A Novel Biologically Inspired Target Detection Method based on Saliency Analysis in SAR Imagery

Fei Ma[a], Fei Gao[a], Jun Wang[a,*], Amir Hussain[b,c], Huiyu Zhou[d]

[a]*School of Electronic and Information Engineering, Beihang University, Beijing 100191, China*
[b]*Cyber and Big Data Research Laboratory, Edinburgh Napier University, Edinburgh, UK*
[c]*Taibah Valley, Taibah University, Madinah, Saudi Arabia*
[d]*Department of Informatics, University of Leicester,Leicester,LE1 7RH,UK*

Abstract

Saliency Object Detection (SOD) models driven by the biologically-inspired Focus of Attention (FOA) mechanism can obtain highly accurate saliency maps. However, their application in the high-resolution Synthetic Aperture Radar (SAR) images faces some intractable problems due to complex background. In this paper, we propose a novel hierarchical self-diffusion saliency (HSDS) method for detecting vehicle targets in large scale SAR images. To reduce the influence of cluttered returns in saliency analysis, we learn a weight vector from the training set to capture the optimal initial saliency of the superpixels during saliency diffusion. Considering the background objects have multiple sizes, the saliency analysis is implemented in multi-scale space, and a saliency fusion strategy then is employed to integrate the multi-scale saliency maps. Simulation experiments demonstrate that our proposed method with these improvements can achieve more accurate detection and lead to less false alarms, compared to benchmark approaches.

Keywords: Biological Vision System, Saliency Detection, Focus of Attention (FOA), Target Detection, Synthetic Aperture Radar (SAR)

[*] Corresponding Author
Email address: wangj203@buaa.edu.cn (Jun Wang)

1. Introduction

SAR systems can implement all-weather observations for the targets of interest by providing better quality images than optical remote sensors [1]. At present, the widely used target detection algorithms based on SAR images include two-parameter constant false alarm rate (Two-CFAR) [2], Order Statistic CFAR (OS-CFAR) and Variability Index CFAR (VI-CFAR), etc. Recently, Gao et al. proposed an adaptive and fast CFAR algorithm for SAR target detection based on automatic censoring (AC)[3]. Cui et al. further proposed a target detection scheme that is based on iterative censoring [4]. The implementation of these pixel-wise methods is not efficient because they mainly search targets pixel by pixel according to the difference between the targets and the background clutters in the statistical models. Further, with the explosive growth of SAR data [5], these algorithms cannot afford high-speed data processing.

Compared with the above algorithms, the vision systems of the primates can process $10^8 - 10^9$ bit data per second, using their unique Focus of Attention (FOA) mechanism. FOA refers to the system in which the retinas can be directed to the most salient parts of the scenes using the scarcity of scene information. This mechanism can avoid the interference caused by target displacement or background clutters, demonstrating unparalleled efficiency and stability in signal processing. With the rapid development of neurophysiology and cognitive psychology in the last two decades, researchers have achieved encouraging results on FOA, such as Change Blindness, Attentional Blindness [6], Attentional Blink [7] and Central Fixation Bias[8], which have demonstrated the feasibility of exploiting the biological visual mechanism for SAR target detection.

The vision models driven by the FOA mechanism can be divided into two main categories. The first category, namely Fixation Prediction (FP) mechanism [9], generally achieves target detection by predicting the initial distribution of visual focus [10] within the first 3-5 seconds when the primates start observing a scene [11]. Representative models for processing optical images are Itti model [12] proposed by Itti et al. in 1997 and the later models such as Spectral Residual (SR)[13] and GBVS [14]. Harel et al. use spatial and temporal analysis to generate spatiotemporal saliency [15]. Yu et al. [16] introduce FP models in ship detection in SAR images. The papers [17][18] attempt to apply modified FP models in SAR vehicles detection. Wang et al. [19] propose a modified FP model to extract candidate target chips, which employs task-dependent scales, clustering and modified gist features to effectively detect and discriminate the targets. However, due to the existence of noises in eye tracking or observers' saccade landing, the estimation errors of these models are typically around 1-30 pixels. In this case, the FP models can only detect the approximate positions of the targets and cannot obtain their accurate locations.

Saliency Object Detection (SOD) is the another category driven by FOA [20], which defines the saliency detection as a binary segmentation problem on the basis of highly accurate saliency maps. It is commonly interpreted in detection as a process with two stages: 1) segmenting the images into a plurality of sub-regions (commonly referred to as superpixels [21]) and getting the initial saliency of each sub-region; 2) calculating the saliency of each sub-region using the spatial propagation algorithms. The representative SOD models include Saliency Optimization from Robust Background Detection (RBD)[22], GMR [23], HDCT [24]. Although the SOD models can significantly improve the detection precision for the optical images, they cannot obtain satisfactory results on the SAR images.

The performance degeneration of SOD for SAR images is caused by the differences between SAR images and optical images. First, the coherence between the radar echo signals leads to the strong speckle noises in SAR images. These noises will disturb the saliency initialization of superpixels during saliency diffusion and further affect the results of saliency detection. Second, the sizes of the objects (vehicles, ships, artificial buildings, mountain etc.) in the SAR images vary greatly. There have been several attempts to deal with these problems. For example, Wang et al. [25] propose an improved SOD models to detect the vehicles in SAR images. This method constructs a morphological saliency map, which can highlight the targets and suppress both the natural and man-made clutters via the targets' prior information.

This study attempts to present a hierarchical self-diffusion saliency (HSDS) detection method for detecting vehicle targets in SAR images, as depicted in Fig.1. For the speckle noise problem, an initial saliency optimization rule is introduced, inspired by [26] during the saliency diffusion via a graph model. Specifically, an optimal weight vector is learned from the training set, whose elements represent the contributions of different kinds of features to the initial saliency of the superpixels. The optimal initial saliency of each superpixel is the product of its feature vector and the learned weight vector. Then Manifold Ranking is employed for saliency propagation.

To address the problem caused by various sizes of objects, more scale spaces are considered in the process of the saliency detection, and we employ a hierarchical fusion rule inspired by [27] to integrate the saliency maps. In this strategy, the multiple saliency maps from different scale spaces are modelled as a 3D graph structure. During the fusion process, the saliency value of each pixel is updated based on the pixels in the adjacent lower and upper layers. From the top layer to the bottom layer, the saliency map is updated layer-by-layer following the proposed regulation scheme. Then a similar fusion operation starts from the bottom layer and ends in the top layer. In this case, the top saliency map will absorb the information from all the other layers and is used as the final saliency map.

As shown in Fig.1, for large-scale input SAR images, a proposal detection stage based on the edge cues are first employed to generate proposal chips that may contain vehicle targets. Then HSDS aims to precisely segment the targets from the proposal chips. Afterwards, the geometric properties of the targets, such as the areas, are introduced as the prior information to further reduce false alarms. The chips containing too large or too small targets will be removed from the detection results. Multiple sets of comparison verify that our proposed method outperforms state-of-the-art methods.

The remainder of this paper is organized as follows: Section 2 details the proposal detection stage; Section 3 describes the HSDS stage in detail, including the construction of the graph model, feature extraction, node initial saliency optimization, saliency propagation and hierarchical saliency fusion; Section 4 presents comparative experimental results, and Section 5 finally concludes this paper.

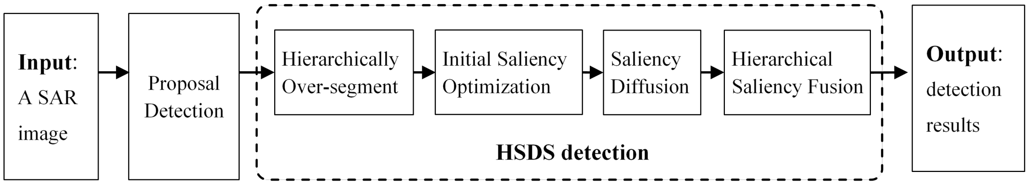


Figure 1: Framework of the proposed detection.

2. Proposal Detection Based on Edge Cues (EC)

Compared with the background objects (i.e., farmland, roads, tress and mountains) in SAR imagery, vehicle targets commonly contain sharp edges.

In this paper, we employ edge cues to quickly locate the suspicious targets in the large scene.

Specifically, a set of Difference of Gaussian (DOG) filters are first used to preprocess the large-scale input SAR images R_{in} to suppress the clutter and identify edges. Then we construct the edge cue-based saliency maps from the filtered images. The preprocessed results are named intensity maps, i.e., $F\left(s\right), s = [1, ..., S]$, empirically $S = 3$ in this paper. Since the non-linear distortion has been made during the DoG filtering process, a linear distortion correction is necessary for each intensity map as follows:

$$\tilde{F}\left(s\right) = alog_{10}\left[b \cdot F^2\left(s\right)\right] \tag{1}$$

where the constant coefficients a and b can be chosen with empirical values. Then we highlight the stimulus in the corners or edges while suppress the clutters in the maps as follows:

$$H\left(s,x,y\right) = \begin{cases} 0 & \tilde{F}\left(s,x,y\right) < 0 \\ \tilde{F}\left(s,x,y\right) & \tilde{F}\left(s,x,y\right) \geq 0 \end{cases} \tag{2}$$

$$I\left(s,x,y\right) = \frac{1}{\left\|H\left(s,x,y\right)\right\|_1} H\left(s,x,y\right) \tag{3}$$

where $\tilde{F}\left(s,x,y\right)$ denotes the intensity maps and $\left\|\cdot\right\|_1$ represents 1-norm. Eq. (2) means that all the negative pixels will be replaced by 0 and Eq. (3) realizes the normalization. $I\left(s,x,y\right), s = 1, 2, 3$ denotes the final intensity maps.

