Optik - International Journal for Light and Electron Optics 269 (2022) 169944

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.com/locate/ijleo

# Content-aware dynamic filter salient object detection network in multispectral polarimetric imagery

Suining Gao [a,b,c], Xiubin Yang [a,b,c,*], Li Jiang [d], Ziming Tu [a,b,c], Mo Wu [a,b,c], Zongqiang Fu [a,b,c]

[a] Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China
[c] Key Laboratory of Space-Based Dynamics and Rapid Optical Imaging Technology, Chinese Academy of Sciences, Changchun 130033, China
[d] School of Physics,Changchun University of Science and Technology, Changchun 130022, China

ARTICLE INFO

*Keywords:*
Salient object detection
Multispectral polarimetric imagery
Dynamic filter model

ABSTRACT

Salient object detection (SOD) is widely applied in image segmentation, image fusion, and adaptive compression. However, the SOD of visible images in complex scenes remains a prominent problem due to the lack of low-level features. To solve this problem, a Content-aware Dynamic Filter salient object detection Network using visible and polarized mask images is proposed. It can use the prior information on polarization dimension to guide the SOD of visual features. First, to extract information from the MSPI, a salient polarization mask M composed of three channels is generated. Secondly, deep fused features of the M and RGB images are generated by Encoder and DenseNet fusion structures with receptive fields. Finally, the fused features guide the decoder to generate saliency maps through the content-aware dynamic filtering model. The indexes of the salient object detection results given in this paper are superior to the state-of-the-art algorithms, especially for objects in dim, low contrast, or high transparency and other complex scenarios.

## 1. Introduction

Salient object detection aims to identify the most attractive regions in images automatically just like how human visual systems. It is a common and necessary pre-processing step in various machine vision applications such as image segmentation [1], target localization [2], and image fusion [3], therefore has led to considerable research [4,5]. In general, traditional salient object detection methods determine the local contrast of image regions with their surroundings through features of color and intensity. Those methods are mainly based on objects' uniqueness and compactness [6,7]. Moreover, the boundary and connection priors can provide more clues for the saliency detection of central objects [8]. The saliency detection algorithm based on deep learning improves the capability for extracting features of traditional algorithms. It uses networks [9] to learn multi-scale deep-scale features to break the limits of traditional algorithms and improve the accuracy of saliency detection [10–12]. Although great progress has been made in this field, the problems of complex scenes detections are still not well resolved. This is that the existing saliency detection algorithms focus merely on

*Abbreviations:* SOD, Salient object detection; MSPI, Multispectral polarization images; CDFNet, Content-aware dynamic filter salient object detection network; CDPM, Content-aware dynamic pyramid model; DPR, Dilated pyramid refinement.

* Correspondence to: Changchun Institute of Optics, Fine Mechanics and Physics, Changchun, 130033, China.
*E-mail address:* yangxiubin@ciomp.ac.cn (X. Yang).

https://doi.org/10.1016/j.ijleo.2022.169944
Received 16 May 2022; Received in revised form 1 September 2022; Accepted 5 September 2022
Available online 8 September 2022

the features of visible images which lack front-background differences [13]. To solve this problem, some works[14–16] introduce depth data as an aid to further improve the detection performance. Since depth information can express objects more intuitively, some progress has been made in these works. However, when objects are similar to their surroundings both in space and distance, algorithms based on RGB-D models tend to fail to distinguish them [17].

The problems can be solved substantially if multispectral polarization images (MSPI) are used. The superior recognition ability of MSPI has been well demonstrated in applications such as object detection [18], transparent object detection [19], and background segmentation [20]. MSPI can obtain intrinsic properties of objects, such as degree of polarization, roughness, and other special performance [21,22]. These intrinsic properties are more robust to the environment, which is more stable for the detection of complex objects and can reduce the difficulty of identification [23,24]. For example, in natural scenes, the degree of polarization of artificial or camouflaged targets is significantly higher than that of the background, and their angular polarization will be smoother and more continuous similarly. These features won't vary with changes in light intensity and shooting angle greatly [25]. Therefore, the polarization information in the MSPI image can be a good supplement to the visible information, which increases the algorithm's ability to recognize complex targets and helps to improve the detection accuracy of salient objects. However, MSPI has only been applied in some specific machine vision fields, few works are adopting MSPI for salient object detection currently. How to use MSPI for salient object detection is worth further exploration.

MSPI images are analyzed from the color space and polarization dimensions, which are roughly divided into four categories: 1) The target is easy to distinguish in both dimensions; 2) The target is easy to distinguish in the spectral dimension; 3) The target is easy to distinguish in the polarization dimension; 4) The polarization of the local area of the target is easy to distinguish, and the color of the remaining part is easy to distinguish. For the first three cases, if the single-model saliency detection method is used applied, the correct detection results can be obtained from at least one dimension. However, the single model will fail for the fourth case. If a more reasonable saliency value can be selected from the two dimensions and fused into the final saliency map, the best detection result can be obtained.

Inspired by the analysis, this paper proposed a CDFNet (Content-aware Dynamic Filter salient object detection Network) based on RGB-M data, which performs saliency detection on the RGB image and the polarization mask M respectively, and then fuses the two features stream into Network to generate a high-precision MSPI saliency map. Our main contributions are summarized as follows:

1. A simple and effective preprocessing algorithm is proposed for multi-spectral polarized image information extraction. It can generate a three-channel mask image that can be used directly as input to a deep learning network.
2. This paper presents a multi-feature fusion structure based on a content-aware dynamic pyramid model. Because the coordinate attention weight of visual features and the deep mask features are used as the convolution coefficient of dynamic filtering, the effect of the polarization mask can be well used to highlight the boundary of objects.
3. We constructed different types of polarized salient object detection datasets, and compared our algorithm with 8 state-of-the-art models to verify the feasibility of CDFNet. At the same time, we also verified the practicability and superiority of our proposed mask.

## 2. Related works

### 2.1. Salient object detection

Salient object detection technology usually assumes that the intensity of foreground pixels in multispectral images is different from that of other pixels [26]. For the detection of prominent areas of multispectral polarization, it can be defined as the segmentation of foreground and background:

$$O(x,y) = \begin{cases} 1 & \text{if } \mathrm{S}(x,y,\lambda,o_i) > \text{threshold}, \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where is the predicted prospect, $\mathrm{S}(x,y,\lambda,o_i)$ is the significant value in coordinates, $\lambda$ is spectral length, $o_i$is polarization direction, $threshold = E(\mathrm{S}(x,y,\lambda,o_i))$ is the average value of significance map. In this paper, the significant partition variable space adds a polarized dimension to the conventional CIE-Lab space.

As we all know, visible light images can provide rich details, while depth images contain contour information of objects, which are complementary to some extent. Therefore, many RGB-D SOD algorithms adopt element addition [27] or concatenation [28] to fuse these two features, or directly use 3 d network structure [29] to perform multidimensional feature extraction. However, these algorithms all depend on a large number of fixed coefficients, which leads to poor universality. Moreover, the polarization mask M, which pays more attention to image edge details, is also difficult to apply to the following algorithms. In this way, dynamic filtering [30,31] proves to be the best choice for polarization-guided salient object detection.

The concept of dynamic filtering was originally put forward for recognition and video detection to enhance the expression of self-features. Later, multi-scale dynamic filter layer [32] was developed. Based on this, a dynamic extended pyramid model [33] is proposed and applied to RGB-D salient target recognition. The core idea is to make use of RGB-D fusion features to adaptively adjust the convolution kernel coefficients of different position features as to boost the effect of guiding filtering. This efficient and concise network has achieved good results in the significance test of RGB-D. However, the disadvantages lie in that the parameters of convolution layer depend entirely on the fusion feature, and the lack of convolution-kernel-weight coefficients. The filtering accuracy

tends to drop significantly in the multi-task scenarios. To solve this problem, we propose a content-aware dilated dynamic filtering algorithm based on the Dynamic Dilated Pyramid Module, adding the Coordinate Attention [34] weight coefficient to the parameters of the convolution kernel. This means that the dynamic model can obtain consciously selected channel and location information, while keeping the simplicity of the network architecture. The model is very suitable for filtering tasks such as RGB-M with uneven feature distribution.

### 2.2. Analysis of the MSPI features

Multispectral-polarized images have one more dimension than multispectral images. Different brightness of images with different polarization angles will be the main basis for distinguishing the fore-background of objects. In order to deeply understand the relationship between this polarization feature and salient targets, and to verify the reliability of polarization salient object detection, we establish polarized datasets and analyze their characteristics. The multispectral polarimetric images used in this paper are obtained from publicly available datasets [35], which contain polarimetric and multispectral images at four different polarimetric filter orientation angles (0, 45, 90, 135). The images from the dataset cover three bands in the range 400–800 nm. As shown in Fig. 1, in order to study the light intensity and polarization variation characteristics of multispectral polarized images in complex scenes, the spectral reflectivity and redundancy of foreground and background images were calculated respectively.

#### 2.2.1. Analysis of object reflectance and degree of polarization

In this part, we analyzed the change of original image reflectance and DOLP (Degree of linear polarization) intensity with different wavelengths respectively.

Firstly, the Stokes vector and DOLP are calculated as follows:

$$S_0 = (I_{0^\circ} + I_{45^\circ} + I_{90^\circ} + I_{135^\circ})/2 \tag{2}$$

$$S_1 = I_{0^\circ} - I_{90^\circ}, S_2 = I_{45^\circ} - I_{135^\circ} \tag{3}$$

$$DOLP = \frac{\sqrt{S_1^2 + S_2^2}}{S_0} \tag{4}$$

where is the total average light intensity, $S_1$ is the degree of polarization in 0°and 90° direction. $S_2$ represents the degree of polarization in 45°and 135° direction. Then the reflectance curve of $S_0$ and DOLP are solved as follows:

As shown in Fig. 2, the DOLP has its unique spectral characteristics with higher contrast between foreground and background than the $S_0$. This shows that the DOLP of object in complex scenes is more significant and special. Therefore, when generating the fused polarization components, it is necessary to calculate the polarization components of each band separately, and then fuse the multispectral polarization components to prevent the loss of the important DOLP spectral segment information. After fusion, the polarization component preserves the front-background differences to the greatest extent, making saliency detection relatively easy.

