Received 11 November 2022, accepted 25 November 2022, date of publication 1 December 2022, date of current version 7 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3225918

RESEARCH ARTICLE

Joint Learning of Discriminative Metric Space From Multi-Context Visual Scene for Unsupervised Salient Object Detection

SHIGANG WANG
School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China
e-mail: sgwang@nwpu.edu.cn

This work was supported in part by the National Natural Science Foundation of China under Grant 61906157; in part by the Postdoctoral Innovative Talent Support Program under Grant BX20190282; in part by the China Postdoctoral Science Foundation under Grant 2020M673490; in part by the Young Talent Fund of Association for Science and Technology in Xi'an, China under Grant 095920221371; and in part by the Fundamental Research Funds for the Central Universities under Grant 31020180QD094.

ABSTRACT Mimicking the biological visual attention mechanism to discriminate visually appealing regions in natural scenes has been a hot research topic in recent years. However, designing computational models with self-driven capability for open world scenarios remains a challenging task which deserves to be further studied. In this paper, we propose an unsupervised learning approach to detect salient objects from images by fully exploiting the multi-context semantic information of the scenes. Specifically, a self-driven model combing the idea of discriminative metric learning and structured sparse constraint is designed to find an optimal semantic mapping space for robust scene specific saliency prediction from complex environments. Meanwhile, a heuristic alternating optimization algorithm is developed to remove the ambiguity in the coarse geometric prior to generate a fine-grained discriminative model for saliency. On the basis of this, multi-context visual scenes are jointly modeled and fused to capture the image hierarchical structures for high-quality saliency map generation. Finally, we conduct experiments to verify the effectiveness of the proposed approach on four saliency benchmark datasets and compare it with 18 state-of-the-art saliency detection methods. Both qualitative saliency map and quantitative numerical index results indicate that our method has superior detection performance than the other counterparts under diversified scenes. Also, the proposed approach is applied to model wide synthetic aperture radar images for rapid target detection and promising results are obtained.

INDEX TERMS Heuristic alternating optimization, salient object detection (SOD), SAR target detection, unsupervised learning.

I. INTRODUCTION

Visual attention is an intelligent behavior of primates, which can help to handle massive visual streams under limited brain processing and storage capacity. It is based on the rationality that reducing the redundancy of visual scenes will not influence our understanding and may improve the visual perception efficiency. For decades, researchers from cognitive science, neurobiology and computer science have devoted to explore its underlying mechanism and many theories are proposed for this goal [1], [2], [3]. Despite the great efforts made, there is still a lot more to discover on how this intelligent behavior comes into being. It is generally believed there is a saliency map in the visual pathway, which directs our attention to the most conspicuous region in the scene. Meanwhile, visual attention works both in a top-down task driven and bottom-up scene stimulated manner. Ever since the pioneer work of Itti et al. [4], there has been an increasing interest in predicting this saliency map with computer algorithms [5], [6], [7]. Also, the advancement of saliency modeling technique has benefited a wide range of scientific and engineering fields, such as industrial defect

The associate editor coordinating the review of this manuscript and approving it for publication was Mehul S. Raval.

detection [8], remote sensing interpretation [9], multi-media applications [10], etc.

Till now, researchers from computer science have resorted to different ways to mimic the visual attention mechanism. The study on visual saliency modeling can be classified into two major categories, i.e., eye fixation prediction [11] and salient object detection [12]. The former is based on fixation points acquired by eye-tracking devices for gaze prediction, while the latter is based on object annotations for regions of interest detection. They provide different views to study the modeling of this visual cognition process with specific ground-truth data. In this paper, we will mainly focus on developing robust learning algorithm to detect salient objects from complex scenes. Extensive works can be found in the literature to address this problem [13], [14], [15], but the performance in open world scenarios still needs further improvement. It is usually difficult to learn a universal model from limited prior knowledge suitable for diversified situations. Therefore, it is a meaningful research topic to develop scene adaptive saliency models with better generalization capability.