Next the hierarchical edge features are extracted from the intensity maps via self-information in Shannon theorem. Generally, the edges of the vehicles in the SAR images are close to straight lines. The intensity difference between the pixels at the edges of the vehicles and their adjacent pixels will mainly concentrate on a certain direction (perpendicular to the edges). Conversely, if the pixels are in the speckles, their difference with surrounding pixels will uniformly disperse in all the directions. Specifically, the 5×5 window centered on the pixel $I\left(s,x,y\right)$ are defined as the neighboring pixels of $I\left(s,x,y\right)$ as shown in Fig.2. In this window, the difference between $I\left(s,x,y\right)$ and $I\left(s,x+i,y+j\right)$ is defined as

$$L\left(s,x+i,y+j\right) = \varphi_{i,j} \frac{1}{\sqrt{2\pi}\sigma_s} \exp \frac{\left[I\left(s,x+i,y+j\right) - I\left(s,x,y\right)\right]^2}{2\sigma_s^2} \tag{4}$$

where σ_s denotes the standard deviation and $\varphi_{i,j}$ represents a spatial weighting term. Mathematically, $\varphi_{i,j}$= $\exp\left[-\left(i^2+j^2\right)/2\right]$, where i,j are the relative coordinate of two pixels. Farther neighboring pixel contributes less to edge features of the center pixel.

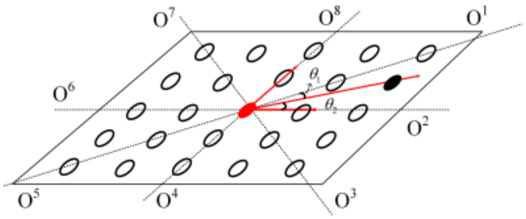


Figure 2: A 5 × 5 window centered around the red pixel.

Then, all the differences between the center pixel $I\left(s,x,y\right)$ and the neighboring pixels will be mapped to the 8 directions. For example, the pixel $I\left(s,x+2,y+1\right)$ (the solid, black circle in Fig. 2) locates between Orientation 1 and 2, and its angles to two directions are θ_1=$\pi/4\ -\arctan\left(1/2\ \right)$ and θ_2= $\arctan\left(1/2\ \right)$. The difference between two pixels can be mapped into these two directions, i.e., $L\left(s,x+2,y+1\right)\cdot\cos\theta_1$ and $L\left(s,x+2,y+1\right)\cdot\cos\theta_2$.

The projections of all neighboring pixels within the window in the direction k are summed up, and expressed as $O_k\left(s,x,y\right)$. The projections on eight directions are then normalized to ensure the sum of eight kinds of projections equals one. Finally, we use the self-information to indicate the edge features of the pixel $I\left(s,x,y\right)$:

$$O\left(s,x,y\right)=-\log\left(1-\tilde{O}\left(s,x,y\right)\right) \tag{5}$$

where $\tilde{O}\left(s,x,y\right)$ denotes the largest projection in eight $O_k\left(s,x,y\right)$. For the pixel $I\left(s,x,y\right)$ located in the edges of the vehicles, its $\tilde{O}\left(s,x,y\right)$ will be closer to one and hence is given higher saliency values. For three intensity maps $I\left(s,x,y\right)$, we can obtain three edge-based saliency maps $O\left(s,x,y\right),s=1,2,3$.

Fig. 3(a) shows a 1300 × 1200 (in pixels) SAR image with a resolution of 0.3m, where 10 vehicle targets are marked with the white rectangles. Fig. 3(b)-3(d) reports the three edge maps. As can be seen, these maps accurately maintain the shape, area and other topological features of objectives in different scale spaces and effectively reduce the interference from bushes and trees. Especially, many important boundaries and areas of interest are highlighted.

Three edge maps are summed to obtain the edge-based saliency map $S_e(x, y)$. As the regions with higher saliency usually considered as the suspicious areas (targets), an appropriate threshold T is used to carry out the binary processing for $S_e(x, y)$ and generate the proposal chips. Inspired by [20], T is defined as $T = \mu_s + c \cdot \sigma_s$, where μ_s and σ_s are the mean and variance of the saliency map, respectively. c is a constant empirically adjusted, setting as 0.3 in this paper. The regions whose saliency values are higher than threshold are classified as the suspicious targets. Then two kinds of geometric priori knowledge (area and major minor axis ratio) of targets are used to weed out the false alarms and obtain the candidates. As the false alarms from roads and buildings generally have far larger major minor axis ratio than that of vehicles, these regions can be removed from the results. Finally, a set of potential chips can be extracted from the input SAR images. In our experiment, the size of the proposal chips is set as 128×128, which is the same for the training images used in the subsequent node initial saliency optimization stage. As can be observed from the Fig.4(b), the whole 10 targets are detected (labeled by the red rectangles), while 19 false alarms are retained after the morphological operation.

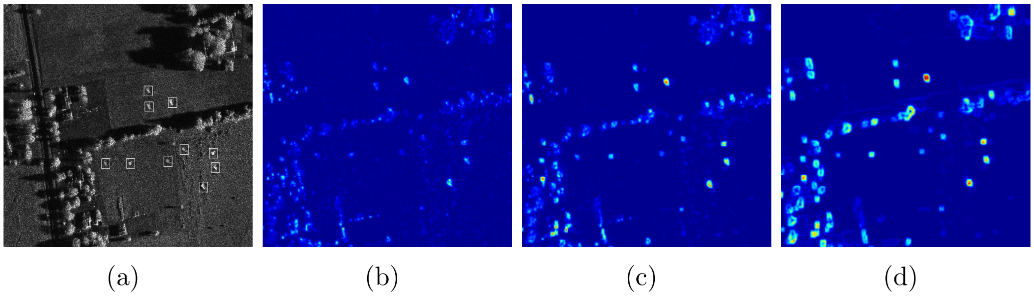


Figure 3: A high-resolution SAR image and its three edge maps. (a) A SAR image with 10 real vehicle targets, (b)-(d) edge maps $O(1)$, $O(2)$ and $O(3)$.

3. Hierarchically Self-Diffusion Saliency (HSDS) Detection

After the proposal detection, a set of potential chips $R_m, m = 1, ..., M$ are extracted from the input. In this stage, a hierarchical self-diffusion saliency

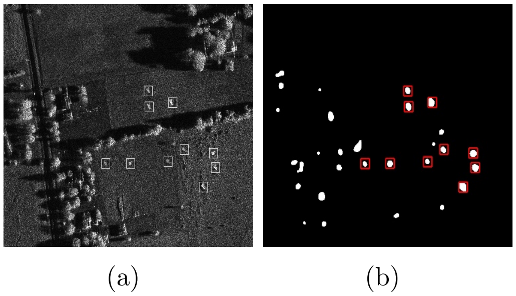

(a) (b)

Figure 4: The input SAR image and the results of proposal detection. (a) Input SAR image, (b) Results of the proposal detection stage (the red rectangles indicate the real targets, and the white regions represent the false alarms). In the results, the 10 targets are all detected, while 19 false alarms are retained.

detection method is designed to accurately detect the targets from these proposal chips. The graph diffusion is a common saliency detection algorithm, such as conditional random fields [28], quadratic energy models[29], random walks [30] and manifold ranking [31]. In these methods, the images are first partitioned into the graphs. Saliency information then propagates throughout the graph. In this section, we mainly promote the performance of the saliency propagation by improving the node initial saliency optimization strategy and the saliency map fusion rule.

3.1. Regional Features and Graph Construction

Each potential chip is first hierarchically over-segmented to the superpixels. Let R_m^g denotes the g_{th} layer of the over-segmentation results of m_{th} proposal chips. R_m^g can be modeled as a undirected graph $G_m^g\left(V,E\right)$, where $v_i \in V, i=1,...,N$ denotes a node (superpixel) and $w_{ij} \in E$ represents the edge linking the node pairs (i,j). In this paper, only the nodes sharing the boundaries are connected by edges. The edges can be expressed by an affinity matrix $W=\left[w_{ij}\right]_{N\times N}$, where w_{ij} is defined by

$$w_{ij} = \exp\left(-\sqrt{\left(T^i - T^j\right)^2}\right) \tag{6}$$

where T^i and T^j represent the feature vectors of two superpixels.

Most of previous SOD methods [23][32] usually adopt the average value in the CIELAB color space as the feature vector of the superpixel. But it may be ineffective for SAR images. In this section, we adopt a 38-dimensional feature vector, i.e., $T^i = [t_1^i, ..., t_{38}^i]^T$, which includes two types of features, i.e., regional contrast descriptor $X^{C,i}$ and backgroundness descriptor $X^{B,i}$, as shown in Table 1.

Intensity and context features			Contrast descriptor $X^{C,i}$			Backgroundness descriptor $X^{B,i}$		
Definition	Dim		Definition	Dim	T^i	Definition	Dim	T^i
Average gray value	1	a	$X_a^{C,i}$	1	t_1^i	$X_a^{B,i}$	1	t_{20}^i
Gray histogram	256	HG	$X_{HG}^{C,i}$	1	t_2^i	$X_{HG}^{B,i}$	1	t_{21}^i
LM texture vector	15	Lm	$X_{Lm}^{C,i}$	15	$t_3^i - t_{17}^i$	$X_{Lm}^{B,i}$	15	$t_{22}^i - t_{36}^i$
LM histogram	15	HL	$X_{HL}^{C,i}$	1	t_{18}^i	$X_{HL}^{B,i}$	1	t_{37}^i
LBP histogram	256	HB	$X_{HB}^{C,i}$	1	t_{19}^i	$X_{HB}^{B,i}$	1	t_{38}^i

Table 1: The elements of the regional feature vector.