#### 2.2.2. Pearson's correlation coefficient of multispectral polarized image

Higher correlations among spectra and polar exist in highly redundant images. By calculating the Pearson's correlation coefficient (PCC) [36], we can compare the differences between the foreground and background information of each band image. For images $X_{mn}$, $Y_{mn}$

$$r = \frac{\sum_m \sum_n (X_{mn} - \overline{X})(Y_{mn} - \overline{Y})}{\sqrt{\left(\sum_m \sum_n (X_{mn} - \overline{X})^2\right)\left(\sum_m \sum_n (Y_{mn} - \overline{Y})^2\right)}} \tag{5}$$

As shown in Fig. 1, the foreground and background parts of the image are obtained by multiplying Ground Truth and the cluster of multi-spectral polarized images. The Pierce correlation coefficients of the foreground and background in several scenes are compared

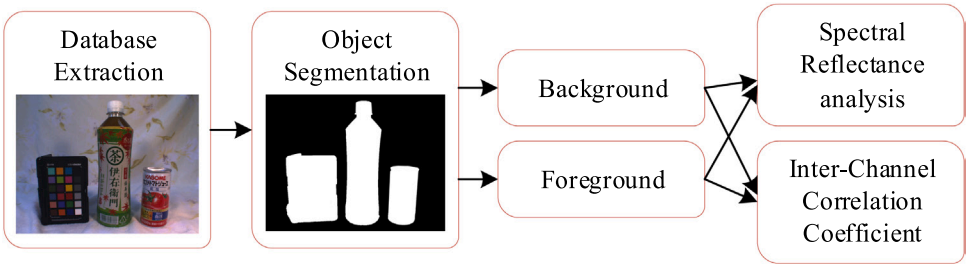


**Fig. 1.** Flowchart of Dataset Analysis. The multispectral polarized images in the datasets are divided into foreground and background. The spectral reflectivity and the internal correlation coefficient are calculated respectively, and the difference between them is analyzed. The results of the calculations can be found in 2.2.1 and 2.2.2.

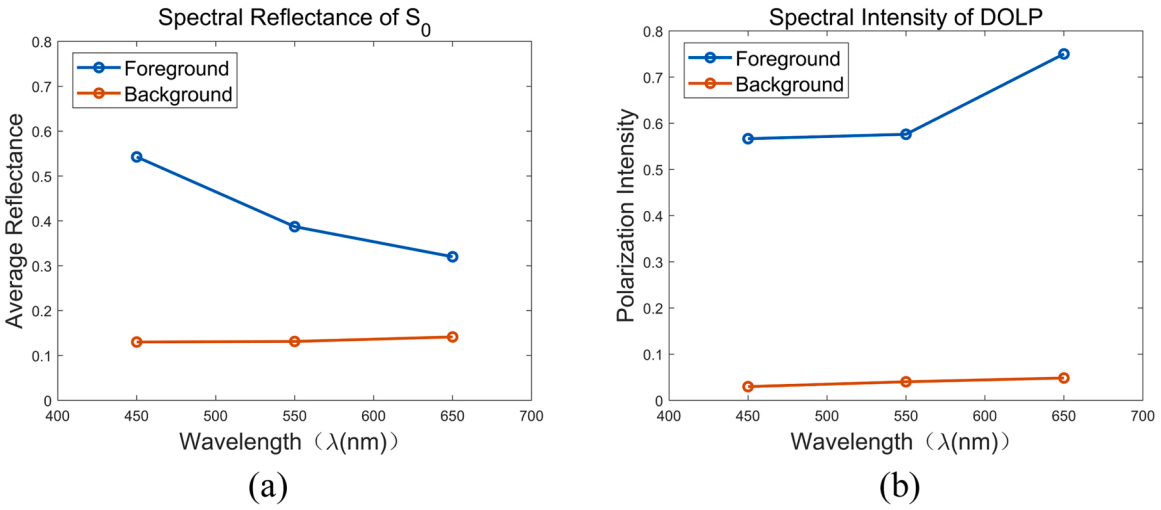


**Fig. 2.** Spectral curve distribution of foreground and background. (a) The spectral reflectance of RGB image; (b) The DOLP intensity curve. The vertical axis in Fig. 2(a) represents the average pixel intensity of the object area (blue line) and the background area (red line) in the calculated images, and the size ranges from 0 to 1; The more blue and red lines separate, the greater the difference between the foreground and background.

respectively, and then the redundancy of the image in spectral dimension and polarization dimension is analyzed.

As shown in Fig. 3, the foreground regions of images have obvious intensity differences in both spectral dimension and polarization dimension. The biggest difference lies between 0° and 90° in polarization dimension, while the background of passive imaging is relatively average. In the spectral dimension, the R channel and B channel differ significantly.

The information of RGB image and polarized image are complementary, while RGB image can provide texture details and color difference, and polarized image can provide intensity difference. In conclusion, the applied dataset is very suitable for the experiment of SOD.

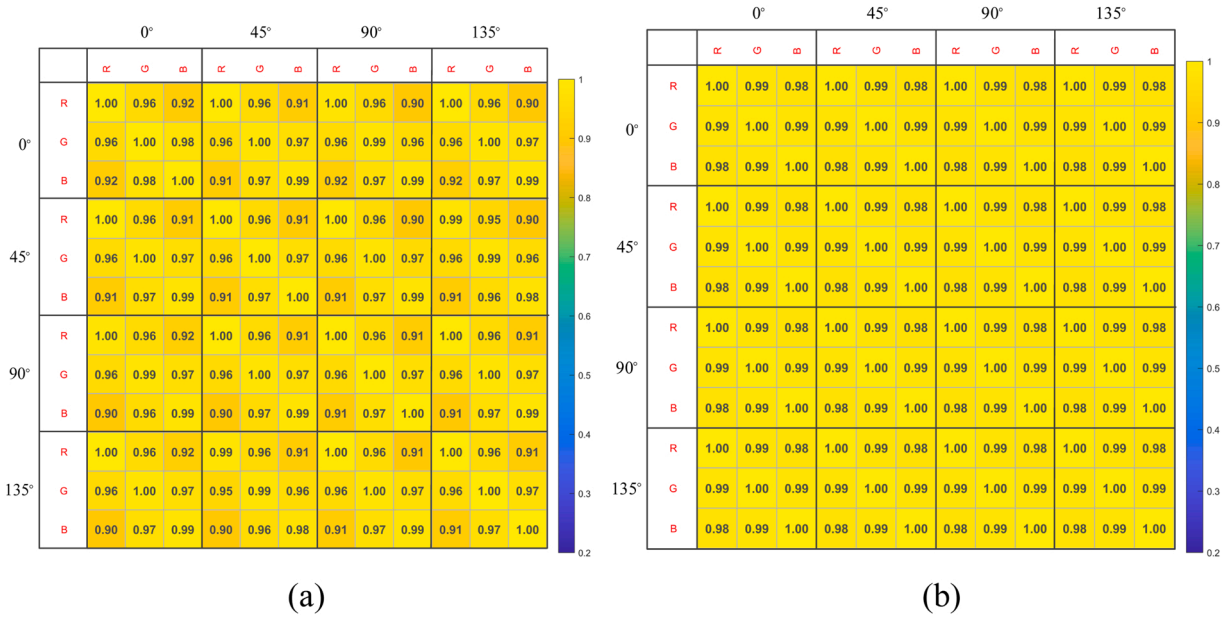


**Fig. 3.** Correlation coefficient of passive RGB polarization imaging. (a) correlation coefficient of foreground; (b) correlation coefficient of background. In the diagram (a), the maximum difference of the target in the color space can reach 0.1 (between R and B channel in super-pixel (0,90)), and the maximum difference in the polarization space is 0.02 (Super-Pixel (45,135) and Super-Pixel (135,135) in B channel; In diagram (b), the difference in color space is 0.02 at most (between R and B channel in the super pixel (0,90)), while the difference in polarization space is 0.

## 3. RGB-M saliency detection algorithm

Through the analysis in Section 2.2, the RGB image and polarized image are complementary in nature, and the combination of the two kinds of information will make it easier to identify salient objects. Therefore, a proposed dual RGB-M salient object detection framework based on distinguishing features of MSPI clues is discussed and presented in Fig. 4. Firstly, a new mask generation method is proposed to enhance the contrast of the object area, and then the combined salient area is obtained by applying our proposed network application mask into its RGB image. Process details are in Algorithm 1.

**Algorithm 1.** : Multispectral Fusion and saliency detection.

**Requires:**
Multiband Polarimetric Image Dataset
**Ensures**
Calculate Spectral Reflectance of shadow and Background
Calculate DOLP Spectral Reflectance and PCC
**if** Information Differs Significantly among Bands, **then**
    Compute Stokes Vector: $S_0$, $S_1$ and $S_2$
    Calculate the degree of polarization, the degree of angular polarization
**else**
    Compute Stokes Vector by mind-spectral image
**end**
Calculate three different masks $m_1$, $m_2$, $m_3$
Cascade $\mathbf{M} = \text{cat}(m_1, m_2, m_3)$
**for** each image **RGB** & **M**, **do**
    Generate Salient maps by proposed CDFNet
**end**

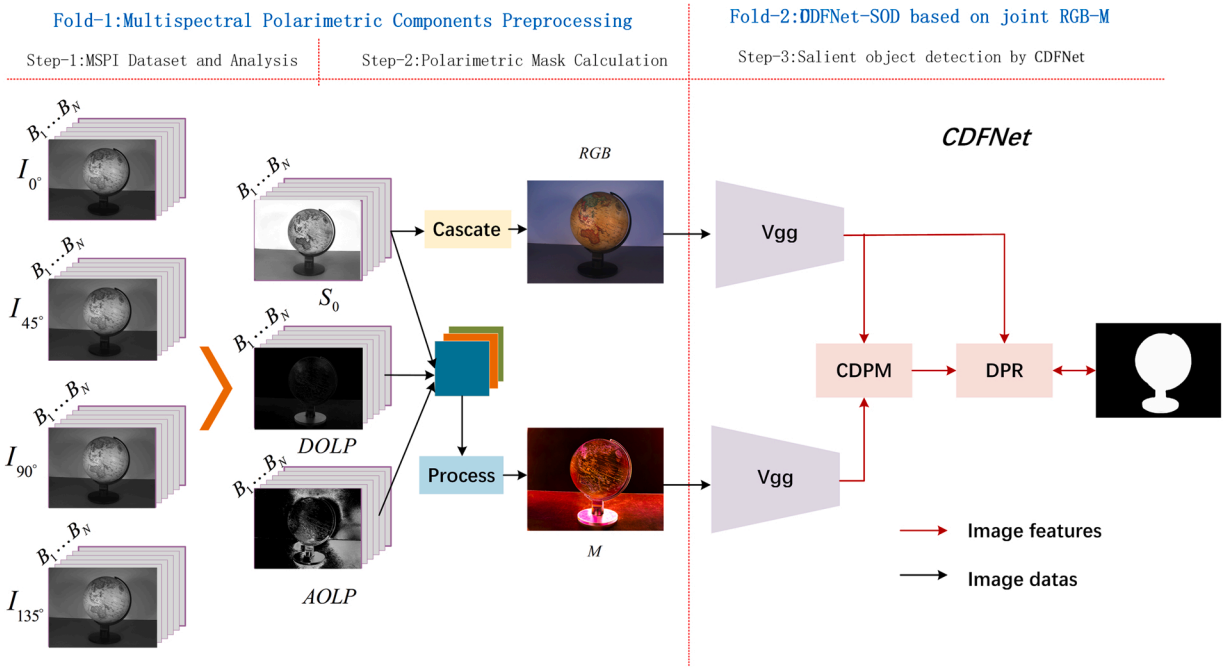


**Fig. 4.** The framework of the proposed two-fold RGB-M SOD algorithm. The first Fold of the algorithm is to solve the Stokes components by solving the multispectral images with four polarization directions, and then calculating the polarization mask based on the Stokes components; See Section 3.1 for details. In the second stage, **CDFNet** is used to detect RGB-M salient objects; See Section 3.2 for details;.