Prior knowledge plays a vital role in saliency modeling by providing inspiration sources to tackle this highly ill-posed problem. A comprehensive literature review shows most existing saliency methods get their modeling inspirations either from cognitive rules or supervision data. Cognitive rules provide biological basis for researchers to follow for saliency model design. Also, supervision data can be used to learn saliency prediction model for new scenes. Prior knowledge adds momentum in formulating computationally plausible models and meanwhile constrains the model adaptive capacity in varying environment. However, as direct and reliable modeling basis, the multi-context visual information of the scene under evaluation has not yet been fully studied for scene driven saliency learning (Fig. 1).

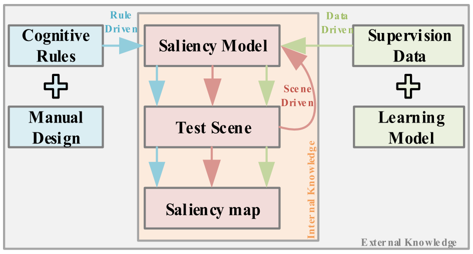


FIGURE 1. Schematic diagram of different ways of saliency modeling process.

Cognitive rules are primary source of inspirations for early saliency detection models. Inspired by neural structure in superior colliculus, a central-surround contrast operator is proposed to mimic the biological vision for saliency generation [4]. A psychological model is designed in [16] to exploit global distance and color information of the scene for saliency estimation. Also, the photographic preferences and scene layout are used to provide location and structure priors for salient object detection [17], [18], [19]. Later on, supervision data is more widely used to build learning based saliency models. A saliency optimization model is proposed by Zhu et al. to combine existing saliency measures to obtain improved detection performance [20]. From the regression perspective, a random forest model is designed by Jiang et al. to learn from multi-view features to saliency scores [21]. Meanwhile, the strong learning capability of deep neural network is explored to build end-to-end models for saliency prediction [22], [23], [24]. A comprehensive review on recent advances in CNN-based encoder-decoder networks for salient object detection is made in [25], which provides both empirical study and method investigation in this direction.

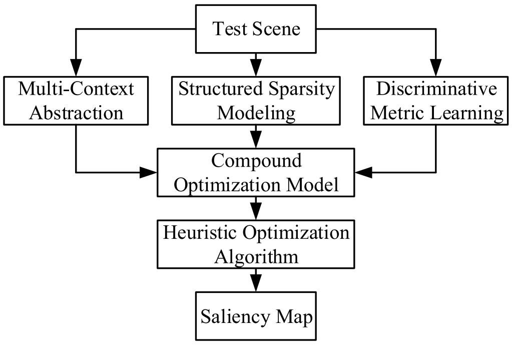

Test Scene
Multi-Context Abstraction
Structured Sparsity Modeling
Discriminative Metric Learning
Compound Optimization Model
Heuristic Optimization Algorithm
Saliency Map

FIGURE 2. Implementation flowchart of the proposed salient object detection method.

Both cognitive rules and supervision data provide external knowledge for saliency detection model design. However, external knowledge usually leads to universal rule or data driven models that are not scene specific and adaptive. To gain better generalization capability, multi-context visual information of the observed scene needs to be fully utilized for scene driven saliency modeling. Inspired by this, in this paper we propose an unsupervised model to learn discriminative metric space from multi-context visual scene for joint salient object detection. The implementation flowchart of the proposed method is shown in Fig. 2. Different from rule or data driven methods, our model is learned directly from the observed scene and does not rely on artificial design and outside scenes. Specifically, a compound optimization model based on structured constraints is designed for simultaneous saliency label disambiguation and metric learning. The semantic correlation among visual patterns is modeled to learn discriminative saliency detector from weakly labelled data. Meanwhile, a heuristic alternating optimization algorithm is proposed to iteratively search for the optimal solution of the non-convex problem. Through this scene driven unsupervised learning process, high-quality saliency maps can be generated for diversified scenes. The primary motivation of our work is to develop a scene driven unsupervised salient object detection method applicable to complex open world scenarios.