3.1.1. Regional Contrast Descriptor

$R_m^{g,i}$ is the i_{th} superpixel of R_m^g. The regional contrast descriptor $X^{C,i}$ represents the uniqueness and scarcity of superpixel $R_m^{g,i}$ relative to all its neighboring nodes. First, five kinds of features are extracted for each superpixel $R_m^{g,i}$, i.e., average gray values $a_m^{g,i}$, gray histogram $HG_m^{g,i}$, LM (Leung-Malik) texture vector $Lm_m^{g,i}$, LM histogram $HL_m^{g,i}$ [33], and LBP histogram $HB_m^{g,i}$ [34]. For the remainder of the paper, we will drop the superscript g and subscript m of five classes of features for notational simplicity, i.e., a^i, HG^i, Lm^i, HL^i and HB^i. Among them, Lm^i is the mean of results from LM filtering. A LM filter set consists of fifteen 19 × 19 filters, including 2 Gaussian Laplacian filters, 1 Gaussian filter and 12 Gabor filters (6 directions, 2 phases). The filtering results for $R_m^{g,i}$ can be represented as $F\left(R_m^{g,i}\right) = \left[f_1\left(R_m^{g,i}\right), ..., f_{15}\left(R_m^{g,i}\right)\right]^T$ and then the k_{th} element of Lm^i is calculated by

$$Lm^i\left(k\right) = mean\left[f_k\left(R_m^{g,i}\right)\right] \tag{7}$$

To obtain the LM histogram HL^i, 15 filtered superpixels in $F\left(R_m^{g,i}\right)$ first emerge into one superpixels via a max-pooling operation, and then the pixel values are replaced by the corresponding sequence number of the filter. Hence, the dim of HL^i is 15.

The regional contrast descriptor from a^i is as follows:

$$X_a^{C,i} = \sum\nolimits_{j=1}^{N_{nei}^i} \beta_{ij} d\left(a^i, a^j\right) \tag{8}$$

where $d\left(a^i, a^j\right)$ is defined as $\left|a^i - a^j\right|$ to captures the differences between a^i and a^j. a^j is the average gray value of superpixel $R_m^{g,i}$ and N_{nei}^i denotes the number of neighboring nodes of $R_m^{g,i}$. $\beta_{ij}{=}\exp\left[-|p_i-p_j|^2/2\sigma_\beta^2\right]$ is a spatial weighting term, where p_i and p_j are the mean positions of $R_m^{g,i}$ and $R_m^{g,j}$. This means that the farther superpixel from $R_m^{g,i}$ has smaller effect on its saliency. σ_β controls the strength of spatially weighting effect, empirically set as 1 in our experiments.

The regional contrast descriptor from Lm^i is defined as

$$X_{Lm}^{C,i}{=}\sum\nolimits_{j=1}^{N_{nei}^i}\beta_{ij}d\left(Lm^i,Lm^j\right) \tag{9}$$

where $d\left(Lm^i,Lm^j\right)=\{\left|Lm^i\left(1\right)-Lm^j\left(1\right)\right|,...,\left|Lm^i\left(15\right)-Lm^j\left(15\right)\right|\}$, so the dimension of $X_{Lm}^{C,i}$ is 15. The elements of regional contrast descriptor from the histogram features (HG^i, HL^i and HB^i) have similar computing methods. With HG^i as an example,

$$X_{HG}^{C,i}{=}\sum\nolimits_{j=1}^{N_{nei}^i}\beta_{ij}d\left(HG^i,HG^j\right) \tag{10}$$

$d\left(HG^i,HG^j\right)$ is defined as

$$d\left(HG^i,HG^j\right)=\frac{\sum_{q=1}^{Q}\left(h_q^i-h_q^j\right)^2}{\sum_{q=1}^{Q}\left(h_q^i\right)^2} \tag{11}$$

where h_q^i is the q_{th} elements of HG^i and Q is the dimension of HG^i.

3.1.2. Backgroundness Descriptor

In addition to the neighboring superpixels, the differences between the superpixels and the background areas are also informative for saliency measurement. In general, the background is closer to the boundaries of the input images. Therefore, the superpixels at the border areas of R_m^g are selected as the background nodes $B_m^{g,j}$. We define the scarcity of a superpixel $R_m^{g,i}$ relative to all the background nodes as the backgroundness descriptor. Likewise, the backgroundness descriptor of the superpixel $R_m^{g,i}$ from a^i is defined as

$$X_a^{B,i}{=}\sum\nolimits_{j=1}^{N_{back}}\beta_{ij}d\left(a^i,a^j\right) \tag{12}$$

where N_{back} is the number of background superpixels and a^j denotes the gray value of a background superpixel $B_m^{g,j}$. The definition of $d\left(a^i,a^j\right)$ is same as Eq. (8). The other four kinds of elements of backgroundness descriptor

($X_{HG}^{B,i}$, $X_{Lm}^{B,i}$, $X_{HL}^{B,i}$,$X_{HB}^{B,i}$) have the similar calculation methods with that of regional contrast descriptor, except that the backgroundness descriptor does not consider the spatial weighting term β_{ij}. In summary, we can obtain a 38-dimensional regional feature vector T^i for each superpixel $R_m^{g,i}$, and then all superpixels in R_m^g can be modeled as a undirected graph $G_m^g\left(V,E\right)$ according to Eq. (6) .

3.2. Initial Saliency Optimization

In the saliency diffusion stage, for all the superpixels from R_m^g, we adopt the manifold ranking method proposed in [23] to propagate the saliency in the graph. Mathematically, the saliency of all the nodes after saliency diffusion can be expressed as:

$$\tilde{S}_m^g = \left(I - \frac{1}{1+u} D^{-1/2} W D^{-1/2}\right)^{-1} S_m^g \tag{13}$$

where S_m^g and $\tilde{S}_m^g$ represent the initial saliency and the final saliency of all the superpixels at the R_m^g, and u is a constant. D is the degree matrix of graph $G_m^g\left(V,E\right)$, which is defined as $D = diag\left(d_{11},\ldots,d_{nn}\right)$ and $d_{ii} = \sum\limits_j w_{ij}$.

The saliency propagation is sensitive to the initial saliency of nodes S_m^g, which contains the vital prior information and should be calculated carefully. In most of previous saliency diffusion methods, the initial saliency of nodes is calculated based on some priors. For example, the boundary prior assumes that the targets located in the central areas of images in general. Hence, the model in [23] selects the nodes on four sides of the images as the background seeds, and the initial saliency of these background seeds is set as zero. While the nodes in the central areas of the input images are regarded as the target seeds, whose initial saliency is set as one. However, this prior cannot provide enough target information for saliency diffusion in SAR chips. Inspired by [26], we propose a new node initial saliency optimization mechanism in this section. In this approach, we attempt to learn a weight vector from the training sets to calculate the optimal initial saliency of the nodes.

The contributions of each element of the feature vectors to the initial saliency of superpixels are diverse. Some features are mainly distributed in the background areas, while some features only occur in the target areas. Hence, we introduce a weight vector V^* to represent the contribution of different kinds of features to the superpixels' initial saliency. The initial saliency of each superpixel can be calculated by $S_m^{g,i} = T_m^{g,i} \times V^*$, where $T_m^{g,i}$

is the feature vector of superpixel $R_m^{g,i}$. The weight vector V^* can be learned by minimizing the following objective function:

$$V^* = \arg\min_V \left\{ \max\left(0, \left(1 - \sum_{i \in Tar} T_m^{g,i} * V^*\right)\right) + \max\left(0, \sum_{j \in Bg} T_m^{g,j} * V^*\right)\right\} \tag{14}$$

where $i \in Tar$ and $j \in Bg$ respectively indicate the superpixels belonging to the target regions and the background regions from the training set. The first item on the right side of the Eq. (14) ensures that the initial saliency of all target superpixels is close to one, while the second attempts to make that of background superpixels close to zero. Hence, the weight V^* can distinguish the targets from the background as much as possible.

In the experiment, the training set includes 20 vehicle chips from the Moving and Stationary Target Acquisition and Recognition (MSTAR) database and the corresponding manually labeled Ground Truth. These Ground Truth images are binary images where the target areas are one and the background areas are zero.

Once learned the optimal weight vector V^*, the initial saliency of the superpixels is the product of their feature vectors $T_m^{g,i}$ and V^*. The boundary priors are also considered to further optimize the initial saliency. The superpixels along the four sides of the images will be regarded as background seeds and their initial saliency is set to zero. Then, the saliency propagates in the graph $G_m^g(V, E)$ according to Eq. (13) , resulting in the optimal saliency values of all the nodes.

3.3. Hierarchical Saliency Fusion

In this section, the saliency maps $\tilde{S}_m^g$ from multi-scale spaces are integrated to the final saliency map. An average or sum operation cannot achieve ideal results [35]. Inspired by [27][36], we employ a fusion rule for combining the multi-layer saliency maps to achieve better detection results.

Let $\tilde{S}_{m,(k,l)}^g$ denote the pixel located at (k, l) of $\tilde{S}_m^g$. For notational simplicity, we drop the subscript m of the saliency maps $\tilde{S}_{m,(k,l)}^g$ in the remainder of the paper. During the fusion process, an objective function is designed to capture the final saliency maps. Integrating all the saliency maps at the same time in the objective function requires a large amount of calculation, so we first simplify the relation of multiple saliency maps as a 3D graph model. The salient maps are arranged in order of the superpixels' sizes, with the top layer having the largest superpixels. In this 3D graph model, pixel $\tilde{S}_{k,l}^{g}$ is only connected to the corresponding pixels in the upper and lower layer, i.e., $\tilde{S}_{k,l}^{g-1}$ and $\tilde{S}_{k,l}^{g+1}$. We optimize the saliency map layer by layer to realize the propagation of saliency information in the 3D graph. Specifically, our saliency propagating rule includes two stage: top-down fusion and bottom-up fusion. The top-down fusion stage starts from the second top layer, and the saliency map of each layer is updated layer-by-layer by minimizing the loss function until the bottom layer is updated. After the top-down stage, the bottom-up fusion starts from the second bottom layer until the top layer is optimized. After that, the top layer has successfully integrated the saliency information from other scale spaces. In each propagation, only adjacent layers need to be considered, which can decrease the difficulty of finding the optimal solution of the objective function. In the end, an updated saliency map will be selected as the final saliency map.