### 3.1. Proposed polarization mask

Because the intensity difference of the four-angles images is very small, it is difficult for the network to directly provide effective information in polarized image clusters. To enhance the polarization characteristics better, it is necessary to generate a single polarization mask based on the polarization image. Pre-processing is to re-integrate the polarization dimension, spectral dimension, and spatial dimension into three-channel mask images.

As shown in Fig. 5, the polarization of the salient object regions compared with the background has three main priors, polarization prior, entropy prior and intensity and spectral prior, to generate three corresponding mask M.

#### 3.1.1. The strong polarization channel of mask

The role of the high-polarized area mask is to sort out the parts of the image that are highlighted by *DOLP* (formula (4)) intensity. According to the principle of image dichroism, specular reflection often appears in a small area, while the distribution of diffuse reflection is usually more uniform. Diffuse reflection and specular reflection are easy to separate due to their different distribution characteristics. Areas with strong polarization can be sharply normalized. The polarization modulation coefficient $\ln(DOLP+e-1)$ can stretch the contrast of*DOLP*, making it more continuous in the spatial region. Finally, the mask for the strong polarized area can be described as:

$$m_1 = DOLP \cdot \ln(DOLP + e - 1) \tag{6}$$

Where $e$ is Euler number, which is used as a non-zero bias. The strong polarization area mask $m_1$ removes the interference of diffuse reflection to a great extent, which can enhance the contrast of materials with different polarization characteristics such as glass and plastic.

#### 3.1.2. Entropy fusion polarized region mask

The angular polarization degree of an object is linearly related to the normal angle of its surface microfacet, $AOLP = -\eta_r$, which indicates that the richer the texture information of the object is, the higher the local entropy of *AOLP* will be [37]. Besides, the entropy value of the *DOLP*of the artificial target in a large uniform area is small, while the entropy value will be relatively large at the edge of the object and in complex material areas. area. Based on this feature, this paper proposes an entropy fusion mask that combines *AOLP* and. information to enhance the contrast between artificial targets and backgrounds, which is shown as follows.

$$m_2 = DOLP \cdot [entropy(AOLP) + entropy(DOLP)] \tag{7}$$

where $AOLP = \frac{1}{2}\arctan(S_1/S_2)$, $S_1$ and $S_2$ are defined in formula (3). The first item of the $m_2$filters out the complex region and retains its polarization degree information. The second item can preserve the edge region and eliminate the influence of the background significantly.

#### 3.1.3. Weak light and spectral feature mask

The two masks above are generated based on the calculation of multi-spectral images fused by IFT. The fusion image mixes the useful spatial information of each band to the greatest extent, so that the accuracy of the first two masks can be improved greatly. However, the difference information in the visible spectral dimension of MSPI is eliminated. In order to preserve the spectral information, the RX inspection algorithm is used to effectively extract the spectral anomalous regions and convert them into intensity information, which is essential to the concatenation and merging of the three masks. Spectral feature RX filter, defined as the RX anomaly detection algorithm, is usually applied for local target detection [38].

$$m_{rx} = RX[DOLP(\lambda)], \lambda \in (400nm, 700nm) \tag{8}$$

In addition, to find out the object regions with prominent spectrum but weak light intensity in complex scenes, it is necessary to consider the influence of light intensity. The light intensity is affected by the observation azimuth $\beta$, which is determined by the electric field components of the two vibration directions of the light.

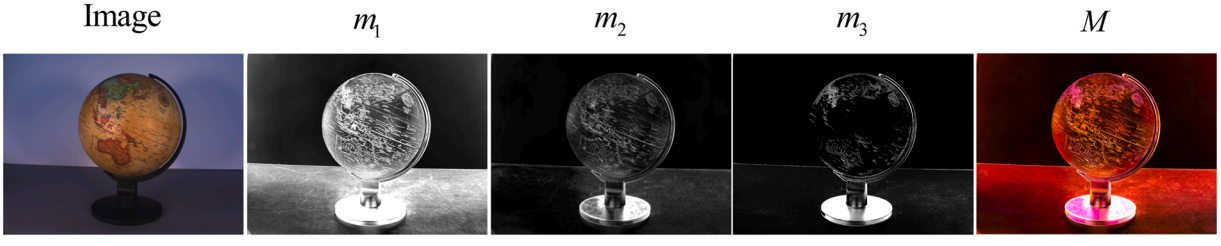


**Fig. 5.** The comparison between visual image and *M*.The $m_1$, $m_2$, and, $m_3$ are used as red, green and blue channels to generate a pseudo-color image *M*; The spherical part of the globe in the picture is visually prominent, while the bracket part is polarized prominently.

$$E_x = \sqrt{(S_0 \times DOLP) + S_1} \tag{9}$$

$$E_y = \sqrt{(S_0 \times DOLP) - S_1} \tag{10}$$

$$\beta = \tan^{-1}\left(E_y / E_x\right) \tag{11}$$

The intensity of the light in the area can be calculated by dividing the total light intensity $S_0$ by the logarithmic function of the azimuth angle $\beta$ [39]. Use a logarithmic function to stretch the contrast to get the mask with the light intensity:

$$m_{st} = \frac{S_0}{\log(1+\beta)} \tag{12}$$

Finally, the intensity of light is adjusted by the RX anomaly, so that the regions with the darkest light intensity but special spectrum are identified as:

$$m_3 = DOLP \cdot (1 - m_{st}) \cdot \exp[-k \cdot (1 - m_{rx})] \tag{13}$$

The cascade of three masks representing different attributes reduces the risk of false detection and facilitates dual-source fusion detection with RGB images. In theory, the three mask functions are the same as the RGB channels, since the intensity of each area represents the strength of the spatial polarization characteristics, and different objects have significant and unique polarization characteristics, all of which can help to identify significant targets better. Thus, the final combined 3-channal mask M is defined as:

$$M = Concatenate[m_1, m_2, m_3] \tag{14}$$

### 3.2. Proposed RGB-M SOD method

In this section, we first introduce the overall structure of our proposed method, and then detail the content-aware dynamic pyramid model (CDPM), and finally propose the dilated pyramid refinement (DPR) structure.

#### 3.2.1. Two stream structure

As shown in Fig. 6, we have built a dual branch network. It has two inputs, one is the RGB image and the other is the polarization mask M. Through a two-pass encoder network composed of 5-layer convolutional blocks of VGG-19, we obtain the features of different scales of two kinds of images respectively. The semantic information contained in deeper features is more important. In order to eliminate the errors of RGB depth features, by using PAC structure [40], we deliberately utilize $f_m^5$ to guide $f_{rgb}^5$ to generate DPR input feature $f_{Drgb}^5$. To balance the calculation efficiency and accuracy, we choose deeper encoder features $f^3, f^4, f^5$ to calculate the fusion features of two streams for the subsequent dynamic filter layers. We use the DenseNet block [41] to construct a feature translator layer [33], and perform interactive fusion and dimension reduction of depth features between M flow features $f_m^3, f_m^4, f_m^5$ and the RGB image

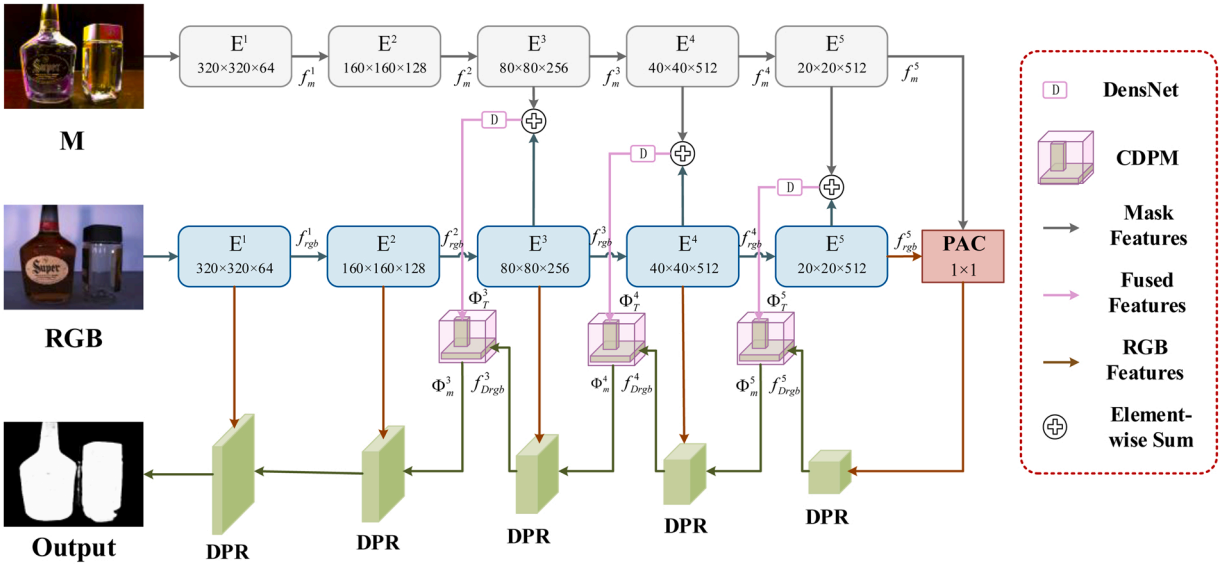


**Fig. 6.** In the whole process of CDFNet, we selected three deeper layers of RGB information and M information, and deeply fused them with the Sum+DenseNet structure, which served as the kernel of the CDPM convolution window. We use M of the deepest layer to guide the RGB information to generate the input characteristics of the DPR through the PAC structure.

feature $f^3_{rgb}, f^4_{rgb}, f^5_{rgb}$ streams to generate fused features $\Phi^3_T, \Phi^4_T, \Phi^5_T$. These fused features will enter the CDPM model to monitor the $f_{Drgb}$ stream and guide the generation of salient regions. After the multi-layer decoder recovers the resolution of the image, we get the final prediction of salient objects under the supervision of Ground Truth (GT).

