 10(a)–(c). Cross-T and Concat represent the use of cross-transformer with multiple attention and direct concatenation instead of DMC, and Dynamic denotes the use of the dynamic parameter fusion structure proposed by as Cong et al. [30] by computing light intensity. The results show that DMC can lead the compared ablation methods in 3 metrics, the Em ranks is second level (only 0.5%0 behind the first place). Although the Cross-T method can achieve the optimum in Em, the remaining indicators all lag behind our method.

4) Effectiveness of CCM: For multistage CCM, there are two keys in this part: the pooling layer of each CCM$_i$ (i=1, 2, 3, 4) and the contextual features from the DMC. The pooling layers and contextual features can deeply mine features, which are conducive to restoring fine edges and improving the robustness of the model. In Table VI, CCM outperforms the ablation method in all four metrics. Meanwhile, the variance and median of the results also demonstrate the stability of the images used in the test.

TABLE VI
PERFORMANCE COMPARISON BETWEEN PROPOSED METHOD AND ABLATION EXPERIMENTS OF CCM

Methods	MAE	Fm	Em	Sm	Var(Median)
w/o contextual features	.0338	.9180	.9510	.9233	.00288(.0166)
w/o pooling layers	.0309	.9258	.9532	.9312	.00298(.0238)
w/o three outputs	.0313	.9158	.949	.9237	.00291(.0163)
Proposed	**.0275**	**.9345**	**.9596**	**.9345**	**.00256(.0119)**

The bold values denote the best results.

TABLE VII
OBJECTIVE PERFORMANCE OF DIFFERENT METHODS ON USOD10K [52]

Metrics	HAINet [54]	D3Net [10]	HQSOD [55]	SVAM-Net [56]	CSNet [57]	Proposed
Pub	TIP_{21}	TNNLS_{21}	ICCV_{21}	RSS_{22}	TPAMI_{22}	-
MAE	.0279	.0374	.0552	.0915	.0548	**.0242**
m-F	.8937	.8408	.7677	.6251	.7825	**.8957**
m-E	.9466	.9107	.8267	.7466	.8652	**.9536**
Sm	.9123	.8931	.8111	.7465	.8559	**.9129**

The best and second results are in bold and underscored.

TABLE VIII
OBJECTIVE RESULTS ON VDT-CHALLENGE DATASET [9]

	V-SA		V-SSO		V-BSO		V-LI		V-SI		V-NI	
Methods	m-F	m-E	m-F	m-E	m-F	m-E	m-F	m-E	m-F	m-E	m-F	m-E
EDN	.785	.911	.916	.987	.681	.908	.706	.899	.633	.845	.588	.836
MIDD	.711	.896	.913	.986	.434	.675	.702	.889	.709	.893	.590	.793
HAINet	.844	**.972**	.940	.991	**.683**	.902	.791	.962	.791	.958	.667	.908
proposed	**.858**	.971	**.944**	**.996**	.680	**.911**	**.808**	**.977**	**.863**	**.980**	**.700**	**.947**

The best and second results are in bold and underscored.

D. Extended Applications and Experiments

1) Extended Application in Underwater Scene: As our method can effectively mitigate the impact of environmental lighting on images, we attempted experiments in more complex scenarios. Taking an underwater scene as an example, we conducted experiments using the existing USOD10 K dataset [52] and compared the results with eight state-of-the-art methods. The specific results are shown in Table VII. Our method demonstrates excellent results in underwater image datasets. In Table VII, our method can achieve the optimal performance. Compared with methods specifically used for underwater image detection such as HQSOD, our method can also outperform it in four metrics.