In the top-down fusion stage, for the g_{th} saliency map, we believe the optimized map $\hat{S}_{k,l}^{g}$ should be as consistent as possible with its initial saliency $\tilde{S}_{k,l}^{g}$. Moreover, $\hat{S}_{k,l}^{g}$ also should match its adjacent upper saliency map $\hat{S}_{k,l}^{g+1}$ to ensure $\hat{S}_{k,l}^{g}$ contains the saliency information from other scale spaces. Hence, the loss function of top-down fusion is as follows:

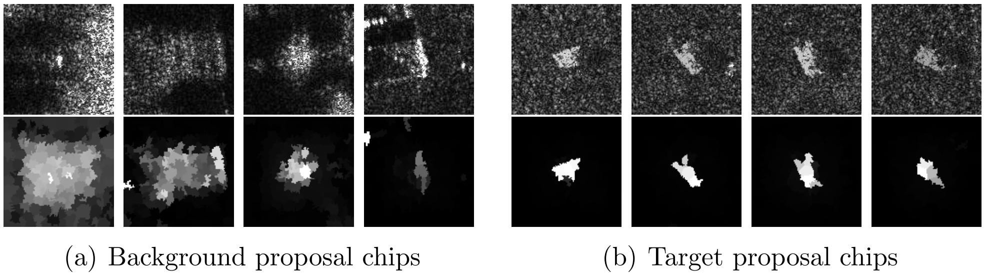
$$E_{T-B}\left(\hat{S}_{k,l}^{g}\right) = argmin\left\{\sum_{k,l}\left\|\hat{S}_{k,l}^{g} - \tilde{S}_{k,l}^{g}\right\|_2^2 + \sum_{k,l}\left\|\hat{S}_{k,l}^{g} - \hat{S}_{k,l}^{g+1}\right\|_2^2\right\} \tag{15}$$

The loss function comprises two parts, the first term ensuring that the updated saliency map $\hat{S}_{k,l}^{g}$ maintains maximum agreement with the original saliency map $\tilde{S}_{k,l}^{g}$. The second term can guarantee $\hat{S}_{k,l}^{g}$ maintains the maximum consistency with the updated saliency map of upper layer $\hat{S}_{k,l}^{g+1}$. The updated saliency map $\hat{S}_{k,l}^{g}$ is then used to optimize the its lower layer. The top-down fusion achieves saliency information propagating from the higher layers to lower layers.

In the bottom-up phase, the lower layers' saliency information propagates to the higher layers. Similarly, the loss function of bottom-up fusion is defined as:

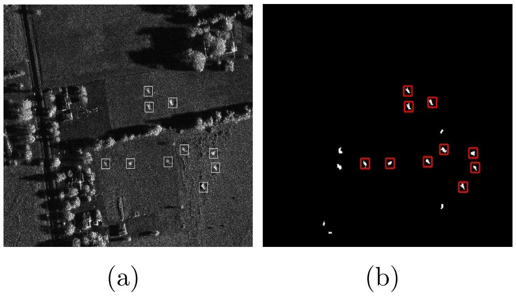
$$E_{B-T}\left(\hat{S}_{k,l}^{g}\right) = argmin\left\{\sum_{k,l}\left\|\hat{S}_{k,l}^{g} - \tilde{S}_{k,l}^{g}\right\|_2^2 + \sum_{k,l}\left\|\hat{S}_{k,l}^{g} - \hat{S}_{k,l}^{g-1}\right\|_2^2\right\} \tag{16}$$

In our method, the top saliency map layer is adopted as the result of the fusion stage. The first row of Fig.5 shows eight proposal chips from the proposal detection stage, including four real targets and four false alarms. The results of hierarchical saliency fusion are shown in the second row of Fig.5. Our method well preserves the contour features of the targets and suppresses the irregular clutters.


(a) Background proposal chips

(b) Target proposal chips

Figure 5: The results of HSDS detection for eight proposal chips. The first row: eight proposal chips, second row: results of saliency fusion.


(a)

(b)

Figure 6: The input SAR image and the detection results of the proposed method . (a) Input SAR image, (b) detection results of our method (the red areas indicate real targets). All the 10 targets are detected, while the number of false alarms drops from 19 to 6.

After acquiring the final saliency map, binarization is first implemented, and some geometric prior properties are then used to further remove background chips. In particular, for each proposal chip, maximum between-class variance method (OTSU) is first utilized to perform automatic image thresholding, which returns a single intensity threshold that separate the pixels of the saliency maps into two classes: targets and background. Second, the binary images are processed using morphologic operator (open operation) to remove the small regions caused by clusters or speckle noises. Third, prior knowledge about targets' sizes are introduced to further remove the background chips. In other words, the chips whose target regions are too large or too small will be deleted. The remaining proposal chips are considered as the real targets, forming the final detection results. As the HSDS's saliency map can accurately predict the regions of the targets, the postprocessing stage described above can effectively reduce the false alarms of the detection results. Fig.6(b) shows the detection results of our method, where the red rectangles indicate the real targets, and the white regions represent the false alarms. The false alarms drop from 19 to 6 and no target is missed.

To summarize the process, the particular flow of the proposed method is shown in Fig. 7. In the proposal detection stage, the edge cues are firstly extracted from input images. Then a set of proposal chips are selected from the input based on the edge feature maps. The following HSDS detection stage aims to locate the precise regions of the targets in the proposal chips. These chips are hierarchically over-segmented into superpixels by SLIC algorithm, which is followed by regional feature extraction and graph construction. Then a learned weight vector is used to capture the optimal initial saliency of the superpixels. After saliency propagating throughout the graph, we employ a fusion rule to combine the multiple saliency maps into the final results. Next, we postprocess the chips' saliency maps to remove the background chips in all potential chips. Specifically, binarization and open operation are performed on the saliency maps to remove the small regions caused by clusters or speckle noises. According to the geometric prior information of the targets, those chips whose target areas are too large or too small will be considered as the background chips and further removed from the detection results, thereby reducing the false alarm.

4. Experiments

In this section, we first describe the data sets used in the experiments and then discuss the detection performance of the proposed method for the large-scale high-resolution SAR images.

4.1. Description of Database

The images used in our experiments are from MSTAR with a resolution of 0.3m. The MSTAR public database was collected using the Sandia National Laboratories Twin Otter SAR sensor payload operating at X band, spotlight

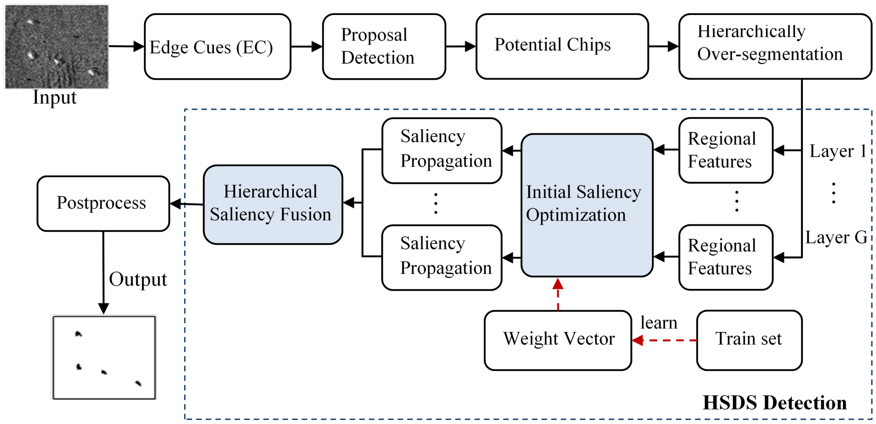


Figure 7: The detailed flowchart of the proposed method.

mode and HH single polarization. The MSTAR database includes 10 kinds of vehicle targets. Their optical and SAR images are shown in Fig.8.

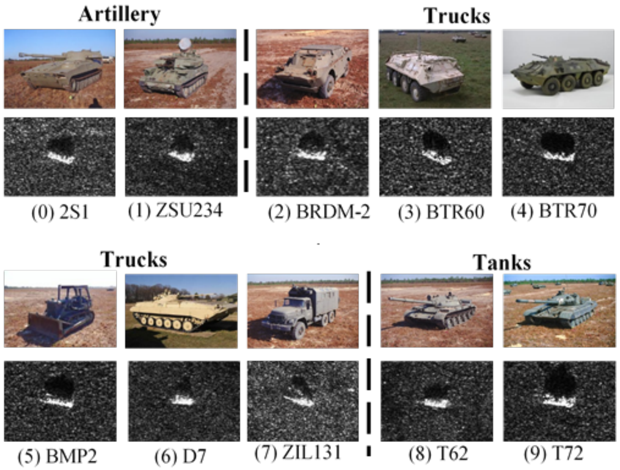
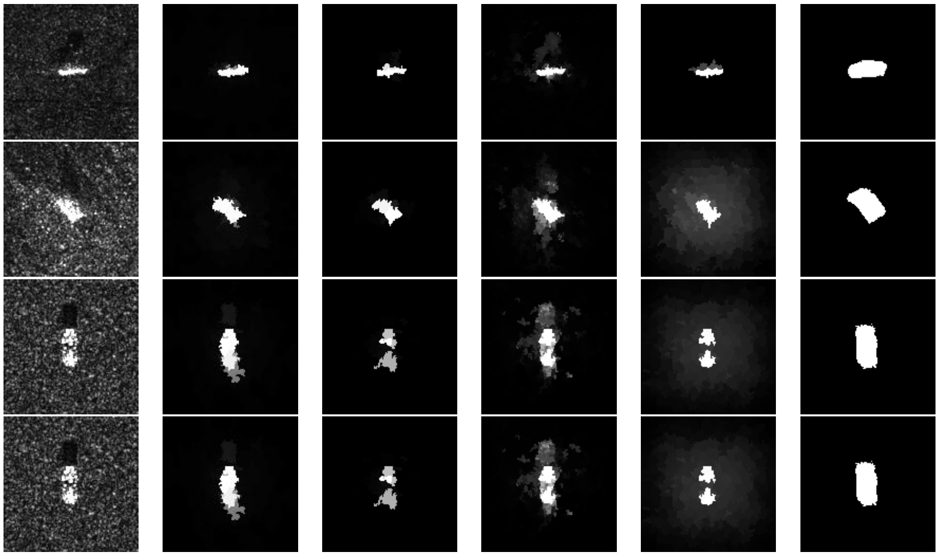


Figure 8: Optical and SAR images of 10 classes of targets in MSTAR database.