### 3.2.2. Content-aware dynamic pyramid model

We propose a content-aware dynamic pyramid model, which can filter the RGB reconstructed features to form an adaptive convolution kernel and guide the RGB information to reconstruct the final saliency map. The CDPM has two inputs, that is, the reconstructed features $f_{Drgb}$ obtained from DPR and the deep fusion features $\Phi_T$ output from DenseNet. As shown in Fig. 7, CDPM uses three sub-models, namely, kernel generation units (KGUs), kernel transformation units (KTUS) [33] and coordinate attention units (CAU) [34], where KGUs is used to generate an independent conv weight Tensor of different size dilation (d = 1,3,5),like $3 \times 3(KGU_1)$, $7 \times 7(KGU_2)$,$11 \times 11(KGU_3)$. KGUs is also a DenseNet structure. KTUs is used to transform the weight windows generated by KGUs into different dilation scales. CAU is a self-attention mechanism capable of generating additional window weights. The attention information in X and Y spatial directions that can be extracted by CAU. It can capture long-range dependence along one spatial direction, while keeping the accurate position information along the other spatial direction. The obtained feature maps are then separately encoded into a pair of orientation-aware and position-sensitive attention maps, which can be applied complementarily into the input feature maps to enhance the representation of objects of interest.

The mathematical relationship between the pixel value $\mathbf{v} = (\mathbf{v}_1, \ldots, \mathbf{v}_n), \mathbf{v}_r \in \mathbb{R}^C$ of each image $f_{Drgb}$ and the convolution kernel weight of scale $\mathbf{A} \in \mathbb{R}^{c' \times c \times s \times s}$ , $\mathbf{W} \in \mathbb{R}^{c' \times c \times s \times s}$is as follows.

$$v'_r = \sum_{q \in \Omega(r)} A[p_r - p_q] \cdot W[p_r - p_q] \cdot v_q \tag{15}$$

Where $\boldsymbol{p}_r = (x_i, y_j)^T$ represents the coordinates of pixels, and$\Omega()$represents convolution kernel window with size of $s \times s$.$[\boldsymbol{p}_r - \boldsymbol{p}_q]$is the position in the convolution kernel window, while the features of the output $\mathbf{v}'_r \in \mathbb{R}^{c'}$ correspond to the $r$ point. **A**is obtained by calculating the weight of CAU, and **W**is obtained by fusing $\Phi^i_T$features through KGUs and KTUs transform.

To speed up the operation, the specific method adopted by the algorithm is to decompose the depth feature $\Phi^i_T$ into $3 \times 3$-channel dynamic filter coefficients of the convolution kernel through KTUs. These $3 \times 3$ uncorrelated coefficient tensors are multiplied by the weights of $3 \times 3$ position attention values to get finally convolution wights. Three adaptive convolution layers with different convolution dilation sizes are applicated on $f^i_{Drgb}$to get a three channel dynamic features. Finally, it combines the input feature level

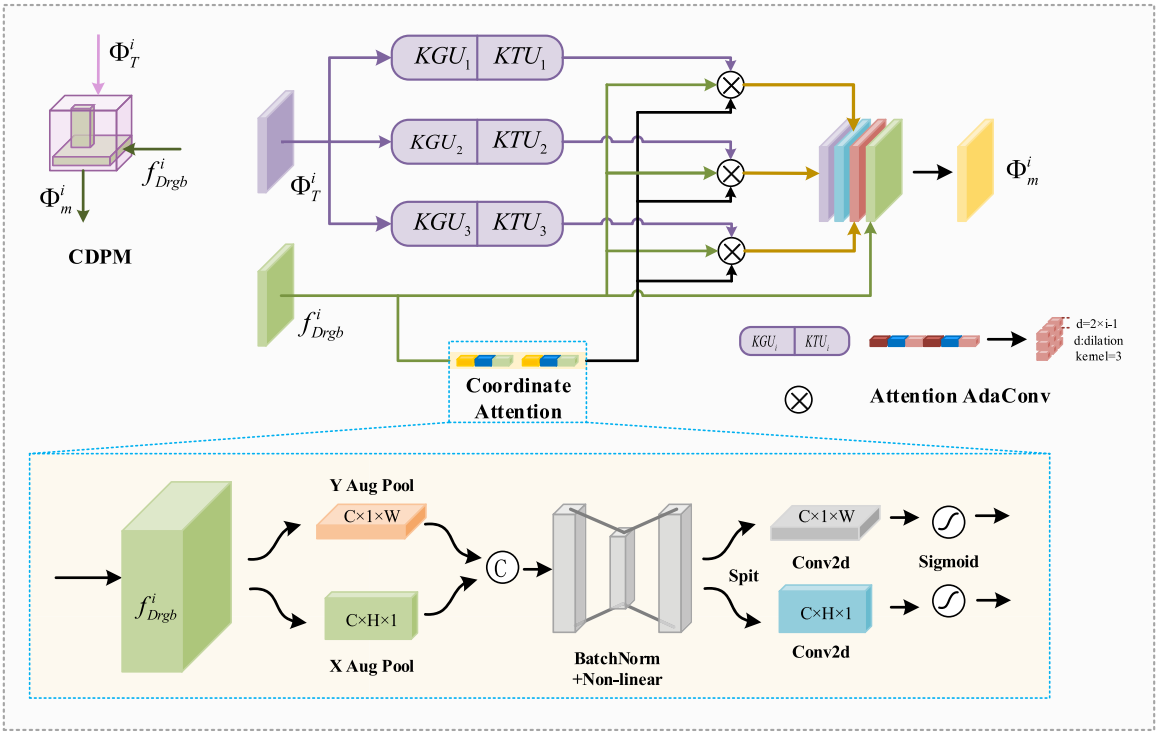


**Fig. 7.** CDPM structure diagram. AdaConv convolution kernel k = 3. There are three configurations of dilation= 1, 3, 5, the convolution kernel weight of each configuration is the product of Coordinate Attention and KTUs value. After performing the AdaConv convolution operation on $f_{Drgb}$, the final fusion feature $\Phi^i_m$ is generated after cascade fusion.

$f^i_{Drgb}$ to generate the fusion feature $\Phi^i_m$.

$$\begin{aligned}\Phi^i_m &= CDPM^i\left(f^i_{Drgb}, \Phi^i_T\right)\\ &= \mathscr{F}\left(\mathcal{R}\left(f^i_{Drgb}\right), f^i_{C^1}, f^i_{C^2}, f^i_{C^3}\right)\\ &= \mathscr{F}\Big(\mathcal{R}\left(f^i_{Drgb}\right), KTU^i_1\left(KGU^i_1\left(\Phi^i_T\right)\right) \otimes \mathscr{C}^i_1\left(f^i_{Drgb}\right) \otimes \mathcal{R}\left(f^i_{Drgb}\right),\\ &\quad KTU^i_2\left(KGU^i_2\left(\Phi^i_T\right)\right) \otimes \mathscr{C}^i_2\left(f^i_{Drgb}\right) \otimes \mathcal{R}\left(f^i_{Drgb}\right),\\ &\quad KTU^i_3\left(KGU^i_3\left(\Phi^i_T\right)\right) \otimes \mathscr{C}^i_3\left(f^i_{Drgb}\right) \otimes \mathcal{R}\left(f^i_{Drgb}\right)\Big)\end{aligned} \tag{16}$$

Where $\mathcal{C}^i_j()$ is the operation of calculating the coordinate attention value of different features $f^i_{Drgb}$, $\mathcal{R}()$represents the $1 \times 1$ convolution operation, and $\mathscr{F}()$ is the cascade fusion operation. See Algorithm 2 for the AdaConv $\otimes$ operation, where s = 3 is selected.

**Algorithm 2.** : The operation process of Attention AdaConv $\otimes$ of KTU in CDPM.

**Requires:** $f^i_r = \mathrm{R}\left(f^i_{Drgb}\right) \in \mathbb{R}^{N\times C\times H\times W}$, $f^i_{a^j} = \mathrm{C}^i_j\left(f^i_{Drgb}\right) \in \mathbb{R}^{N\times(9\times C)\times 1\times 1}$,

$f^i_{\varphi^j} = \mathrm{KTU}^i_j\left(\mathrm{KGU}^i_j\left(\Phi^i_T\right)\right) \in \mathbb{R}^{N\times(9\times C)\times H\times W}$

**Output:** $f^i_{C^j} \in \mathbb{R}^{N\times C\times H\times W}$

$d = 2j-1$ ;

pad $f^i_r$ with 0 from $(H,W)$ to $(H+2\times d, W+2\times d)$ ;

**for** $n$ **in** $0 \to N-1$

  **for** $c$ **in** $0 \to C-1$

    **for** $h$ **in** $0 \to H+d-1$

      **for** $w$ **in** $0 \to W+d-1$ **do**

$$\begin{aligned}f^i_{C^j}[n,c,h,w] = \sum_{l=-1}^{1}\sum_{m=-1}^{1}\Big\{ & f^i_{\varphi^j}[n,(l+1)\times 3+(m+1),h,w]\\ & \times f^i_{a^j}[n,(l+1)\times 3+(m+1),1,1]\\ & \times f^i_r[n,c,h+l\times d, w+m\times d]\Big\}\end{aligned}$$

### 3.2.3. Dilated pyramid refinement

To improve the utilization ratio of multi-scale fusion features, we propose a dynamic multi-scale feature reconstruction network, the structure of which is shown in Fig. 8. We configure a three-stream network to further fuse features $f^i_{Drgb}$at various dilation scales and dynamic filtering features to enhance the performance of unsampling on the network.

### 3.2.4. Loss function

Our loss function for optimizing $L$ is combined by binary cross entropy (BCE) and Dice loss [42], its effectiveness has been well

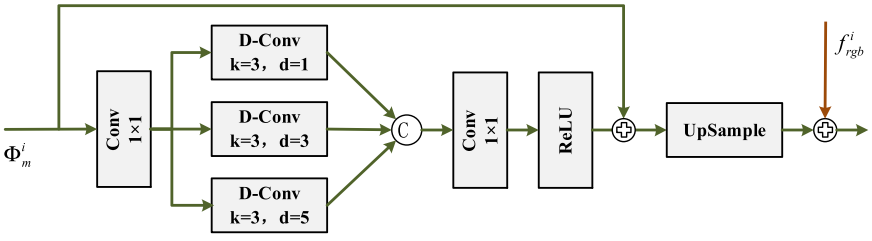


**Fig. 8.** The Dilated pyramid refinement (DPR) structure can further fuse the information of three dilation scales and enhance the use of features. The function of the ReLU layer is to accelerate the convergence of the network.

proven [43], and its expression is

$$L_{bce}(P,G) = G\log P + (1-G)\log(1-P) \tag{17}$$

$$L_{dice}\left(P,G\right) = 1 - \frac{2\cdot G\cdot P}{\|G\| + \|P\|} \tag{18}$$

$$L(P,G) = L_{bce} + L_{dice} \tag{19}$$

Where $P$ and $G$ represent prediction and real significance maps respectively. || || stands for the first-order norm. Losing dice is an effective way to alleviate the category imbalance of foreground and background.