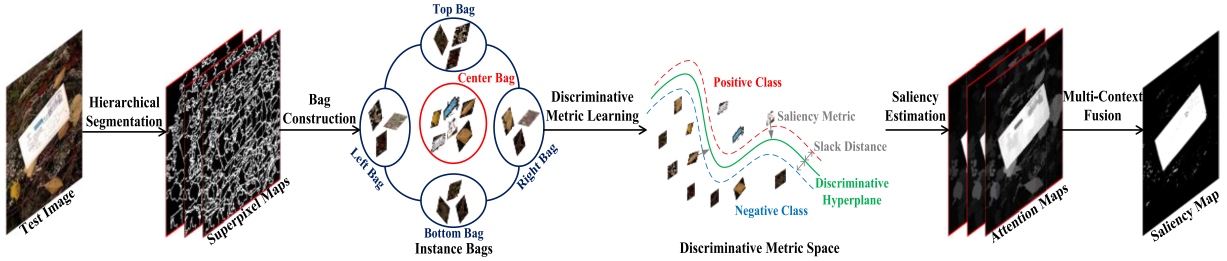


FIGURE 3. **General pipeline of the proposed salient object detection method.**

The main contributions of this paper are summarized as follows. First, we propose a novel scene driven unsupervised learning paradigm to combine multi-context visual semantics for adaptive saliency detection from diversified scenes. Secondly, a structured sparse constraint based joint optimization model is established to learn discriminative metric space for saliency estimation. Thirdly, we design a heuristic alternating optimization algorithm to efficiently search for the optimal saliency metric space from the above model. In general, we contribute a novel unsupervised saliency learning framework to fully exploit the test scene information for robust salient object detection from diversified open world scenes. The rest of this paper is organized as follows. In Section II, we will introduce the proposed unsupervised discriminative metric learning model and heuristic alternating optimization algorithm in detail. Experimental results and performance evaluation on benchmark datasets as well as model application study will be given in Section III. Finally, we will briefly conclude this paper in Section IV.

II. PROPOSED METHOD

Our saliency detection method is mainly inspired by the following observations. First, existing methods are mostly driven by external rule or data, which ignore the use of internal scene information in model design. To achieve more adaptive modeling capability, a scene driven unsupervised discriminative metric learning model is proposed for saliency detection. Secondly, to avoid the modeling bias caused by insufficient visual semantic correlation, a structured sparse constraint based multi-context fusion scheme is developed for thorough scene analysis. General pipeline of the proposed salient object detection method is illustrated in Fig. 3. As can be seen, the proposed method is mainly composed of three components, i.e., hierarchical instance bag (HIB) construction, discriminative metric learning (DML) model deduction, and multi-context fusion (MCF) scheme design. In the following subsections, we will give detailed descriptions of these components.

A. UNSUPERVISED DISCRIMINATIVE METRIC LEARNING (DML) MODEL

Deriving semantic perception results directly from the observed visual scenes conforms to the basic biological cognition rules. Existing methods mostly depend on external rule or data for model design, which however are not customized for each individual scene and thus not well adapted to changing environment. Concerning the observed internal scene, its semantic information is seldomly studied for unsupervised saliency modeling. In this paper, we propose a discriminative metric learning method to build closer connection between the observed scene and its specific saliency model.

Given a test color image $I \in \Re^{m\times n\times 3}$, we first over-segment it into several uniform superpixel regions $\{s_1, s_2, \cdots, s_k\}$ using LSC algorithm [26]. Since color space transform can provide feature representation closer to human visual system (HVS), the RGB, Lab and HSV color spaces are connected in cascade manner for superpixel description. The average feature values of the pixels inside each superpixel are assigned as the superpixel feature values. In this way, we can build a superpixel feature matrix $Fs = [fs_1, fs_2, \cdots, fs_k] \in \Re^{8\times k}$ with each column being an 8-dimensional feature vector spanned by R, G, B, L, a, b, H and S color channels.