4.2. HSDS Detection Experiment

This section focuses on evaluating the HSDS's performance in saliency target detection. In the experiment, 20 images from the MSTAR dataset are selected as the potential chips, two images each class. And the HSDS method is used to calculate the saliency maps of the chips. The size of the superpixel determines the resolution of the detection results. Too small size will make the saliency propagation vulnerable to the background clutters. Instead, too large superpixels will decrease the precision of saliency detection. According to the geometric prior properties of the targets, each 128×128 proposal chip is respectively over-segmented into $\{100, 120, 140, 160\}$ superpixels in the experiment.

Fig.9 compares the results of HSDS and other three state-of-the-art saliency object detection models for four chips, i.e., Saliency Detection via Dense and Sparse Reconstruction (DSR) [37], RBD [22] and Saliency Detection via Absorbing Markov Chain (MC) [38]. Due to the application of node initial saliency optimization and hierarchical saliency fusion, HSDS tends to delineate the whole target region accurately. The heterogeneous clutters in the background regions have been effectively suppressed and corresponding saliency values are close to zero.



(a) Input (b) HSDS (c) RBD (d) DSR (e) MC (f) GT

Figure 9: Comparison of saliency maps of HSDS with that of three state-of-the-art SOD algorithms, MC, RBD and DSR. (a) Four potential chips from the MSTAR dataset. (b)-(e) Detection results of HSDS, RBD, DSR and MC. (f) Manually labeled Ground-truth.

We also introduce the precision-recall (PR) curve shown in Fig. 10 to quantitatively evaluate the performance of HSDS method. The precision value is defined as the ratio of the detected real targets assigned to all the detected targets, while the recall value corresponds to the percentage of the detected targets in relation to the ground-truth. Specifically, a threshold increasing from 0 to 255 is used to convert the saliency map to a binary image, then a pair of precision and recall values can be computed. Fig.10 shows the average PR curves of HSDS and MC, RBD and DSR on 20 chips. As can been seen, thanks to the hierarchical framework and the learned weight vector, HSDS method has a more competitive performance. The areas under the curve of HSDS method are larger than the other three methods.

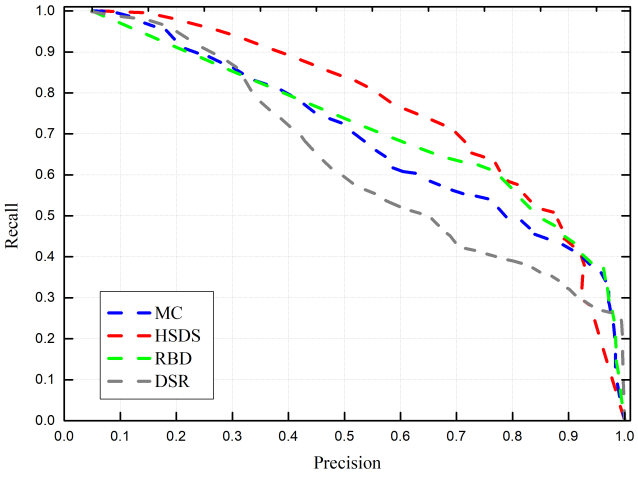


Figure 10: Comparison of the Precision-Recall curve of HSDS with that of three state-of-the-art algorithms on 20 potential chips from MSTAR.

4.3. Vehicle Target Detection Experiment

To verify the adaptability of the proposed method in vehicle detection, we select 20 targets from three kinds of tanks randomly (BMP, BRT70, T72) and add them into two SAR scene images with different SCRs (Signal Clutter Ratio), as shown in Fig.11(a) and Fig.12(a) , where targets are marked with red rectangles. The sizes of the background SAR images are both 1784 × 1476 (in pixel). In this work, SCR is defined as $SCR = u_T/u_B$, where u_T and

u_B are the mean value of the target region and background pixels in the target-centered 100×100 area, respectively.

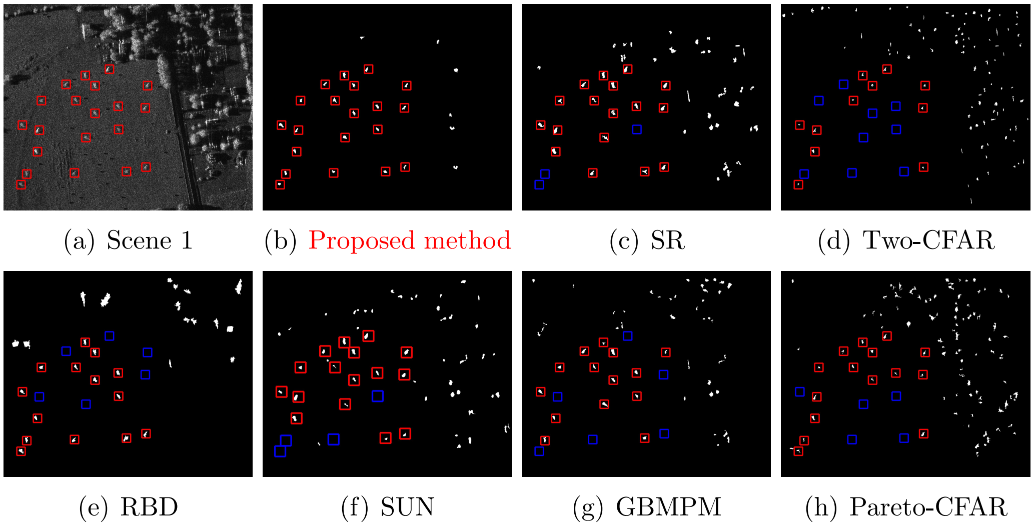
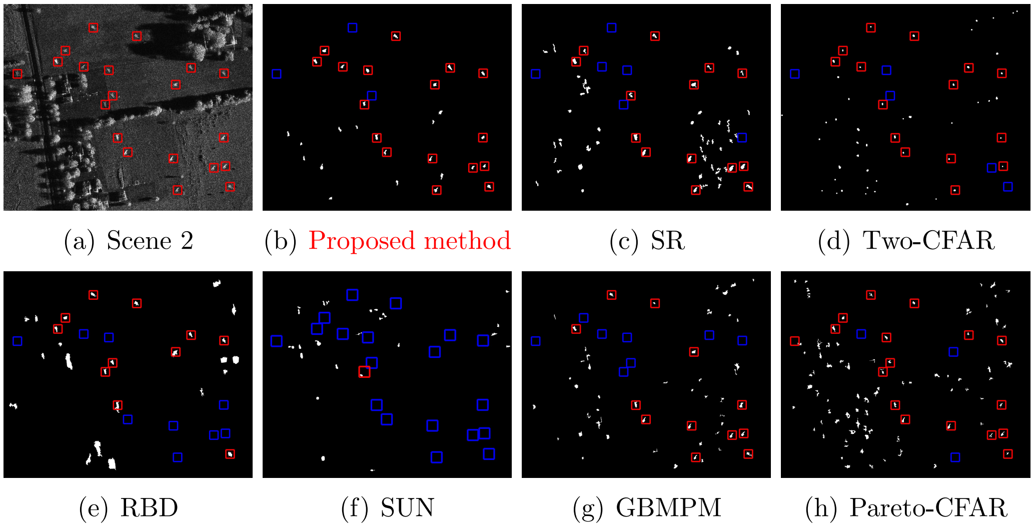


Figure 11: The detection results of four methods for Scene 1. (a) Scene 1: 20 targets are marked by red rectangles (SCR=2.2), (b)-(h) show the detection results of the proposed method, SR, two-parameter CFAR, RBD, SUN, GBMPM and Pareto-CFAR, respectively. The detected targets are marked by red rectangles, and the white regions represent the false alarms. The missed targets are labeled by blue rectangle rectangles.

The detection results of our method for the two scenes are given in Fig.11 and Fig.12, respectively. For the comparison, the results of other methods are also presented, i.e., two-parameter CFAR (Two-CFAR), SR[13], RBD, SUN [39], Gated Bi-directional Message Passing Module (GBMPM) [40] and Pareto-CFAR[41]. For Two-CFAR method, the adaptive threshold is calculated by keeping the false alarm rate constant. RBD is an unsupervised SOD method. SR model is a classical fixation prediction model, which extracts the spectral residuals of the images and constructs the corresponding saliency maps in the spectral domain. SUN employs a Bayesian framework to incorporate top-down information with bottom-up saliency for predicting human fixations. GBMPM is a supervised deep learning SOD model based on the Fully Convolutional Neural Network (FCN) [42]. GBMPM employs a bi-directional structure to pass messages among different layers of deep networks, so the output features can simultaneously encode semantic information and spatial details. The multi-level features are further utilized to produce the final saliency maps. In our experiments, we utilize the dataset used in the node initial saliency optimization stage to train GBMPM network. The hyperparameters of GBMPM are consistent with the original network shown in [40]. For example, learning rate is 0.001 and epoch is 10. Pareto-CFAR is a new CFAR scheme proposed by Graham [41],which introduces Pareto distribution into the radar community as a suitable model for X-band clutter returns. In order to ensure the consistent experimental conditions, a similar postprocessing is implemented for the results of above six methods, including binarization, morphological opening operation and removing background chips based on the geometric prior properties.