## 4. Experiments

### 4.1. Implementation details

Our proposed models are based on the PyTorch library. Adam [44] optimizer was chosen to optimize the model. The selected parameter is $\beta_1 = 0.9$, $\beta_2 = 0.99$,weight $= 2$, decay $=10^{-4}$and the batch size $= 16$. The learning rate of our design is $lr = 5 \times 10^{-5}$, $power = 0.9$. We use 8 Tesla K 80 (24 g) as a training platform. The training lasts for a total of 30 epochs.

### 4.2. Datasets

Due to the limited number of existing polarized images, we have collected 120 groups of polarized photos from [45], in addition to some of our photos taken with a polarimetric camera. There are images with four polarization angles, and the sizes are 2048 * 2048. Using the method in Section 3.1, generate the polarization mask M of the images, and form the RGB-M dataset together with the RGB images and the GT images. As the number of images is too small, we first use the NJUD [46] RGB-D data set to pre-train our network, so that it can extract bidirectional information initially. Then, we use the polarization dataset for further fine-tuning training to make it completely suitable for salient object detection in biased RGB-M images.

As shown in Fig. 9, we have also constructed a classified testing dataset for polarized image SOD, which is divided into three categories: ***S*** (submerged), ***T*** (transparent) and ***O*** (ordinary). The objects in the submerged group are strong camouflage. The objects in the transportation group are transparent and difficult to distinguish. The last one is a common object, which is not prominent in the polarization dimension. Our dataset can be accepted the address https://github.com/Sci-Epiphany/RMD_Datasets.

### 4.3. Selection of SOD metric

The evaluation criteria for salient region detection will be compared using seven indicators: mean **E-measure** $E_m$ [47], **F-measure** ($F_{avg}$,$F_{\max}$)[48], **S-measure**( $S_m$) [49], Precision-Recall(**PR**) **curve** and **MAE** [49]. **PR curve** plots the precision rate and the recall rate.

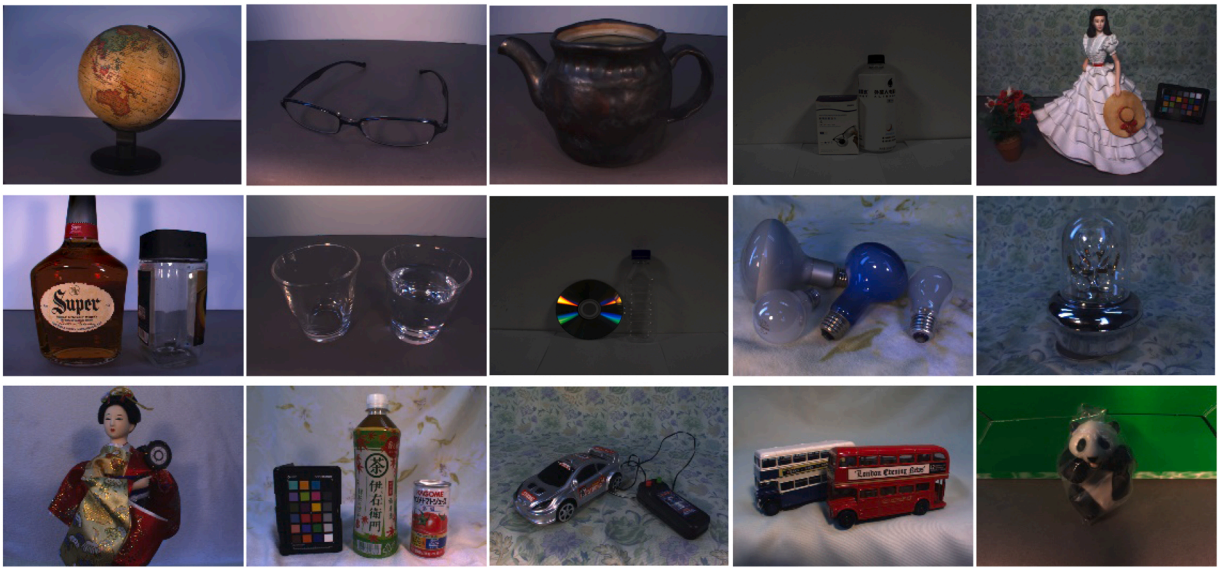

**Fig. 9.** Part of the RGB-M dataset. From top to bottom, each row corresponds to **S** (Submerged), **T** (Transparent), and **O** (Ordinary) polarization images of three different object features.

The horizontal axis represents the recall rate, and the vertical axis represents the precision rate. PR Curves represent the proportion of samples predicted to be positive samples. The more the curve is outside, the better the detection effect will be. **E-measure** uses the predictions and basic facts of removing the mean to calculate the similarity, which represents image-level statistics and local pixel matching. **F-measure** is a similarity measure based on regions, which is expressed as a weighted average of Precision and Recall. We employ the threshold changing from 0 to 1 to get $F_{\max}$, and use the mean value of the prediction as the threshold to obtain $F_{avg}$. Since F-measure reflects the performance of the binary prediction under different thresholds, we evaluate the consistency at the regional level according to F-measure threshold curves. **S-measure** calculates the similarity of regional perception structure between prediction and ground truth. **MAE** is used to calculate the method of the absolute value of the error between the observed value and the true value.

### 4.4. Salient object detection experiment

For a more comprehensive and effective comparison of the algorithm proposed in this paper, we compare it with eight state-of-the-art CNNs-based salient object detection algorithms, including EGNet [50], BASNet [51], MINet [52], GateNe [53], END [54], U2Net [55], F3Net [56], and VST [57]. In order to ensure the accuracy and fairness of the results, all indexes of the classified dataset are calculated by the public codes provided by the algorithm authors. In addition, the encoders of all algorithms are based on the VGG-19 model.

As shown in Table 1, we illustrate the results of all the comparison algorithms on five indexes. The overall performance of our proposed algorithm is the best. In the transparent dataset, our algorithm is the best in $F_{\max}$, $E_\phi$ and *MAE*, while in the submerged

**Table 1**
Results of different CNNs-based SOD methods across three datasets. The best results are highlighted in **red** and the second-best results are marked in **blue**. The three types of datasets are ***S*** (submerged), ***T*** (transparent), and ***O*** (ordinary) respectively, and the number of pictures in each type remains the same. ***Avg*** contains all pictures of our testing dataset.

| Type | Metric | EGNet [49] | BASNet [50] | MINet [51] | GateNet [52] | EDN [53] | U2Net [54] | F3Net [55] | VST [56] | CDFNet (Ours) |
|---|---|---|---|---|---|---|---|---|---|---|
| ***T*** | $S_m\uparrow$ | 0.69 | 0.862 | 0.874 | 0.89 | 0.886 | 0.904 | 0.895 | **0.921** | **0.916** |
| | $F_{avg}\uparrow$ | 0.682 | 0.882 | 0.887 | 0.897 | 0.897 | 0.9 | 0.899 | **0.916** | **0.912** |
| | $F_{\max}\uparrow$ | 0.709 | 0.893 | 0.892 | 0.911 | 0.907 | 0.915 | 0.904 | **0.931** | **0.935** |
| | $E_\phi\uparrow$ | 0.741 | 0.886 | 0.907 | 0.915 | 0.899 | 0.921 | 0.909 | **0.932** | **0.943** |
| | $MAE\downarrow$ | 0.153 | 0.078 | 0.061 | 0.062 | 0.071 | 0.056 | 0.058 | **0.048** | **0.046** |
| ***S*** | $S_m\uparrow$ | 0.642 | 0.782 | 0.749 | 0.768 | 0.765 | 0.773 | **0.834** | 0.806 | **0.869** |
| | $F_{avg}\uparrow$ | 0.639 | 0.824 | 0.762 | 0.787 | 0.811 | 0.793 | **0.878** | 0.833 | **0.906** |
| | $F_{\max}\uparrow$ | 0.729 | 0.837 | 0.777 | 0.814 | 0.832 | 0.811 | **0.894** | 0.847 | **0.919** |
| | $E_\phi\uparrow$ | 0.656 | 0.804 | 0.775 | 0.770 | 0.776 | 0.789 | **0.854** | 0.829 | **0.900** |
| | $MAE\downarrow$ | 0.229 | 0.138 | 0.155 | 0.157 | 0.140 | 0.134 | **0.108** | 0.109 | **0.073** |
| ***O*** | $S_m\uparrow$ | 0.815 | 0.958 | 0.945 | 0.928 | 0.947 | **0.969** | 0.947 | 0.956 | **0.962** |
| | $F_{avg}\uparrow$ | 0.821 | 0.966 | 0.961 | 0.929 | 0.966 | **0.970** | 0.96 | 0.946 | **0.967** |
| | $F_{\max}\uparrow$ | 0.832 | 0.978 | 0.968 | 0.941 | 0.974 | **0.983** | 0.968 | 0.962 | **0.979** |
| | $E_\phi\uparrow$ | 0.858 | 0.975 | 0.963 | 0.936 | 0.965 | **0.979** | 0.963 | 0.972 | **0.980** |
| | $MAE\downarrow$ | 0.088 | 0.017 | 0.024 | 0.032 | 0.027 | **0.014** | 0.023 | 0.019 | **0.016** |
| ***Avg*** | $S_m\uparrow$ | 0.722 | 0.872 | 0.859 | 0.89 | 0.863 | 0.884 | 0.894 | **0.895** | **0.917** |
| | $F_{avg}\uparrow$ | 0.721 | 0.894 | 0.887 | 0.872 | 0.895 | 0.890 | **0.915** | 0.899 | **0.931** |
| | $F_{\max}\uparrow$ | 0.746 | 0.905 | 0.879 | 0.911 | 0.907 | 0.902 | **0.924** | 0.912 | **0.945** |
| | $E_\phi\uparrow$ | 0.757 | 0.892 | 0.884 | 0.874 | 0.883 | 0.899 | 0.911 | **0.912** | **0.942** |
| | $MAE\downarrow$ | 0.154 | 0.075 | 0.079 | 0.083 | 0.078 | 0.067 | 0.062 | **0.058** | **0.044** |

dataset, all the indexes are obviously ahead of other algorithms. On the ordinary dataset where the polarization feature is not obvious, our performance is only slightly lower than that of U2Net. Finally, on the all testing dataset, our algorithm is 2.41%, 1.72%, 2.27%, 3.28% and 24.14% ahead of the second-ranked algorithm in $S_m$, $F_{avg}$, $F_{\max}$, $E_\phi$ and *MAE* indicators respectively. In Fig. 10 and Fig. 11, we construct PR curves and F-measure curves, respectively. Our measurement is smoother under most of thresholds, indicating that our detection results are completer and more continuous.