2) Extended Application in Challenging Scene: As a common challenging scene, strong and low light scene often contains a lot of noise, so we have carried out special experiments for this kind of scene. The fourth and fifth sets of images in Fig. 7 show a subjective comparison of low-light and strong scenes. In addition, we also give the test index results of the proposed method under challenging scenarios. We tested with VDT challenging dataset [9], which contains data from six challenging scenes. We select V-SA (similar appearance), V-SSO (small salient object), V-LI (low illumination), V-SI (side illumination), V-NI (no illumination), and V-BSO (big object) to verify proposed model with compared methods EDN [3], MIDD [29] and HAINet [53], these three methods are the SOTA methods based on RGB, RGBT, and RGBD, respectively. Table VIII presents the experimental results. We find that our method can achieve the optimum on the m-F, with only a small gap in V-SA and V-BSO on m-E metrics.

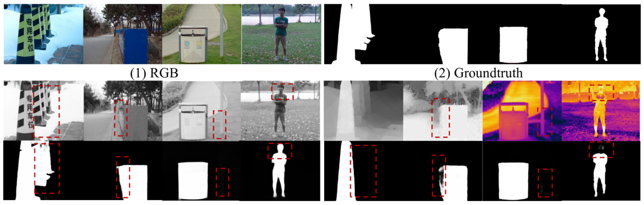
(1) RGB
(2) Groundtruth
(3) Illumination maps and the results
(4) Multi-modal maps and the results

Fig. 11. Compared results on multimodal inputs.

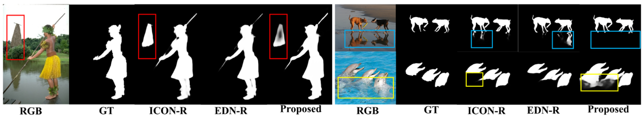

Fig. 12. Limitations of proposed and compared methods.

Fig. 11
shows the visual results. The problem of depth maps is that the results are fuzzy in the face of objects with different depth of field. The problem with infrared maps as an input is that thermal cross (in the second line sample), which leads to missed detection or missing objects. Proposed method cannot only locate the position of the objects, but also effectively reduce the interference of background details by smoothing the background (such as the first line sample and the head of the second image).

E. Analysis and Discussion

1) Limitations: To explore the underlying meanings and roles among the models, we conducted a thorough analysis of the results. Fig. 12 illustrates the comparisons between our method and the other two top-performing methods, highlighting two main characteristics of our approach. First, it can still detect and recognize locally salient objects. This is a drawback, as shown in the comparison image in the first row of Fig. 12. Our method still segments distant building (even though it is not on the label image). Second, when faced with an object that has both underwater and above-water parts, it exhibits edge blurriness, as demonstrated in the comparison image in the third row of Fig. 12.

2) Discussion: For the current issues, we have considered two possible reasons. First, in Fig. 12, ICON-R [2] employs a mechanism of integrity learning, with a strong focus on modeling contextual features, which enables it to discover all salient objects in the image. Therefore, for the image in the first row, it can recognize and detect distant building, as these buildings are highly salient relative to the local background. As for our method, we have embraced this idea extensively, not only in the DMC module but also in the subsequent CCM module, where we use skip-feature connections to introduce context information. Consequently, some characteristics of building are retained. However, this does introduce a visual flaw, which the EDN method excels in addressing.

EDN-R [3] employs an extremely downsampled model, utilizing downsampling to achieve a larger receptive field within the model. This not only accurately locates salient objects but also avoids false positives caused by "local saliency." The result of EDN-R [3] in Fig. 12 does not consider distant buildings as salient objects (global aspect). This is another reason why we introduced downsampling in the CCM module.

Second, we will analyze the reason leading to the blurriness of objects on the surface of water. The comparison results in the second and third rows of Fig. 12 demonstrate the performance of SOD in scenes with water surfaces. In the image of the second row, our method accurately segments the salient object without being affected by shadows, whereas ICON and EDN are influenced by the reflection of water. The main reason is the introduction of the illumination maps based on Retinex theory, which is a method that utilizes the attribute of the object to extract illumination information. Regarding reflections on the water surface, they are not real objects but "false objects," and our method can accurately identify them. However, in the comparison results of the third row, blurriness appears in our method. The primary reason is that the blurred part under the water is still a part of the "real object." Due to the unevenness of the water surface and light refraction, Retinex theory extracts incorrect illumination information when dealing with the object under such conditions. For the Retinex theory, underwater dolphins are still considered a part of the entire object. Therefore, the model will "try" to recover the position and edges of the object, leading to the blurriness at the boundaries.