For scene1, the false alarms of our method can be controlled to 6 without missing real targets. While the false alarms of SR, two-CFAR, RBD, SUN, GBMPM and Pareto-CFAR, are 81, 31, 16, 36, 38 and 86 , respectively. Our method also has the least missed targets in all seven methods. For scene2, missed targets of seven methods are 3, 6, 5 ,9 19, 8 and 3. Our method has 11 false alarms, which is far less than the SR (25 false alarms) CFAR (38 false alarms), RBD (15 false alarms), SUN(21 false alarms), GBMPM (37 false alarms) and Pareto-CFAR (68 false alarms). The above results demonstrate that our model can obviously improve the performance of detection for detecting vehicles in the SAR images.


(a) Scene 2 (b) Proposed method (c) SR (d) Two-CFAR

(e) RBD (f) SUN (g) GBMPM (h) Pareto-CFAR

Figure 12: The detection results of six methods in scene 2. (a) Scene 2: 20 targets are marked by red rectangles (SCR=2.6). (b)-(h) show the detection results of the proposed method, SR, two-parameter CFAR, RBD, SUN, GBMPM and Pareto-CFAR, respectively.

To further verify the detection performance of the proposed method under various SCR, we employ F-measure to comprehensively evaluate the performance of our method. F-measure is computed by the weighted harmonic of precision and recall as follows:

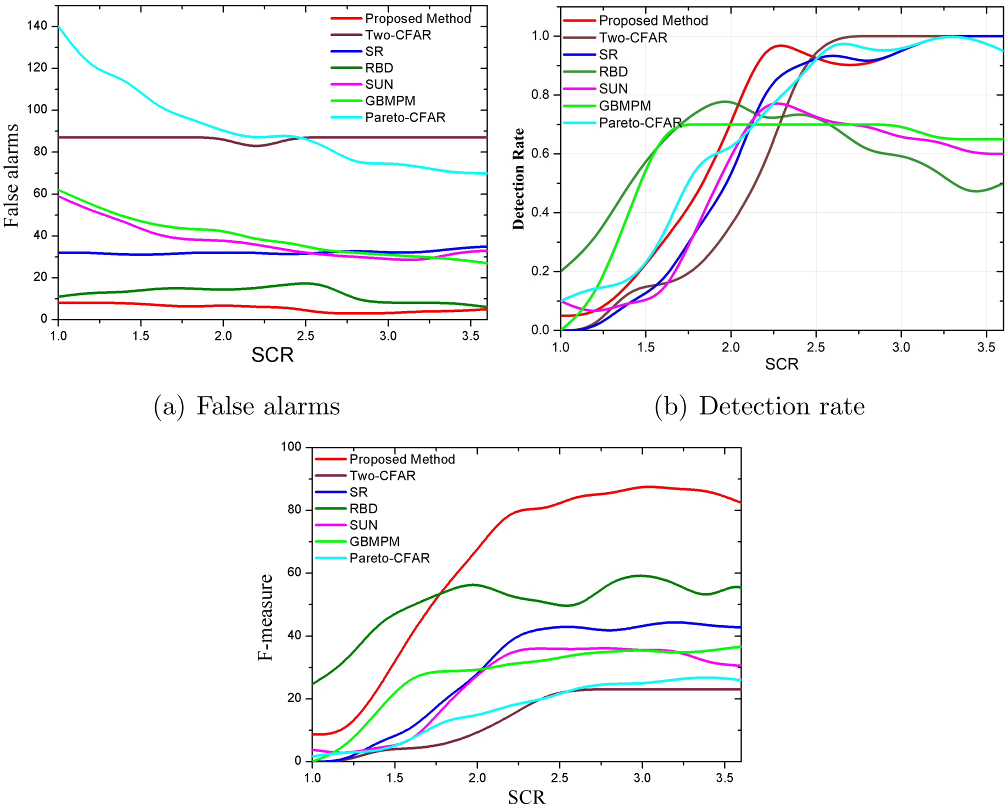
$$F_{\beta} = \frac{(1+\beta^2)\cdot Precision \cdot Recall}{\beta^2 \cdot Precision + Recall} \tag{17}$$

The constant β^2 represents the importance of the precision value in the F-measure, which is set to 0.3 inspired by [43].

Fig.13 compares the false alarms, detection rate and F-measure values of our method with that of SR, two-parameter CFAR, RBD, SUN, GBMPM and Pareto-CFAR. Fig.13(a) shows the false alarms of seven approaches when the SCR increases from 1 to 3.6. It can be seen that our method has the least false alarms for diverse SCRs. In Fig.13(c), our method has the largest F-measure values for $SCR > 1.6$. When SCR is less than 1.6, its F-measure is only worse than RBD and better than other five baseline methods. GBMPM based on the deep convolutional networks requires a large amount of training data [44]. As the training dataset contains only 20 vehicles chips, the performance of GBMPM is unsatisfactory. Two-parameter CFAR and Pareto CFAR are both pixelwise detection methods. They ignore the spatial relationship between the neighboring pixels during detection and both generate a large number of false alarms. These quantitative comparisons demonstrate that our method has better robustness.

4.4. Ship Detection Experiment

In order to verify the practicability of the proposed method, we utilize the proposed method to detect the ships on a SAR dataset [45] in this section. This dataset is proposed by Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences in March 2019, which is constructed using 108 Sentinel-1 images and 102 Chinese Gaofen-3 images. For Gaofen-3, the images have resolutions of 3 m, 5 m, 8 m, and 10 m with Ultrafine Strip-Map (UFS), Fine Strip-Map 1 (FSI), Full Polarization 1 (QPSI), Full Polarization 2 (QPSII), and Fine Strip-Map 2 (FSII) imaging modes, respectively. For Sentinel-1, the imaging modes are S3 Strip-Map (SM), S6 SM, and IW-mode. These SAR images are cropped to acquire ship chips 256 × 256 pixels in size, which are then labeled by SAR experts with LabelImg. Each ship chip corresponds to an



(a) False alarms

(b) Detection rate

(c) F-measure

Figure 13: Quantitative comparisons of the proposed method and six baseline methods under various SCRs for scene 1.

Extensible Markup Language (XML) file like that in the PASCAL VOC detection dataset, indicating the ship location, the ship chip name, and the image shape, respectively. This dataset totally includes 43,819 ship chips of 256 pixels in both range and azimuth.We randomly select 1000 images from the dataset, and 100 of them are used to train the weight vector during node initial saliency optimization stage.

We first utilize the edge cues to extract the potential chips on this dataset. Then HSDS is employed to segment the precise regions of the ships in the proposal chips. Next, binarization and open operation are performed on the saliency maps from HSDS method to remove the small regions caused by speckle noises. Finally, prior knowledge about targets' sizes are introduced to further remove the background chips. These chips with too large or too small target regions will be regarded as the background chips and removed from the detection results. Fig.14 compares some detection results of our method and several common saliency detection methods. It can be seen that our method can accurately get the shapes of ships. Some targets close to the boundaries of the images are also be detected. The results of RBD, DSR and MC are more susceptible to the strong speckle noises in SAR images. It is because they define the salient targets from the perspective of pixel value contrast, while the proposed method employs a lot of texture feature information. We also attempt to detect ships using Two-CFAR, SR, SUN, GBMPM and Pareto-CFAR methods. However, the strong sea clutters and speckle noises of SAR images make their detection results unsatisfactory, so we do not show them in Fig.14.

We also summarize the number of false alarms and detection rate of the proposed method on this ship dataset in Table 2. As a comparison, the detection results of the other three methods are also given. As can be seen, our method has the highest detection rate and least false alarms, which is consistent with the results of Fig.14. These testing experiments on the ship database prove that the proposed method can reduce the interference of clutters and speckle noises on the detection results by comprehensively utilizing various features of the targets.

Methods	Ours	RBD	DSR	MC
false alarms	102	235	461	258
detection rate	0.856	0.737	0.511	0.749

Table 2: Comparison of the performance of different detection methods on the ship dataset.

4.5. The Role of Initial Saliency Optimization

In order to clearly reflect the function of the node initial saliency optimization stage for the performance of the proposed method, we give the detection results of our method for scene 1 without the learned weight vector. In this case, all the elements of the weight vector will be set to 1 instead of learning from the training set.