The basic idea is to learn a discriminative metric model from structured visual scenes for accurate saliency prediction of superpixel regions. This problem can be mathematically expressed as the following form

$$ss_i \propto \Phi_I(fs_i), \tag{1}$$

where, $fs_i \in \Re^{8\times 1}$ is the feature vector of the i-th superpixel region and its corresponding saliency value is denoted by ss_i. Φ_I is the scene driven saliency model learned from the test image I. Previous methods whose saliency models are independent of the test image, mostly adopt cognitive rules or supervision data as their modeling driven force. They usually result in fixed detection models with limited adaptive capacity for changing scenarios. To promote the model adaptability, a joint discriminative metric learning framework is proposed in this paper, which has the following expression form

$$\begin{aligned}(w^*, b^*) = \arg\max_{w,b} &\frac{1}{||w||_2}\\ \text{s.t.}\quad &\underbrace{ls_i}_{\substack{\text{unknown}\\ \text{label}}} \times (w \cdot \underbrace{fs_i}_{\substack{\text{known}\\ \text{feature}}} + b) \geq 1,\\ &i = 1, 2, \cdots, k\end{aligned} \tag{2}$$

where, $w \in \Re^{8\times 1}$ and b are the normal vector and intercept of the discriminative metric hyperplane, and $ls_i \in \{-1, +1\}$ is the saliency label of the i-th superpixel region which is not known in advance. After the optimal hyperplane variables w^* and b^* are learned from the test scene, we can predict the saliency label ls_i and saliency value ss_i of the i-th superpixel region as follows

$$\begin{cases} ls_i = \text{sign}(w^* \cdot fs_i + b^*) \\ ss_i = \dfrac{w^* \cdot fs_i + b^*}{||w^*||_2}, \end{cases} \quad i = 1, 2, \cdots, k \tag{3}$$

Since the saliency label information is unknown, the above optimization problem is intractable in practice. To guarantee unique solution to best favor saliency modeling, we enforce structured sparse constraints to the superpixel saliency label vector $Ls = [ls_1, ls_2, \cdots, ls_k] \in \{-1, +1\}^{1\times k}$. We divide all the superpixels into different groups according to their spatial positions in the image. Superpixel regions that are in touch with the image boundaries are respectively grouped into four instance sets, which we refer to as top bag B_t, bottom bag B_b, left bag B_l and right bag B_r. For those that have no touch with the image boundaries, they are grouped into one instance set named as center bag B_c. Each instance bag carries specific visual semantic elements that are closely related to the image geometry. For salient object detection, saliency labels of the instance bags show structured sparse characteristic, which is explored for adaptive model design.

For description convenience, we denote the saliency label vectors of the instance bags by Ls_t, Ls_b, Ls_l, Ls_r and Ls_c respectively. Since superpixel instances in the four bags B_t, B_b, B_l and B_r are mostly non-salient, saliency label vectors of these bags Ls_t, Ls_b, Ls_l and Ls_r should be expected to be sparse enough. Also, the center bag B_c covers nearly all the visual elements of salient objects and therefore its saliency label vector Ls_c should contain at least one positive element. In this paper, the above structured sparse constraints on the saliency label vectors are embedded into the learning framework in (3) to form a computationally feasible model as follows.

$$\begin{aligned} (w^*, b^*) &= \arg\max_{w,b} \frac{1}{||w||_2} - \lambda \sum_{i\in\{t,b,l,r\}} ||Ls_i + \mathbf{1}||_0 \\ \text{s.t.} \quad & \underbrace{ls_i}_{\substack{\text{unknown}\\ \text{label}}} \times (w \cdot \underbrace{fs_i}_{\substack{\text{known}\\ \text{feature}}} + b) \geq 1, \\ & i = 1, 2, \cdots, k \quad ||Ls_c + \mathbf{1}||_0 \geq 1 \end{aligned} \tag{4}$$

where, λ is the weighting coefficient used to balance between the two optimization objectives and $\mathbf{1}$ is an all one vector used to convert from label domain to sparse domain. The second term in the objective function seeks to minimize the zero norm of converted saliency label vectors of the four boundary bags, and the second constraint condition enforces the zero norm of the converted saliency label vector of the center bag to be greater than or equal to 1. The rationality behind the use of the two zero norms comes from the intrinsic distribution property of salient objects in the image. In practice, the four boundary bags seldomly cover salient objects and thus their saliency label vectors should be sparse enough. Meanwhile, the center bag basically covers all the salient objects and thus the sparsity of its saliency label vector should at least be 1. The two zero norms are specifically designed to describe the above structured sparse characteristic and they together provide loose but universal guidance for the optimization. By introducing the two zero norms, we aim to add slack label learning momentum to the optimization model for adaptive saliency detection.