3) Analysis of Potential Disadvantages of CCM: While the multistage resampling strategy in the CCM module improves the network's ability to capture edge structures, it may also increase the risk of noise amplification or minor information loss due to interpolation and quantization effects. In our current design, we suppress possible noise and lost semantic information through advance monitoring mechanisms and context-jumping links. Table V shows the results of our approach when contextual feature skip is not used and advance loss supervision (three outputs) is not applicable. To verify the possibility of noise statistically, we add the variance and maximum and minimum values of all data results. However, further investigation into more robust resampling techniques could be explored in future work.

V. Conclusion

This article presents a novel SOD method that leverages the Retinex theory and shadowed set theory to extract illumination features. All the modules we designed are aimed at better extracting and fusing illumination maps and RGB features. Due to the uniqueness of Retinex theory, our method performs exceptionally well in underwater scenarios.

Furthermore, we will explore more applications of shadowed sets in the field of artificial intelligence. By using shadowed set theory and deep learning, we focus on unbalanced lighting scenes and camouflage scenes, and fully explore the uncertainty of salient area. For the data of multimodal, such as infrared images, which are auxiliary maps based on temperature, it is a very meaningful work to apply the shadowed set on the temperature-aware model, and it can more effectively solve the effects of light, such as low visibility, color degradation, and others.

References

[1] P. Han, J. Huang, J. Yang, and X. Li, "Uncertainty-guided Siamese transformer network for salient object detection," *Expert Syst. Appl.*, vol. 272, 2025, Art. no. 126690.

[2] M. Zhuge, D.-P. Fan, N. Liu, D. Zhang, D. Xu, and L. Shao, "Salient object detection via integrity learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3738–3752, Mar. 2023.

[3] Y. Wu, Y. Liu, L. Zhang, M. Cheng, and B. Ren, "EDN: Salient object detection via extremely-downsampled network," *IEEE Trans. Image Process.*, vol. 31, pp. 3125–3136, 2022.

[4] C. Jiang, Y. Liu, J. Sun, J. Guo, and W. Lu, "Illumination-based adaptive saliency detection network through fusion of multi-source features," *J. Vis. Commun. Image Representation*, vol. 79, 2021, Art. no. 103192.

[5] J. Li, Z. Pan, Q. Liu, Y. Cui, and Y. Sun, "Complementarity-aware attention network for salient object detection," *IEEE Trans. Cybern.*, vol. 52, no. 2, pp. 873–886, Feb. 2022.

[6] Z. Wu et al., "Synthetic data supervised salient object detection," in *Proc. 30th ACM Int. Conf. Multimedia*, 2022, pp. 5557–5565.

[7] G. Li, Z. Liu, D. Zeng, W. Lin, and H. Ling, "Adjacent context coordination network for salient object detection in optical remote sensing images," *IEEE Trans. Cybern.*, vol. 53, no. 1, pp. 526–538, Jan. 2023.

[8] R. Cong et al., "A weakly supervised learning framework for salient object detection via hybrid labels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 2, pp. 534–548, Feb. 2023.

[9] K. Song, J. Wang, Y. Bao, L. Huang, and Y. Yan, "A novel visible-depth-thermal image dataset of salient object detection for robotic visual perception," *IEEE/ASME Trans. Mechatron.*, vol. 28, no. 3, pp. 1558–1569, Jun. 2023.

[10] D. Fan, Z. Lin, Z. Zhang, M. Zhu, and M. Cheng, "Rethinking RGB-D salient object detection: Models, data sets, and large-scale benchmarks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 5, pp. 2075–2089, May 2021.