Table 3 shows the numbers of the missed targets, false alarms and F-measure values of our method with and without the node initial saliency optimization stage for scene1. As can be seen, all 20 targets are detected in

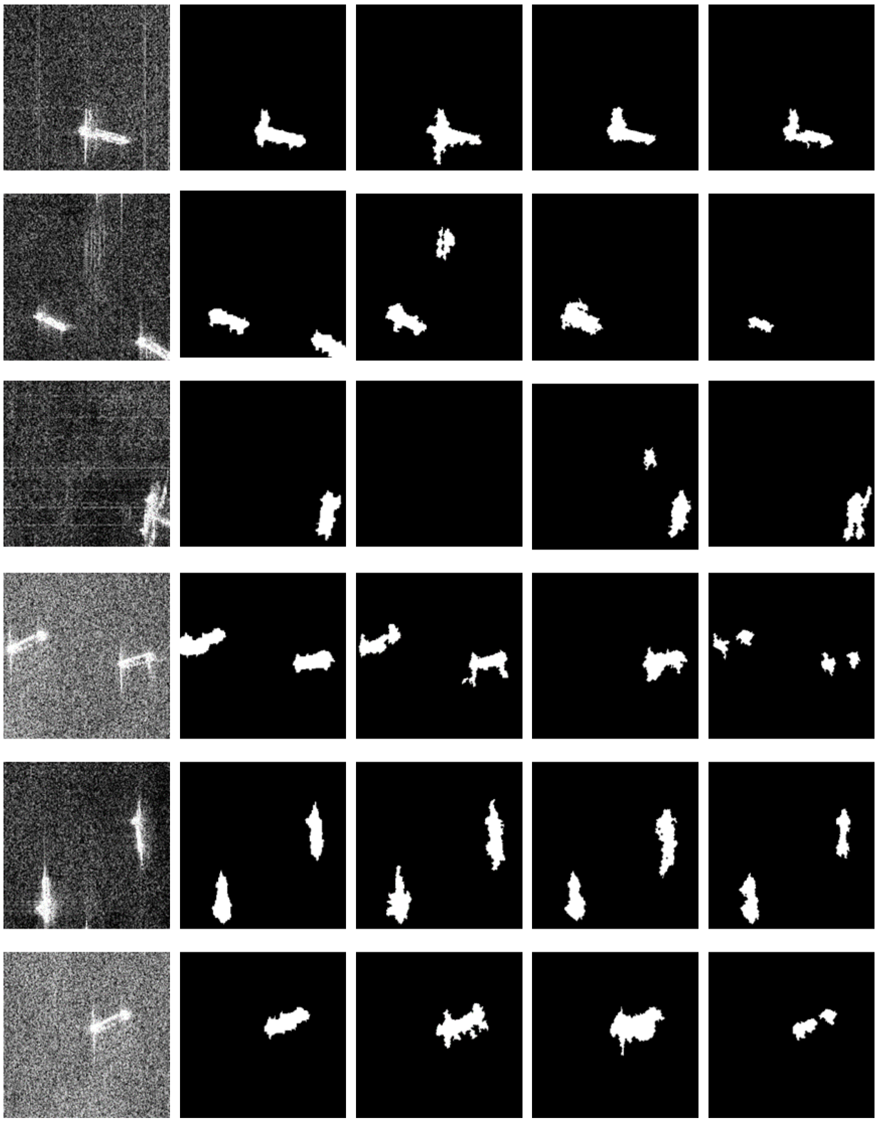


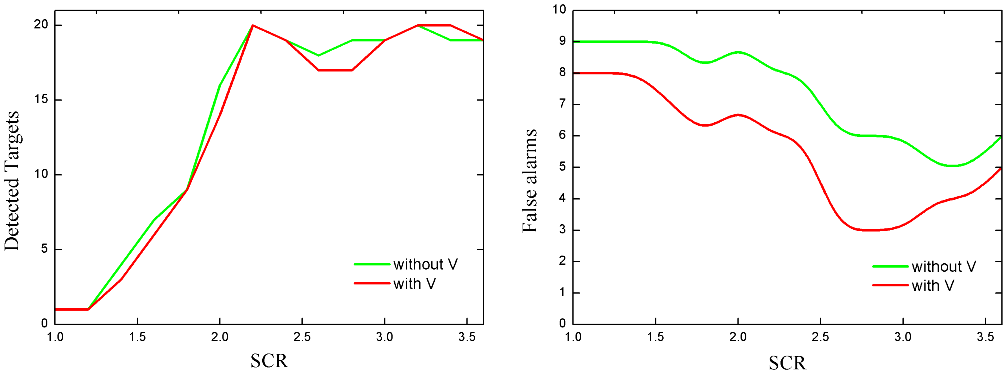
Figure 14: Comparison of the detection results on ship dataset of our method with three state-of-the-art algorithms. (a) SAR images (b)-(e) Detection results of Ours, RBD, DSR and MC.

two cases but the initial saliency optimization stage can decrease the false alarms from 8 to 6.

Furthermore, we give the numbers of the detected targets and false alarms in two cases when the SCR increases from 1 to 3.6 in Fig.15. For the approach without node initial saliency optimization stage, its number of detected targets is seen to be as well as the proposed method for diverse SCRs, but its false alarms are always greater than that of ours. This proves that the prior acknowledge learned from the training set contributes to reducing the false alarms caused by the clutters and improving the detection performance.

	Missed targets	False alarms	F-measure(%)
PM with V^*	0	8	81.3
PM without V^*	0	6	76.5

Table 3: Change in performance of the proposed method upon removal of the node initial saliency optimization stage for scene 1 (SCR=2.2).



(a) Detected targets

(b) False alarms

Figure 15: Change in detected targets and false alarms of the proposed method when initial saliency optimization stage is removed from the proposed method for scene 1 under various SCRs.

4.6. The Role of Hierarchical Saliency Fusion

This section aims to verify whether the fusion rule can enhance the detection results. First, Table 4 gives the precision, recall and F-measure values of the proposed method for scene 1 (SCR=2.2) when each potential chip is over-segmented in single scale space. The number of the superpixels is respectively 100, 120, 140 and 160. We also give the detection results of our method taking average operation as the fusion strategy in Table 4, which is termed PM-Average. We can see that the proposed method (PM) obviously has the highest precision, recall and F-measure than the other four single-scale models and PM-Average.

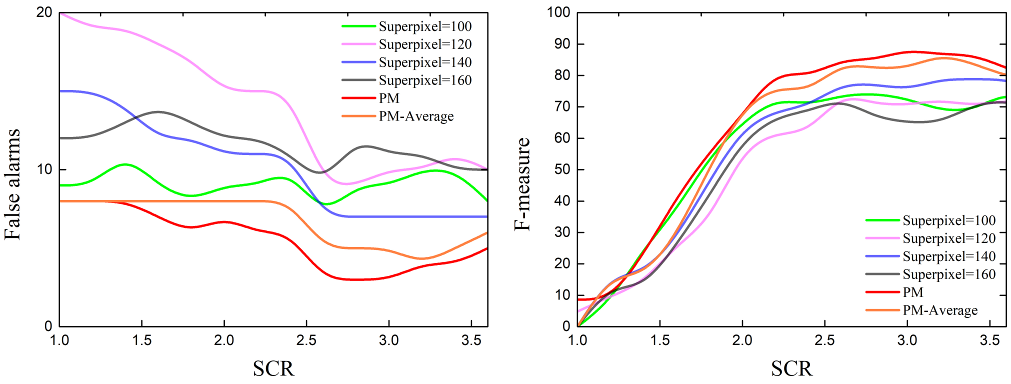
Number of Superpixel	Precision	Recall	F-measure
100	67.9	95	72.6
120	55.9	95	61.8
140	63.3	95	68.6
160	61.3	95	66.8
PM-Average	71.4	100	76.5
PM	76.9	100	81.3

Table 4: Change in performance when the proposed method only considers single-scale saliency map for scene 1 (SCR=2.2) or takes the average operation as its fusion strategy.

Moreover, Fig.16 shows the variations of false alarms and F-measure values of five methods for scene 1 under different SCRs. Fig.16(a) demonstrates that our fusion operation can significantly reduce the number of the false alarms, outperforming average operation and single layer methods. In Fig. 16(b), the performance of the four single layer methods is seen to be especially vulnerable to SCR. For example, the approach with 100 superpixels has better F-measure values than the other three single layer methods when SCR is less than 2. However, when the SCR is in range of 2 to 3.6, the best single layer method becomes the approach with 140 superpixels. On the contrary, our approach can keep the best F-measure values compared to the other five methods. These results further prove that the hierarchical saliency fusion stage can improve detection performance by more effectively integrating background and targets features from multiple scale spaces.

4.7. Running Time

To evaluate the efficiency of our method, we compare the average computational time for per SAR image of our method with six state-of-the-art methods, as shown in Table 5.The simulation software is MATLAB 2017a, and the main configuration of the computer includes 8GB RAM and Intel Core i5-8265U CPU. The computational time of our method is less than that of Two-CFAR and SUN. CFAR takes each pixel as a processing unit, seriously affecting its computational efficiency. Pareto-CFAR significantly reduces its running time by employing the Pareto distribution to model the clutter returns. Although SR saves much running time by calculating saliency in the



(a) False alarms

(b) F-measure

Figure 16: Change in false alarms and F-measure values of the proposed method when the hierarchical saliency fusion stage is removed and only single-scale saliency map is considered for scene 1 under various SCRs.

frequency domain space, its detection results are unsatisfactory. RBD and the proposed method both take superpixel as the basic processing unit, and hence have the similar running time.

Method	Ours	SR	Two-CFAR	RBD	SUN	GBMPM	Pareto-CFAR
Times(s)	30.91	2.58	633.37	12.62	214.61	4.79	13.87

Table 5: Comparison of average run time (seconds per image).

5. Conclusion

We have presented a novel hierarchical self-diffusion saliency (HSDS) method for detecting vehicle targets in large-scale SAR images. During the HSDS based detection, a weight vector learned from the training set is introduced to calculate the optimal initial saliency of the nodes. We further design a saliency fusion rule to integrate multiple saliency maps in order to locate accurate regions of the targets. Simulation experiment results verify that these improvements can effectively decrease false alarms and increase the stability of detection performance. Benchmark comparisons with CFAR and other state-of-the-art saliency detection methods demonstrate the wider applicability of our proposed model.

In terms of limitations, it should be noted that the node initial saliency optimization stage proposed in our method requires a certain number of vehicle chips to train. However, it is hard to obtain enough training chips in many practical application, hence the detection results of our method for these SAR images are not ideal. In future research, we will attempt to reduce the algorithm's reliance on training samples. We also plan to integrate task-related information in saliency calculations in order to further improve the accuracy of target detection.