Since the saliency label vector $ls_i, i = 1, 2, \cdots, k$ is an intermediate variable, it is further expressed by the two optimization variables w and b via sign function. Correspondingly, we introduce a new equality constraint into the model as follows.

$$\begin{aligned} (w^*, b^*) = \arg\max_{w,b} & \frac{1}{||w||_2} - \lambda \sum_{i\in\{t,b,l,r\}} ||Ls_i + \mathbf{1}||_0 \\ \text{s.t.} \quad & ls_i \times (w \cdot fs_i + b) \geq 1, \\ & i = 1, 2, \cdots, k \\ & ||Ls_c + \mathbf{1}||_0 \geq 1 \\ & ls_i = \text{sign}(w \cdot fs_i + b), \\ & i = 1, 2, \cdots, k \end{aligned} \tag{5}$$

By reorganizing the above model, we can obtain the standard constrained optimization problem in the following.

$$\begin{aligned} (w^*, b^*) = \arg\min_{w,b} & \frac{1}{2}||w||_2^2 + \lambda \sum_{i\in\{t,b,l,r\}} ||Ls_i + \mathbf{1}||_0 \\ \text{s.t.} \quad & \mathbf{1} - Ls \odot (w^T \times Fs + b \times \mathbf{1}) \leq \mathbf{0} \\ & 1 - ||Ls_c + \mathbf{1}||_0 \leq 0 \\ & Ls - \text{sign}(w^T \times Fs + b \times \mathbf{1}) = \mathbf{0} \end{aligned} \tag{6}$$

where, $\odot$ is element-wise Hadamard product and $\mathbf{0}$ is an all zero vector. As can be seen, the major merit of this model is the design of a discriminative metric learning paradigm that matches perfectly with the practical problem. The sparse patterns inside the test image are thoroughly explored for scene driven unsupervised salient object detection. Different from rule or data driven models, our saliency model is learned directly from the test image and can keep dynamic adjusting to the diversified scenes. Therefore, it possesses more reliability and flexibility in modeling saliency from open world scenarios. From a mathematical perspective, this is a discrete multi-objective constrained optimization problem with hidden variables, which cannot be effectively solved with traditional numerical methods. In this paper, we propose an efficient heuristic alternating optimization algorithm to find the optimal solution of the above problem for reliable saliency estimation. Detailed descriptions of the proposed optimization algorithm will be given in the following subsection.

B. EFFICIENT HEURISTIC ALTERNATING OPTIMIZATION (HAO) ALGORITHM

Given the superpixel feature matrix Fs, we aim to learn the optimal hyperplane variables w and b with the above constrained optimization model for scene driven saliency detection. The superpixel saliency label vector Ls is a hidden variable that depends on the optimization variables w and b. Its sub-vectors Ls_t, Ls_b, Ls_l, Ls_r and Ls_c together apply zero norm based structured sparse constraints to the model. As a unique design of our model, different subsets of the hidden variable are respectively introduced into the objective function and constraint conditions, making it a relatively complex optimization problem in practice. In this paper, we propose a heuristic alternating optimization algorithm to find the optimal solution of the above problem in an efficient manner.