In Fig. 12, we list some representative results. These graphics contain objects in various complex backgrounds, including submerged (the first and second lines), transparent (the third and fourth lines), small objects (the fifth and sixth lines) and the complex object (the fifth and seventh lines). As can be seen from the results in the figure, our algorithm has the best performance, which proves the advantage of our algorithm for complex target recognition, such as dim and transparent.

### 4.5. Masks qualities comparison experiment

In order to prove the advanced nature of our proposed mask and its feasibility for salient object detection, we compared several best mask generation methods Zhao [18], Zhou [37],and Islam [41] to evaluate ours mask in terms of image visual quality and salient recognition accuracy. We selected four non-reference image quality evaluation indexes, namely, DBCNN [58], NIMA [59], Brisk [60], and Irnik [61]. DBCNN mainly uses the CNN network to judge the degree of distortion of the image. It is of great significance to evaluate the reliability of the membrane treatment system. NIMA focuses on evaluating the technical and aesthetic qualities of images, while BRISQUE is an effective method to evaluate the naturalness and distortion of images. Finally, ILNIQE is also one of the common image evaluation methods. Its biggest feature is the use of the opinion-unaware method, which has the characteristics of being "completely blind". The four methods adopt feature-extraction methods based on the CNN network and image-evaluation methods based on traditional NSS features, which are of great significance for image evaluation in multiple application scenes.

From the results in Table 2, it can be found that, except that the NIMA rate is lower than that of the Islamic algorithm, the polarization mask comparison algorithm proposed by us is leading in various indicators. This is because exponential or logarithmic stretching is widely used in our proposed mask, which makes the details of the image clearer. Secondly, the introduction of spectral information and spatial features also greatly enhances the of image information density.

As shown in Fig. 13, the mask image we generated is more complete, while the target is more prominent and the overall brightness of the image is higher. This is very useful for the recognition of salient objects.

In Table 3. We compared the effects of different masks on salient object recognition. We used the baseline composed of common encoders and decoders as the recognition network, and the mask features as the input features of the network to conduct the salient target recognition experiment, and compared the results. It shows that our method also has the best performance in SOD.

### 4.6. Ablation study

In order to further verify the contribution and importance of each part of the network, we conducted ablation experiments. Our basic network is still based on the fundamental structure of encoder-decoder. In the basic network, the fusion features of three networks enter the decoding layer through $1 \times 1$ convolution for final prediction. Then, we added $T_{rgb}$, $T_m$, CDPM structure, and DPR structure respectively, and observed their effects on the results.

It can be seen from Table 4 that the RGB fusion characteristics of M (Model 3 and Model 1) $S_m$, $F_{avg}$,$F_{\max}$, $E_\phi$,and MAE are improved by 2.21%, 2.01%, 2.06%, 2.87% and 35.82% respectively compared with the single visible light characteristics. The comparison between model 4 and model 3 shows that adding CDPM can also significantly improve the detection performance, and the indexes are improved by 0.62%, 1.31%, 0.53%, 0.75% and 19.05% respectively. In addition, DPR has made great contributions to the improvement of network performance. By adding a few features, the reusability of the fused features can be greatly enhanced.

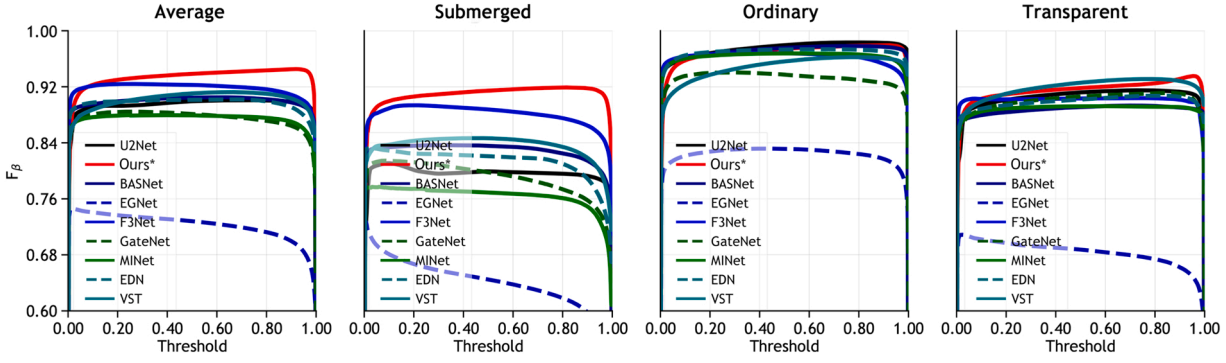


**Fig. 10.** F-measure(vertical axis) – threshold(horizontal axis) curves on three RGB-M SOD datasets. The red line represents our results. Average represents means result of all pictures of testing datasets.

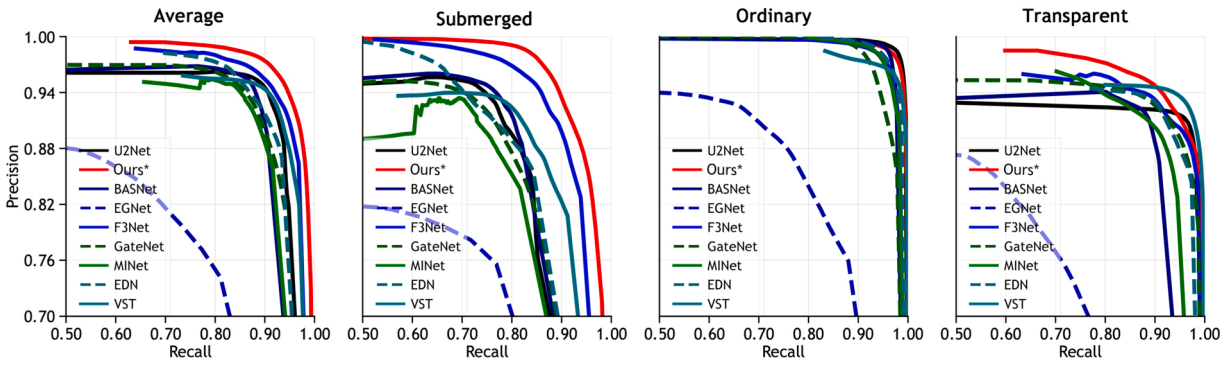


**Fig. 11.** Precision (vertical axis) recall (horizontal axis) curves on three RGB-M SOD datasets.

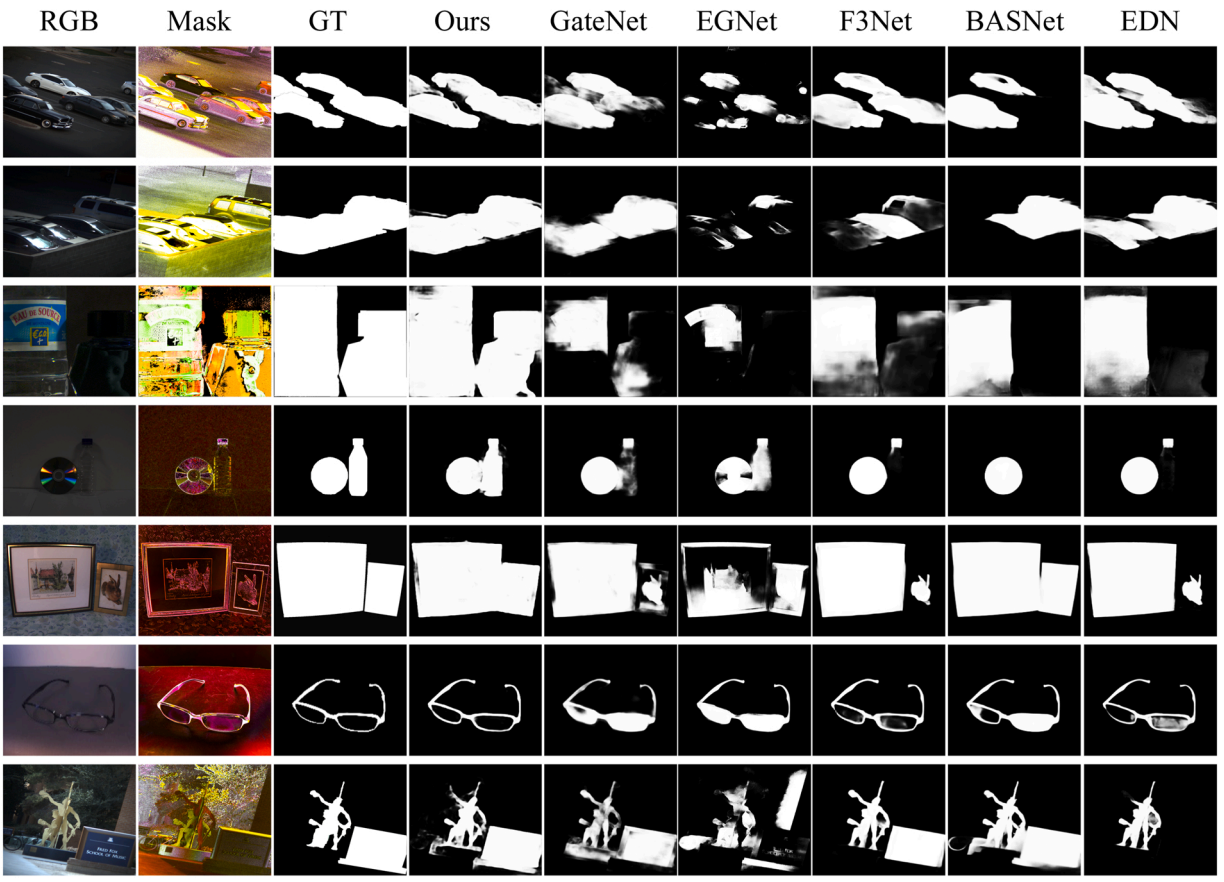


**Fig. 12.** Comparison of visual results between some state-of-the-art methods and ours.

## 5. Conclusions

In this paper, an RGB-M method for salient object detection based on multispectral polarization images is proposed. It combines complicated information from multispectral and multi-polarimetric cues to improve the foreground's contrast and details. The proposed three-channel polarization mask M based on the information entropy theory has achieved ideal results in image quality and remarkable recognition effect. In addition, our dual-stream input CDFNet can extract the useful information of the polarization mask M and the RGB image and generate accurate saliency maps. The CDPM structure can well realize the guiding filtering effect of m for the RGB stream, and the DPR structure also plays the function of information multiplexing in the reconstruction process. The experimental results illustrate the RGB-M has higher indicators accuracy such as E-measure, S-measure, and F-measure than the state-of-the-art algorithms. Salient object detection based on multispectral polarization images proves to be very effective for objects in complex