[11] K. Song et al., "SIA: RGB-T salient object detection network with salient-illumination awareness," *Opt. Lasers Eng.*, vol. 172, 2024, Art. no. 107842.

[12] R. Cong et al., "Does thermal really always matter for RGB-T salient object detection?," *IEEE Trans. Multimedia*, vol. 25, pp. 6971–6982, 2023.

[13] P. Lyu, P.-H. Yeung, X. Yu, X. Cheng, C. Wu, and J. C. Rajapakse, "Efficient fourier filtering network with contrastive learning for UAV-based unaligned bi-modal salient object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5003312.

[14] H. Wang et al., "R-Net: Recursive decoder with edge refinement network for salient object detection," *Expert Syst. Appl.*, vol. 261, 2025, Art. no. 125562.

[15] Z. Yang, S. Soltanian-Zadeh, and S. Farsiu, "BiconNet: An edge-preserved connectivity-based approach for salient object detection," *Pattern Recognit.*, vol. 121, 2022, Art. no. 108231.

[16] C. An, K. Song, L. Bao, D. Zhao, Z. Zhou, and Y. Yan, "A novel edge detection method of blade with multisupervision for Foreground-background confusion caused by extreme illumination," *IEEE Sensors J.*, vol. 24, no. 18, pp. 29429–29440, Sep. 2024.

[17] M. Li, H. Zhang, K. Cai, W. Pedrycz, D. Miao, and Y. Gao, "IFA: Illumination-aware feature aggregation model for salient object detection," *Pattern Recognit.*, vol. 171, 2025, Art. no. 112118.

[18] Y. Yuan, P. Gao, Q. Dai, J. Qin, and W. Xiang, "Uncertainty-guided refinement for fine-grained salient object detection," *IEEE Trans. Image Process.*, vol. 34, pp. 2301–2314, 2025.

[19] W. Pedrycz, "Shadowed sets: Representing and processing fuzzy sets," *IEEE Trans. Systems, Man, Cybern., Part B (Cybern.)*, vol. 28, no. 1, pp. 103–109, Feb. 1998.

[20] Y. Zhou, A. Mao, S. Huo, J. Lei, and S.-Y. Kung, "Salient object detection via fuzzy theory and object-level enhancement," *IEEE Trans. Multimedia*, vol. 21, no. 1, pp. 74–85, Jan. 2019.

[21] A. Kapoor, K. Biswas, and M. Hanmandlu, "Information set based approach for salient object detection," in *Proc. 5th Nat. Conf. Comput. Vis., Pattern Recognit., Image Process. Graph.*, 2015, pp. 1–4.

[22] H. Lu, X. Li, L. Zhang, X. Ruan, and M.-H. Yang, "Dense and sparse reconstruction error based saliency descriptor," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1592–1603, Apr. 2016.

[23] T. Liu et al., "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 353–367, Feb. 2011.

[24] M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, Mar. 2015.

[25] J. Wei, S. Wang, and Q. Huang, “F^{3}Net: Fusion, feedback and focus for salient object detection,” in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 12321–12328.

[26] J. Liu, Q. Hou, Z. Liu, and M. Cheng, “Poolnet+: Exploring the potential of pooling for salient object detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 887–904, Jan. 2023.

[27] H. Zhou, X. Xie, J.-H. Lai, Z. Chen, and L. Yang, “Interactive two-stream decoder for accurate and fast saliency detection,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 9141–9150.

[28] Y. Liu, M.-M. Cheng, X.-Y. Zhang, G.-Y. Nie, and M. Wang, “DNA: Deeply supervised nonlinear aggregation for salient object detection,” *IEEE Trans. Cybern.*, vol. 52, no. 7, pp. 6131–6142, Jul. 2022.

[29] Z. Tu, Z. Li, C. Li, Y. Lang, and J. Tang, “Multi-interactive dual-decoder for RGB-thermal salient object detection,” *IEEE Trans. Image Process.*, vol. 30, pp. 5678–5691, 2021.