We first adopt idea from the sequential unconstrained minimization technique (SUMT) to convert the standard constrained optimization problem into equivalent unconstrained form. A penalty function $P(w, b)$ based on exterior point method (EPM) is accordingly designed as follows.

$$\begin{aligned} P(w, b) = {} & || \max[\mathbf{0}, \mathbf{1} - Ls \odot (w^T \times Fs + b \times \mathbf{1})]||_2^2 \\ & + [\max(0, 1 - ||Ls_c + \mathbf{1}||_0)]^2 \\ & + ||Ls - \mathrm{sign}(w^T \times Fs + b \times \mathbf{1})||_2^2 \end{aligned} \tag{7}$$

where, $\max(x, y)$ is a function that takes the maximum of x and y as its output. The above function can assign positive penalty values to infeasible solutions, while applies no penalty to any feasible solution. By adding the penalty function to the primal objective function, we can obtain the following augmented objective function $F(w, b, \eta)$.

$$\begin{aligned} & F(w, b, \eta) \\ & \quad = f(w, b) + \eta P(w, b) \\ & \quad = \frac{1}{2}||w||_2^2 + \lambda \sum_{i \in \{t,b,l,r\}} ||Ls_i + \mathbf{1}||_0 \\ & \qquad + \eta \Big\{ || \max[\mathbf{0}, \mathbf{1} - Ls \odot (w^T \times Fs + b \times \mathbf{1})]||_2^2 \\ & \qquad + [\max(0, 1 - ||Ls_c + \mathbf{1}||_0)]^2 \\ & \qquad + \ ||Ls - \mathrm{sign}(w^T \times Fs + b \times \mathbf{1})||_2^2 \Big\} \end{aligned} \tag{8}$$

where, $f(w, b)$ is the primal objective function and η is the penalty factor that increases gradually along with iterations. Till now, we derive the unconstrained optimization objective function, which can facilitate the succeeding optimization algorithm design process. However, the augmented objective function is relatively complex in its mathematic expression form, which involves the dynamic nonlinear interactions among the scene features and model variables. It is intractable to find the optimal solution of this unconstrained optimization problem simply with traditional methods. Here, we propose a heuristic alternating optimization algorithm to combine the idea of exterior point method and swarm intelligence theory for efficient optimal solution search.

Following the exterior point method, we gradually increase the penalty factor along with the iterations to enforce the solution to move towards feasible region. Accordingly, the nonlinear augmented objective function will change dynamically during the iterations. As a representative swarm intelligence algorithm, particle swarm optimization (PSO) is proved to be quite suitable for parallel solution search in dynamic environment [27]. Therefore, it is embedded into the overall iteration process to search for the optimal solution in a successive manner. Discriminative metric hyperplane variables w and b are concatenated and encoded as the swarm particle position. A group of particles work collaboratively to search for the position with best fitness function value. Meanwhile, the obtained optimal solution from the previous step is used as initial solution of the next step, making the search process to converge quickly. This heuristic search process will terminate until the final stop criterion is reached. Pseudo code of the proposed heuristic alternating optimization algorithm is summarized in Algorithm 1.

Algorithm 1 Pseudo Code of the Heuristic Alternating Optimization Algorithm

Input: Superpixel feature matrix $Fs \in \Re^{8\times k}$, weighting coefficient λ, swarm particle number N, inertia factor ω, acceleration coefficients c_1 and c_2, termination error ε.

Output: Optimal discriminative metric hyperplane variables $w^* \in \Re^{8\times 1}$ and b^*.

Initialization: Penalty factor η, swarm particle velocity $V \in \Re^{9\times N}$, swarm particle position $P \in \Re^{9\times N}$, particle personal best $P_{best} \in \Re^{9\times N}$, particle global best $G_{best} \in \Re^{9\times 1}$, fitness vector $Fit \in \Re^{N\times 1}$, and number of iterations $t = 1$.

```
do
    for $i = 1: 1: N$
        Calculate fitness value $F(P(1:8, i), P(9, i), \eta)$ with (8).
        if ( $F(P(1:8, i), P(9, i), \eta) < Fit(i)$) % update $P_{best}$
            $P_{best}(:, i) = P(:, i); Fit(i) = F(P(1:8, i), P(9, i), \eta)$
        end
    end
    $Temp = G_{best}$ %store current $G_{best}$ in temporary variable
    Find $i^* = \min_{i \in [1,N]} Fit(i)$ and set $G_{best} = P_{best}(:, i^*)$. %update $G_{best}$
    for $i = 1: 1: N$ %update particle velocity and position
        $V(:, i) = \omega \times V(:, i) + c_1 \times \mathrm{rand}(1) \times [P_{best}(:, i) - P(:, i)]$
            $+ c_2 \times \mathrm{rand}(1) \times [G_{best} - P(:, i)]$
        $P(:, i) = P(:, i) + V(:, i)$
    end
    $t = t + 1; \eta = \sqrt{t} \times \eta$ % increase penalty factor
while ($||G_{best} - Temp||_2 \geq \varepsilon$)
```