**Table 2**
Comparison between the proposed mask and three existing masks on four visual evaluation metrics in three datasets ***S*** (submerged), ***T*** (transparent), and ***O*** (ordinary). The best results are highlighted in **red** and the second-best results are marked in a **black body**.

| Type | Metric | Zhao | Islam | Zhou | Ours |
|---|---|---|---|---|---|
| ***S*** | DBCNN↑ | 37.078 | **53.055** | 45.202 | **53.731** |
| | NIMA↑ | 3.621 | **5.580** | 4.182 | **5.063** |
| | BRISQUE↓ | 65.995 | 182.779 | 45.412 | **33.890** |
| | ILNIQE↓ | 64.293 | 99.160 | 62.832 | **44.644** |
| ***T*** | DBCNN↑ | 30.878 | **52.690** | 45.350 | **54.767** |
| | NIMA↑ | 3.204 | **5.476** | 4.138 | **5.305** |
| | BRISQUE↓ | 55.643 | 186.246 | **43.767** | **40.938** |
| | ILNIQE↓ | **55.394** | 84.989 | 65.480 | **37.843** |
| ***O*** | DBCNN↑ | 40.378 | **52.417** | 49.579 | **56.541** |
| | NIMA↑ | 3.776 | **5.423** | 4.212 | **5.493** |
| | BRISQUE↓ | 62.889 | 182.105 | **45.237** | **25.729** |
| | ILNIQE↓ | 63.132 | 102.952 | **61.789** | **39.599** |

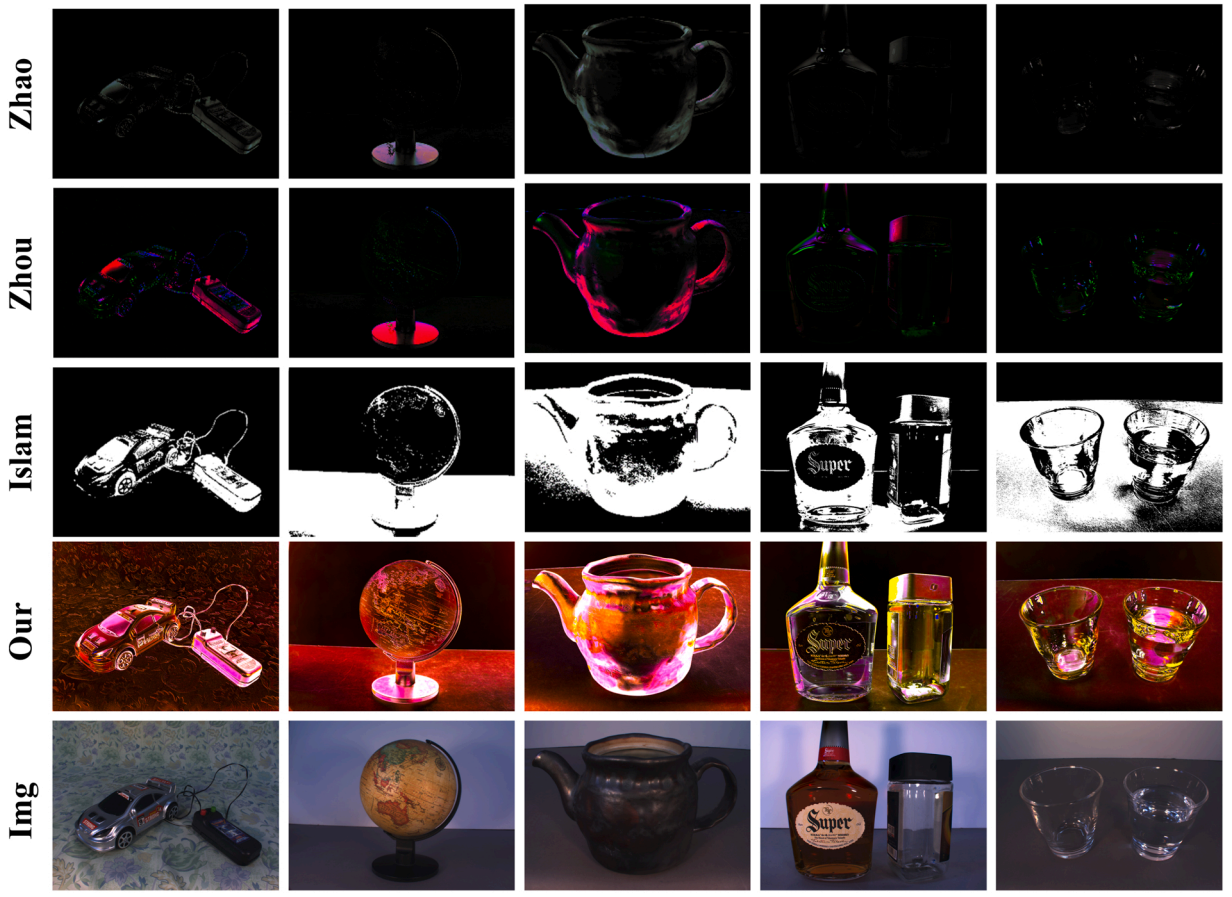


**Fig. 13.** Visual comparison of polarization masks generated in recent methods and ours.

**Table 3**
SOD comparison of the mask proposed in this paper with other masks.

| Methods | $S_m$ | $F_{avg}$ | $F_{max}$ | $E_\phi$ | *MAE* |
|---|---|---|---|---|---|
| Zhao[18] | 0.568 | 0.484 | 0.613 | 0.563 | 0.233 |
| Zhou[37] | 0.717 | 0.683 | 0.699 | 0.738 | 0.163 |
| Islam[41] | 0.718 | 0.71 | 0.73 | 0.759 | 0.172 |
| CDFNet(Ours) | 0.815 | 0.801 | 0.822 | 0.845 | 0.103 |

**Table 4**
Ablation Study results of each submodule in testing datasets.

| Model | Baseline | $T_{rgb}$ | $T_m$ | CDPM | DPR | $S_m$ | $F_{avg}$ | $F_{max}$ | $E_\phi$ | *MAE* |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | √ | √ | | | | 0.887 | 0.896 | 0.918 | 0.906 | 0.078 |
| **2** | √ | | √ | | | 0.815 | 0.801 | 0.822 | 0.845 | 0.103 |
| **3** | √ | √ | √ | | | 0.909 | 0.916 | 0.937 | 0.932 | 0.050 |
| **4** | √ | √ | √ | √ | | 0.915 | 0.928 | 0.942 | 0.939 | 0.042 |
| **5** | √ | √ | √ | √ | √ | **0.917** | **0.931** | **0.945** | **0.942** | **0.044** |

scenes. In the future, efforts will be made to expand the datasets and build a more robust MSPI salient object detection algorithm.

## Funding

This research was Funding by that Natural Science Foundation of Jilin province, grant number 20210101099JC, the National Natural Science Foundation of China (NSFC), grant number 62171430, the National Natural Science Foundation of China, grant number 62101071, and the Innovation and Entrepreneurship Team Project of Zhuhai City, grant number ZH0405190001PWC.

## CRediT authorship contribution statement

**Suining Gao**: Conceptualization, Methodology, Writing - original draft, Software. **Xiubin Yang:** Resources, Data curation. **Ziming Tu:** Visualization, Investigation. **Mo Wu**: Software, Validation. **Li Jiang:** Writing - review & editing. **Zongqiang Fu:** Formal analysis.

## Declaration of Competing Interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

## Data availability

Data will be made available on request.

## References

[1] (a) B. Lai, X. Gong, Ieee, Saliency Guided Dictionary Learning for Weakly-Supervised Image Parsing (Seattle, WA), *2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)* (2016) (Seattle, WA);
(b) Jun 27-30 2016, IEEE Conf. Comput. Vis. Pattern Recognit. (2016) 3630–3639, https://doi.org/10.1109/cvpr.2016.395.
[2] R. Gokberk Cinbis, J. Verbeek, C. Schmid, Multi-fold mil training for weakly supervised object localization, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (2014) 2409–2416.
[3] Y. Liu, X. Chen, H. Peng, and Z.J. I.F. Wang, "Multi-focus image fusion with a deep convolutional neural network," vol. 36, pp. 191–207, 2017.
[4] Y. Jiang, W. Zhang, K. Fu, Q. Zhao, MEANet: Multi-modal edge-aware network for light field salient object detection, Neurocomputing vol. 491 (2022) 78–90, https://doi.org/10.1016/j.neucom.2022.03.056.
[5] Z. Li, C. Lang, T. Wang, Y. Li, J. Feng, Deep spatio-frequency saliency detection, Neurocomputing vol. 453 (2021) 645–655, https://doi.org/10.1016/j.neucom.2020.05.109.
[6] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," 2012. [Online]. Available: https://doi.org/10.1109/CVPR.2012.6247743 http://doi.ieeecomputersociety.org/10.1109/CVPR.2012.6247743.
[7] M. Cheng, N.J. Mitra, X. Huang, P.H. S. Torr, S.J. I.T. o P.A. Hu, and M. Intelligence, "Global Contrast Based Salient Region Detection," vol. 37, no. 3, pp. 569–582, 2015.
[8] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic Saliency Using Background Priors," 2012. [Online]. Available: https://doi.org/10.1007/978-3-642-33712-3_3.
[9] G. Situ, Deep holography, Light: Advanced Manufacturing 3 (2022), LAM2021090031, https://doi.org/10.37188/lam.2022.013, 2689–9620.
[10] P. Zhang, D. Wang, H. Lu, H. Wang, B. Yin, Learning uncertain convolutional features for accurate saliency detection, Proc. IEEE Int. Conf. Comput. Vis. (2017) 212–221.
[11] P. Hu, B. Shuai, J. Liu, G. Wang, Deep level sets for salient object detection, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (2017) 2300–2309.
[12] Y. Liang, G. Qin, M. Sun, J. Qin, J. Yan, Z. Zhang, Multi-modal interactive attention and dual progressive decoding network for RGB-D/T salient object detection, Neurocomputing vol. 490 (2022) 132–145, https://doi.org/10.1016/j.neucom.2022.03.029.
[13] M. Wang, J. Konrad, P. Ishwar, K. Jing, and H.A. Rowley, "Image saliency: From intrinsic to extrinsic context," 2011. [Online]. Available: https://doi.org/10.1109/CVPR.2011.5995743. http://doi.ieeecomputersociety.org/10.1109/CVPR.2011.5995743.

[14] H. Chen, Y.F. Li, and Ieee, "Progressively Complementarity-aware Fusion Network for RGB-D Salient Object Detection," in 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, Jun 18–23 2018, in IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3051–3060, doi: 10.1109/cvpr.2018.00322. [Online]. Available: ://WOS:000457843603020.