[30] R. Cong et al., “Does thermal really always matter for RGB-T salient object detection?,” *IEEE Trans. Multimedia*, vol. 25, pp. 6971–6982, 2023.

[31] C. Wei, W. Wang, W. Yang, and J. Liu, “Deep Retinex decomposition for low-light enhancement,” in *Proc. Brit. Mach. Vis. Conf. 2018*, 2018.

[32] L. Wang et al., “Learning to detect salient objects with image-level supervision,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 136–145.

[33] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[34] P. Zhang, D. Wang, H. Lu, H. Wang, and X. Ruan, “Amulet: Aggregating multi-level convolutional features for salient object detection,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 202–211.

[35] T. Wang et al., “Detect globally, refine locally: A novel approach to saliency detection,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3127–3135.

[36] N. Liu, J. Han, and M.-H. Yang, “PiCANet: Pixel-wise contextual attention learning for accurate saliency detection,” *IEEE Trans. Image Process.*, vol. 29, pp. 6438–6451, 2020.

[37] Z. Deng et al., “R ^{3} Net: Recurrent residual refinement network for saliency detection,” in *Proc. 27th Int. Joint Conf. Artif. Intell.*, 2018, pp. 684–690.

[38] M. Feng, H. Lu, and E. Ding, “Attentive feedback network for boundary-aware salient object detection,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1623–1632.

[39] Z. Chen, Q. Xu, R. Cong, and Q. Huang, “Global context-aware progressive aggregation network for salient object detection,” in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 10599–10606.

[40] X. Zhao, Y. Pang, L. Zhang, H. Lu, and L. Zhang, “Suppress and balance: A simple gated network for salient object detection,” in *Proc. 16th Eur. Conf. Comput. Vis.*, 2020, pp. 35–51.

[41] H. Zhou, B. Qiao, L. Yang, J. Lai, and X. Xie, “Texture-guided saliency distilling for unsupervised salient object detection,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 7257–7267.

[42] X. Wang, Z. Liu, V. Liesaputra, and Z. Huang, “Feature specific progressive improvement for salient object detection,” *Pattern Recognit.*, vol. 147, 2024, Art. no. 110085.

[43] B. Xu, H. Liang, W. Gong, R. Liang, and P. Chen, “A visual representation-guided framework with global affinity for weakly supervised salient object detection,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 1, pp. 248–259, Jan. 2024.

[44] J. Zhu, X. Qin, and A. Elsaddik, “DC-Net: Divide-and-conquer for salient object detection,” *Pattern Recognit.*, vol. 157, 2025, Art. no. 110903.

[45] Q. Yan, L. Xu, J. Shi, and J. Jia, “Hierarchical saliency detection,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 1155–1162.

[46] G. Li and Y. Yu, “Visual saliency based on multiscale deep features,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 5455–5463.

[47] Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille, “The secrets of salient object segmentation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 280–287.

[48] C. Yang, L. Zhang, H. Lu, X. Ruan, and M. Yang, “Saliency detection via graph-based manifold ranking,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 3166–3173.

[49] A. Borji, M. Cheng, H. Jiang, and J. Li, “Salient object detection: A benchmark,” *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5706–5722, Dec. 2015.

[50] F. Dengping, G. Cheng, C. Yang, R. Bo, C. Mingming, and B. Ali, “Enhanced-alignment measure for binary foreground map evaluation,” in *Proc. 27th Int. Joint Conf. Artif. Intell.*, 2018, pp. 698–704.

[51] D. Fan, M. Cheng, Y. Liu, T. Li, and A. Borji, “Structure-measure: A new way to evaluate foreground maps,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 4558–4567.

[52] L. Hong, X. Wang, G. Zhang, and M. Zhao, “USOD10K: A new benchmark dataset for underwater salient object detection,” *IEEE Trans. Image Process.*, vol. 34, pp. 1602–1615, 2025.