As can be seen, the proposed scene driven unsupervised learning approach models saliency detection as a constrained multi-objective optimization problem. Its heterogeneous mathematical expression form provides a complete and adaptive framework for robust saliency modeling from diversified visual scenes. However, the existence of the complex zero norm terms of the hidden variables makes this optimization problem quite challenging in practice. It is intractable to simply use traditional optimization techniques to find the optimal solution in acceptable time. Since the proposed heuristic alternating optimization algorithm has no restriction on the objective function and constraint conditions, it is capable of handling this complex nonlinear optimization problem with great ease. The iterative approximation strategy together with the velocity-displacement model can guarantee a fast global convergence of the optimization algorithm. Therefore, we can expect to obtain satisfactory discriminative metric hyperplane variables for efficient saliency prediction according to (3).

Till now, we accomplish the design of scene driven unsupervised saliency learning model under single semantic context. As is the case, the semantic perception result of a visual element is greatly influenced by its surrounding hierarchical scene layouts. In the following, we will extend our model to fuse multi-context modeling results for more comprehensive saliency detection. Details of the proposed joint multi-context saliency fusion scheme will be discussed in the next subsection.

C. JOINT MULTI-CONTEXT FUSION (MCF) SCHEME AND IMPLEMENTATION DETAILS

It is widely acknowledged that saliency is the competitive outcome of visual elements under multiple observation scales. Inspired by this, we design a saliency fusion scheme to combine multi-context modeling results for high-quality saliency map generation. Using LSC algorithm, we over-segment the test image from multiple semantic levels to capture the hierarchical scene structures. Saliency learning results from multiple visual contexts are fused for fine-grained saliency estimation. Specifically, the final saliency map $S \in \Re^{m \times n}$ of the test image is estimated as follows.

$$S_{r,c} = \frac{\sum_{j=1}^{l}\sum_{i=1}^{k_j} ss_i^j \times \exp\left[-\frac{\|fp_{r,c}-fs_i^j\|_2^2}{2\sigma^2}\right] \times \Theta\left(p_{r,c} \in s_i^j\right)}{\sum_{j=1}^{l}\sum_{i=1}^{k_j} \exp\left[-\frac{\|fp_{r,c}-fs_i^j\|_2^2}{2\sigma^2}\right] \times \Theta\left(p_{r,c} \in s_i^j\right)} \tag{9}$$

where, $S_{r,c}$ is the saliency value in the r-th row and c-th column. $p_{r,c}$ is the pixel in the r-th row and c-th column and $fp_{r,c} \in \Re^{8\times 1}$ is its feature vector. $s_i^j$is the i-th superpixel region in the j-th semantic level, and its feature vector and saliency value are respectively denoted by $fs_i^j \in \Re^{8\times 1}$ and ss_i^j. The number of semantic levels is written as l and k_j represents the number of superpixel regions in the j-th semantic level. Also, exp[·] is the exponential function where σ acts as a smoothing parameter, and $\Theta(\cdot)$ is an indicator function which outputs 1 if the condition in the brackets is satisfied, otherwise 0.