[15] H. Chen, Y.F. Li, D. Su, Multi-modal fusion network with multi-scale multi-path and cross-modal interactions for RGB-D salient object detection, Pattern Recognit. vol. 86 (2019) 376–385, https://doi.org/10.1016/j.patcog.2018.08.007.

[16] J.W. Han, H. Chen, N. Liu, C.G. Yan, X.L. Li, CNNs-based RGB-D saliency detection via cross-view transfer and multiview fusion, IEEE Trans. Cybern. vol. 48 (11) (2018) 3171–3183, https://doi.org/10.1109/tcyb.2017.2761775.

[17] A. Borji, M. Cheng, H. Jiang, J. Li, Salient object detection: a benchmark, IEEE Trans. Image Process. vol. 24 (12) (2015) 5706–5722, https://doi.org/10.1109/TIP.2015.2487833.

[18] Y. Zhao, L. Zhang, D. Zhang, Q. Pan, Object separation by polarimetric and spectral imagery fusion, Comput. Vis. Image Underst. vol. 113 (8) (2009) 855–866, https://doi.org/10.1016/j.cviu.2009.03.002.

[19] M.N. Islam, M. Tahtali, and M.J. R.S. Pickering, "Specular Reflection Detection and Inpainting in Transparent Object through MSPLFI," vol. 13, no. 3, p. 455, 2021.

[20] N. Salamati, D. Larlus, G. Csurka, and S.S.J. Springer-Verlag, "Semantic Image Segmentation Using Visible and Near-Infrared Channels," 2012.

[21] X. Wang, J. Yuan, C. Yan, Y. Qiao, and Y.J. I.A. Dong, "Modified Degree of Polarization Function for Rough Metallic Surface Parameter Estimation Based on Multispectral Polarimetric Measurement," vol. PP, no. 99, pp. 1–1, 2020.

[22] J. Zhang, J. Shao, J. Chen, D. Yang, B. Liang, Polarization image fusion with self-learned fusion strategy, Pattern Recognit. vol. 118 (2021), 108045, https://doi.org/10.1016/j.patcog.2021.108045.

[23] A. Kalra, V. Taamazyan, S.K. Rao, K. Venkataraman, A. Kadambi, Deep polarization cues for transparent object segmentation, 2020 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) (2020).

[24] N. Li, Y. Zhao, R. Wu, Q. Pan, Polarization-guided road detection network for LWIR division-of-focal-plane camera, 2021/11/15, Opt. Lett. vol. 46 (22) (2021) 5679–5682, https://doi.org/10.1364/OL.441817.

[25] M. Yang, W. Xu, P. Xiu, W. Chen, and J.J. O.C. Li, "Degree of polarization modeling based on modified microfacet pBRDF model for material surface," vol. 453, p. 124390, 2019.

[26] X. Hou, L. Zhang, Saliency detection: a spectral residual approach. 2007 IEEE Conference on computer vision and pattern recognition, Ieee,, 2007, pp. 1–8.

[27] W. Ji, et al., DMRA: depth-induced multi-scale recurrent attention network for RGB-D saliency detection, IEEE Trans. Image Process. vol. 31 (2022) 2321–2336, https://doi.org/10.1109/tip.2022.3154931.

[28] H. Chen, Y.F. Li, Three-stream attention-aware network for RGB-D salient object detection, IEEE Trans. Image Process. vol. 28 (6) (2019) 2825–2835, https://doi.org/10.1109/tip.2019.2891104.

[29] Q. Chen, et al., RGB-D Salient Object Detection via 3D Convolutional Neural Networks, Feb 02-09 2021, vol. 35, in AAAI Conference on Artificial Intelligence. 35th AAAI Conference on Artificial Intelligence / 33rd Conference on Innovative Applications of Artificial Intelligence / 11th Symposium on Educational Advances in Artificial Intelligence, Electr Network, 2021, pp. 1063–1071. Feb 02-09 2021, vol. 35, in AAAI Conference on Artificial Intelligence.

[30] X. Jia, B. De Brabandere, T. Tuytelaars, L.V. Gool, Dynamic filter networks, Adv. Neural Inf. Process. Syst. vol. 29 (2016).

[31] N. Liu, J. Han, M.-H. Yang, PiCANet: Pixel-wise contextual attention learning for accurate saliency detection, IEEE Trans. Image Process. vol. 29 (2020) 6438–6451.

[32] J. He, Z. Deng, Y. Qiao, Dynamic multi-scale filters for semantic segmentation, Proc. IEEE/CVF Int. Conf. Comput. Vis. (2019) 3562–3572.

[33] Y. Pang, L. Zhang, X. Zhao, H. Lu, Hierarchical dynamic filtering network for rgb-d salient object detection. European Conference on Computer Vision, Springer, 2020, pp. 235–252.

[34] Q. Hou, D. Zhou, J. Feng, Coordinate attention for efficient mobile network design, Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (2021) 13713–13722.

[35] (a) M. Morimatsu, Y. Monno, M. Tanaka, M. Okutomi, Ieee, Monochrome and color polarization demosaicking using edge-aware residual interpolation. IEEE International Conference on Image Processing (ICIP), Electr Network, 2020;
(b) Sep 25-28 2020, IEEE Int. Conf. Image Process. ICIP (2020) 2571–2575 [Online]. Available: <Go to ISI>://WOS:000646178502136. [Online]. Available: <Go to ISI>://WOS:000646178502136.

[36] G. Courtier, P.-J. Lapray, J.-B. Thomas, and I. Farup, "Correlations in Joint Spectral and Polarization Imaging," vol. 21, no. 1, p. 6, 2021. [Online]. Available: https://www.mdpi.com/1424–8220/21/1/6.

[37] P.-c Zhou, F. Wang, H.-k Zhang, M.-g Xue, Camouflaged target detection based on visible and near infrared polarimetric imagery fusion, in: International Symposium on Photoelectronic Detection and Imaging 2011: Advances in Imaging Detectors and Applications, vol. 8194, SPIE, 2011, pp. 237–243.

[38] .T. Tan and K. Ikeuchi, "Reflection Components Decomposition of Textured Surfaces Using Linear Basis Functions," 2005. [Online]. Available: https://doi.org/10.1109/CVPR.2005.298. http://doi.ieeecomputersociety.org/10.1109/CVPR.2005.298.

[39] M.N. Islam, M. Tahtali, and M.J. R.S. Pickering, "Hybrid Fusion-Based Background Segmentation in Multispectral Polarimetric Imagery," vol. 12, no. 11, p. 1776, 2020.

[40] H. Su, V. Jampani, D. Sun, O. Gallo, E. Learned-Miller, J. Kautz, Pixel-adaptive convolutional neural networks, Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (2019) 11166–11175.

[41] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (2017) 4700–4708.

[42] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation. 2016 fourth international conference on 3D vision (3DV), IEEE, 2016, pp. 565–571.

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

[44] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv Prepr. arXiv 1412 (2014) *6980*.

[45] M. Morimatsu, Y. Monno, M. Tanaka, M. Okutomi, Monochrome and color polarization demosaicking using edge-aware residual interpolation. 2020 IEEE International Conference on Image Processing (ICIP), IEEE, 2020, pp. 2571–2575.

[46] R. Ju, Y. Liu, T. Ren, L. Ge, G. Wu, Depth-aware salient object detection using anisotropic center-surround difference, Signal Process.: Image Commun. vol. 38 (2015) 115–126.

[47] S.-Y. Chiu, C.-C. Chiu, and S.S.-D. Xu, "A Background Subtraction Algorithm in Complex Environments Based on Category Entropy Analysis," vol. 8, no. 6, p. 885, 2018. [Online]. Available: https://www.mdpi.com/2076–3417/8/6/885.

[48] M. Ran, L. Zelnikmanor, A. Tal, How to Evaluate Foreground Maps, IEEE (2014).

[49] Q. Zhai, X. Li, F. Yang, C. Chen, and D.P. Fan, "Mutual Graph Learning for Camouflaged Object Detection," 2021.

[50] J.X. Zhao et al., "EGNet: Edge Guidance Network for Salient Object Detection," presented at the 2019 IEEE/CVF INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV 2019), 2019.

[51] X. Qin, et al., Boundary-aware segmentation network for mobile and web applications, vol. abs/2101.04704, ArXiv (2021). vol. abs/2101.04704.

[52] Y. Pang, X. Zhao, L. Zhang, H. Lu, Multi-scale interactive network for salient object detection, Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (2020) 9413–9422.

[53] X. Zhao, Y. Pang, L. Zhang, H. Lu, L. Zhang, Suppress and balance: a simple gated network for salient object detection, 2020: Springer International Publishing, in Computer Vision – ECCV, Cham, 2020, pp. 35–51.

[54] Y.H. Wu, Y. Liu, L. Zhang, M.M. Cheng, B. Ren, EDN: salient object detection via extremely-downsampled network, IEEE Trans. IMAGE Process vol. 31 (2022) 3125–3136, https://doi.org/10.1109/TIP.2022.3164550.

[55] X.B. Qin, Z.C. Zhang, C.Y. Huang, M. Dehghan, O.R. Zaiane, M. Jagersand, U-2-Net: Going deeper with nested U-structure for salient object detection, PATTERN Recognit. vol. 106 (2020), 107404, https://doi.org/10.1016/j.patcog.2020.107404.

[56] J. Wei, S.H. Wang, Q.M. Huang, and I. Assoc Advancement Artificial, "F(3)Net: Fusion, Feedback and Focus for Salient Object Detection," presented at the Thirty-fourth AAAI conference on artificial intelligence, the thirty-second innovative applications of artificial intelligence conference and the tenth aaai symposium on educational advances in artificial intelligence, 2020.
[57] N. Liu, N. Zhang, K.Y. Wan, L. Shao, J.W. Han, and Ieee, "Visual Saliency Transformer," presented at the 2021 IEEE/CVF INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV 2021), 2021.
[58] W.X. Zhang, K.D. Ma, J. Yan, D.X. Deng, Z. Wang, Blind image quality assessment using a deep bilinear convolutional neural network, IEEE Trans. CIRCUITS Syst. VIDEO Technol. vol. 30 (1) (2020) 36–47, https://doi.org/10.1109/TCSVT.2018.2886771.
[59] H. Talebi, P. Milanfar, NIMA: Neural Image Assessment, IEEE Trans. IMAGE Process vol. 27 (8) (2018) 3998–4011, https://doi.org/10.1109/TIP.2018.2831899.
[60] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. vol. 21 (12) (2012) 4695–4708, https://doi.org/10.1109/TIP.2012.2214050.
[61] Y.Z. Hong, G.Q. Ren, E.H. Liu, A no-reference image blurriness metric in the spatial domain, OPTIK vol. 127 (14) (2016) 5568–5575, https://doi.org/10.1016/j.ijleo.2016.03.077.