[53] G. Li, Z. Liu, M. Chen, Z. Bai, W. Lin, and H. Ling, “Hierarchical alternate interaction network for RGB-D salient object detection,” *IEEE Trans. Image Process.*, vol. 30, pp. 3528–3542, 2021.

[54] L. Bo, T. Lv, Y. Zhong, S. Ding, and M. Song, “Disentangled high quality salient object detection,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 3560–3570.

[55] M. J. Islam, R. Wang, and J. Sattar, “SVAM: Saliency-guided visual attention modeling by autonomous underwater robot,” in *Proc. Robotics: Sci. Syst. XVIII*, K. Hauser, D. A. Shell, and S. Huang, Eds., 2022.

[56] M.-M. Cheng, S.-H. Gao, A. Borji, Y.-Q. Tan, Z. Lin, and M. Wang, “A highly efficient model to study the semantics of salient object detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 8006–8021, Nov. 2022.

Miao Li (Graduate Student Member, IEEE) received the master’s degree in computer science from the School of Information science and Engineering, Yunnan University, Kunming, China, in 2022. He is currently working toward the Ph.D. degree in computer sciences with Tongji University, Shanghai, China.

He spent a year as a joint Ph.D. Student with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. His research interests include computer vision and pattern recognition, image enhancement, salient object detection, and image segmentation.

Dr. Li was the receipient of the funding from the China Scholarship Council in 2024 and

Hongyun Zhang received the Ph.D. degree in pattern recognition and intelligence system from Tongji University, Shanghai, China, in 2005.

She is a Doctoral Supervisor and is currently an Associate Professor with Tongji University. She is the author or coauthor of nearly 70 journal papers and conference proceedings in Principal curves, pattern recognition, granular computing, and rough set. Her current research interests include computer vision and pattern recognition, principal curves, data mining, rough set theory, and granular computing.

Witold Pedrycz (Life Fellow, IEEE) is currently a Professor with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada, the Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah, Saudi Arabia, and the Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland. He has authored 17 research monographs covering various aspects of computational intelligence, data mining, and software engineering. He has authored or coauthored numerous papers in his research field, which include, in particular, granular computing and computational intelligence.

Prof. Pedrycz was the recipient of the IEEE Canada Computer Engineering Medal, Cajastur Prize for Soft Computing from the European Centre for Soft Computing, Killam Prize, and Fuzzy Pioneer Award from the IEEE Computational Intelligence Society. He is intensively involved in editorial activities. He is the Editor-in-Chief of *WIREs Data Mining and Knowledge Discovery* (Wiley) and *the International Journal of Granular Computing* (Springer). He is a Fellow of the Royal Society of Canada and a Foreign Member of the Polish Academy of Sciences.

Zhihua Wei received the dual Doctoral degree from Tongji University, Shanghai, China, and the University of Lyon 2, Lyon, France.

She is currently a Professor and Doctoral Supervisor with the School of Computer Science and Technology. She has authored or coauthored more than 100 papers in top international conferences, such as CVPR, IJCAI, and WWW, and international journals, such as TPAMI, and has published two textbooks. In recent years, she has presided over seven national-level projects, including projects under the National Key Research and Development Program and the National Natural Science Foundation of China. Her research interests include natural language processing, multimodal content analysis and generation, etc.

Dr. Wei was the recipient of second prize of Shanghai Science and Technology Progress Award and the second prize of Wu Wenjun Artificial Intelligence Natural Science Award.

Duoqian Miao is currently a Professor with the College of Electronics and Information Engineering, Tongji University, Shanghai, China. He works with the Department of Computer Science and Technology, Tongji University. He has authored or coauthored more than 180 papers in this area, more than nine books and academic works, and nine national invention patents. His research interests include artificial intelligence, machine learning, Big Data analysis, granular computing and rough sets.

Prof. Miao is a Fellow of International Rough Set Society and the Director of Chinese Association for Artificial Intelligence. He is currently the Associate Editor of *Information Sciences*, *CAAI Transactions on Intelligence Technology*, and *International Journal of Approximate Reasoning*, and so on.