Acknowledgement

Funding: This work was supported by the National Natural Science Foundation of China [grant numbers 61771027, 61071139, 61471019, 61501011, and 61171122]; H Zhou was supported by UK EPSRC under Grants [grant numbers EP/N508664/1, EP/R007187/1 and EP/N011074/1]; Royal Society-Newton Advanced Fellowship [grant number NA160342].Fei Ma was supported by Academic Excellence Foundation of BUAA for PhD Students. A. Hussain was supported by UK Engineering and Physical Sciences Research Council (EPSRC) grant no. EP/M026981/1.

The authors are grateful to the anonymous reviewers for their insightful comments and suggestions, which have helped improve the quality of this paper.

References

[1] F. Zhang, X. Yao, H. Tang, Q. Yin, Y. Hu, B. Lei, Multiple mode sar raw data simulation and parallel acceleration for gaofen-3 mission, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11 (6) (2018) 2115–2126 (2018).

[2] L. M. Novak, S. D. Halversen, G. Owirka, M. Hiett, Effects of polarization and resolution on sar atr, IEEE Transactions on Aerospace and Electronic Systems 33 (1997) 102–116 (1997).

[3] G. Gao, L. Liu, L. Zhao, G. Shi, G. Kuang, An adaptive and fast cfar algorithm based on automatic censoring for target detection in high-resolution sar images, Geoscience and Remote Sens-

ing, IEEE Transactions on 47 (2009) 1685 – 1697 (07 2009). doi:10.1109/TGRS.2008.2006504.

[4] C. Yi, G. Zhou, Y. Jian, Y. Yamaguchi, On the iterative censoring for target detection in sar images, IEEE Geoscience and Remote Sensing Letters 8 (4) (2011) 641–645 (2011).

[5] H. Chen, F. Zhang, B. Tang, Q. Yin, X. Sun, Slim and efficient neural network design for resource-constrained sar target recognition, Remote Sensing 10 (10) (2018) 1618 (2018).

[6] D. J. Parkhurst, K. Law, E. Niebur, Modeling the role of salience in the allocation of overt visual attention, Vision Research 42 (2002) 107–123 (2002).

[7] N. D. B. Bruce, J. K. Tsotsos, Saliency based on information maximization, in: NIPS, 2005 (2005).

[8] T. Ho-Phuoc, N. Guyader, A. Guérin-Dugué, A functional and statistical bottom-up saliency model to reveal the relative contributions of low-level visual guiding factors, Cognitive Computation 2 (2010) 344–359 (2010).

[9] G. M. Underwood, Cognitive processes in eye guidance: Algorithms for attention in image processing, Cognitive Computation 1 (2008) 64–76 (2008).

[10] V. Yanulevskaya, J. B. Marsman, F. Cornelissen, J.-M. Geusebroek, An image statistics–based model for fixation prediction, in: Cognitive Computation, 2010 (2010).

[11] G. Kootstra, B. de Boer, L. Schomaker, Predicting eye fixations on complex visual stimuli using local symmetry, in: Cognitive Computation, 2010 (2010).

[12] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (2009) 1254–1259 (2009).

[13] X. Hou, L. Zhang, Saliency detection: A spectral residual approach, 2007 IEEE Conference on Computer Vision and Pattern Recognition (2007) 1–8 (2007).

[14] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, in: NIPS, 2006 (2006).

[15] J. Tünnermann, B. Mertsching, Region-based artificial visual attention in space and time, Cognitive Computation 6 (2013) 125–143 (2013).

[16] Y. Yu, B. Wang, L. Zhang, Hebbian-based neural networks for bottom-up visual attention and its applications to ship detection in sar images, Neurocomputing 74 (2011) 2008–2017 (2011).

[17] Z. Yue, F. Gao, Q. Xiong, J. Wang, T. Huang, E. Yang, H. Zhou, A novel semi-supervised convolutional neural network method for synthetic aperture radar image recognition, Cognitive Computation (2019) 1–12 (2019).

[18] F. Ma, F. Gao, J. Sun, H. Zhou, A. Hussain, Weakly supervised segmentation of sar imagery using superpixel and hierarchically adversarial crf, Remote Sensing 11 (5) (2019) 512 (2019).

[19] Z. Wang, L. Du, P. Zhang, L. Li, F. Wang, S. Xu, H. Su, Visual attention-based target detection and discrimination for high-resolution sar images in complex scenes, IEEE Transactions on Geoscience and Remote Sensing 56 (4) (2017) 1855–1872 (2017).

[20] F. Huang, J. Qi, H. Lu, L. Zhang, X. Ruan, Salient object detection via multiple instance learning, IEEE Transactions on Image Processing 26 (2017) 1911–1922 (2017).

[21] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, Slic superpixels compared to state-of-the-art superpixel methods, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (2012) 2274–2282 (2012).

[22] W. Zhu, S. Liang, Y. Wei, J. Sun, Saliency optimization from robust background detection, 2014 IEEE Conference on Computer Vision and Pattern Recognition (2014) 2814–2821 (2014).

[23] C. Yang, L. Zhang, H. Lu, X. Ruan, M.-H. Yang, Saliency detection via graph-based manifold ranking, 2013 IEEE Conference on Computer Vision and Pattern Recognition (2013) 3166–3173 (2013).

[24] J. Kim, D. Han, Y.-W. Tai, J. Kim, Salient region detection via high-dimensional color transform and local spatial support, IEEE Transactions on Image Processing 25 (2014) 9–23 (2014).

[25] Z. Wang, D. Lan, H. Su, Target detection via bayesian-morphological saliency in high-resolution sar images, IEEE Transactions on Geoscience and Remote Sensing PP (99) (2017) 1–12 (2017).

[26] S. Lu, V. Mahadevan, N. Vasconcelos, Learning optimal seeds for diffusion-based salient object detection, 2014 IEEE Conference on Computer Vision and Pattern Recognition (2014) 2790–2797 (2014).

[27] F. Gao, F. Ma, J. Wang, J. Sun, H. Zhou, Visual saliency modeling for river detection in high-resolution sar imagery, IEEE Access PP (99) (2017) 1–1 (2017).

[28] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, H. Shum, Learning to detect a salient object, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (2007) 353–367 (2007).

[29] K.-Y. Chang, T.-L. Liu, H.-T. Chen, S.-H. Lai, Fusing generic objectness and visual saliency for salient object detection, 2011 International Conference on Computer Vision (2011) 914–921 (2011).

[30] V. Gopalakrishnan, Y. Hu, D. Rajan, Random walks on graphs for salient object detection in images, IEEE Transactions on Image Processing 19 (2010) 3232–3242 (2010).

[31] D. Zhou, J. Weston, A. Gretton, O. Bousquet, B. Schölkopf, Ranking on data manifolds, in: Advances in neural information processing systems, 2004, pp. 169–176 (2004).

[32] H. Jiang, Z. Yuan, M.-M. Cheng, Y. Gong, N. Zheng, J. Wang, Salient object detection: A discriminative regional feature integration approach, 2013 IEEE Conference on Computer Vision and Pattern Recognition (2013) 2083–2090 (2013).

[33] D. Hoiem, A. A. Efros, M. Hebert, Geometric context from a single image, Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1 1 (2005) 654–661 Vol. 1 (2005).

[34] T. K. Leung, J. Malik, Representing and recognizing the visual appearance of materials using three-dimensional textons, International Journal of Computer Vision 43 (2001) 29–44 (2001).

[35] Q. Yan, L. Xu, J. Shi, J. Jia, Hierarchical saliency detection, 2013 IEEE Conference on Computer Vision and Pattern Recognition (2013) 1155–1162 (2013).

[36] D. A. Klein, S. Frintrop, Center-surround divergence of feature statistics for salient object detection, 2011 International Conference on Computer Vision (2011) 2214–2219 (2011).

[37] X. Li, H. Lu, L. Zhang, X. Ruan, M.-H. Yang, Saliency detection via dense and sparse reconstruction, 2013 IEEE International Conference on Computer Vision (2013) 2976–2983 (2013).

[38] B. Jiang, L. Zhang, H. Lu, C. Yang, M.-H. Yang, Saliency detection via absorbing markov chain, 2013 IEEE International Conference on Computer Vision (2013) 1665–1672 (2013).

[39] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, G. W. Cottrell, Sun: A bayesian framework for saliency using natural statistics, Journal of vision 8 (7) (2008) 32–32 (2008).

[40] L. Zhang, J. Dai, H. Lu, Y. He, G. Wang, A bi-directional message passing model for salient object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1741–1750 (2018).

[41] G. V. Weinberg, Constant false alarm rate detectors for pareto clutter models, IET Radar, Sonar & Navigation 7 (2) (2013) 153–163 (2013).

[42] Y.-L. Huang, B.-B. Xu, S.-Y. Ren, Analysis and pinning control for passivity of coupled reaction-diffusion neural networks with nonlinear coupling, Neurocomputing 272 (2018) 334–342 (2018).

[43] R. Achanta, S. S. Hemami, F. J. Estrada, S. Süsstrunk, Frequency-tuned salient region detection, 2009 IEEE Conference on Computer Vision and Pattern Recognition (2009) 1597–1604 (2009).

[44] G. Fei, H. Teng, J. Sun, J. Wang, A. Hussain, E. Yang, A new algorithm of sar image target recognition based on improved deep convolutional neural network, Cognitive Computation (5) (2018) 1–16 (2018).

[45] Y. Wang, C. Wang, H. Zhang, Y. Dong, S. Wei, A sar dataset of ship detection for deep learning under complex backgrounds, Remote Sensing 11 (7) (2019). doi:10.3390/rs11070765.