The hierarchical scene structures require that its saliency modeling be better discussed in multiple semantic contexts. By performing multi-layer abstraction to the test image, we can obtain different levels of visual semantic representation for joint saliency learning. For each semantic level, we use the proposed discriminative metric learning model to produce its attention map. The multi-context saliency modeling results are combined through the designed saliency fusion scheme for pixel-wise accurate saliency map estimation. Specifically, the saliency of a spatial position is evaluated under multiple visual semantic contexts and is estimated based on its feature similarity to the hierarchical superpixel regions it resides in. Therefore, we can expect to get high-quality saliency maps with more reliability and higher accuracy as will be seen later.

In the experiments, the weighting coefficient λ is fixed to be 0.1 and the initial value of the penalty factor η is chosen to be 1. Also, the swarm particle number N, inertia factor ω, acceleration coefficients c_1 and c_2, and termination error ε in PSO are set to 36, 1, 2, 2 and 0.001 respectively. Note that the parameter tuning of PSO is performed based on empirical guidance as well as problem characteristics so as to balance between search efficiency and solution quality. Concerning the initialization of the swarm particle states, we adopt random function to generate legal vectors for swarm particle velocity V and position P, as well as particle personal best P_{best} and global best G_{best}. Meanwhile, to make the iteration process start properly, all the elements in the fitness vector Fit are initialized to be positive infinity. Finally, the number of semantic levels l is designed to be 3 and correspondingly the number of superpixel regions k_1, k_2 and k_3 are respectively set to 100, 200 and 400. Besides, to get satisfactory saliency fusion effect, we set the smoothing parameter σ to 10. In practice, we implement the proposed method with MATLAB and the code is run on a HP Z8 G4 workstation with 8 core CPU of 1.70 GHz, 64 GB RAM, and 64 bits Windows 10 operating system. It is also worth mentioning that a specific group of optimal hyperplane variables will be acquired for each test image during the scene driven unsupervised learning process.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. EXPERIMENTAL SETUP

In this section, the proposed saliency model (referred to as DML) is tested on four classical saliency benchmark datasets (PASCAL-S [28], ECSSD [29], DUT-OMRON [30] and THUS-10000 [16]) along with 18 state-of-the-art saliency detection methods, which are respectively denoted as BMS [31], CA [32], CB [33], CGVS [34], DSR [35], FCB [36], GMR [30], GR [37], HS [29], LMLC [38], LPS [39], MBS+ [40], MC [41], MNP [42], RC [16], SF [43], ST [44] and WF [45]. For fair performance evaluation, we use the source codes released by the original authors to produce the corresponding saliency maps during the experiments. Detailed parameters involved in the 18 saliency detection methods can be found in their corresponding source codes and we use the default parameter settings provided by the original authors for saliency map generation. In terms of the evaluation metrics, both qualitative and quantitative results are obtained for comprehensive detection performance analysis.

As an intuitive way to show the modeling effect, saliency map is used to provide qualitative results for detection performance evaluation by comparing its closeness with the corresponding ground-truth map. In general, the closer the saliency map is to the ground-truth map, the better the saliency detection performance will be. Saliency map is a subjective evaluation standard and quantitative numerical indexes are also needed for objective performance comparison. In this paper, precision-recall (PR) curve and F-measure curve are both used for this purpose. Precision and recall are two complementary metrics, which are respectively defined to measure the accuracy and completeness of the saliency detection results. Precision is the ratio between the correctly detected and actually detected foreground regions, and recall is the ratio between the correctly detected and manually labelled foreground regions. To provide a unified evaluation standard, F-measure is proposed as a composite index, which is the harmonic mean of precision and recall.

$$F_\beta = \frac{(1+\beta^2) \times \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}} \tag{10}$$

where, $\beta^2 = 0.3$ is used to give more emphasis on precision than recall as suggested in [46]. For an 8-bits grayscale saliency map, we use a threshold ranging from 0 to 255 to binarize it. By comparing the successively segmented binary maps with the corresponding ground-truth map, we can derive 256 groups of precision, recall and F_β values. After plotting the 256 precision and recall point pairs on a 2-D plane, we can obtain the PR curve of the saliency map for performance analysis. Similarly, we can draw the F-measure curve of the saliency map using the 256 calculated F_β values.

It is worth mentioning that the proposed heuristic alternating optimization algorithm has some randomness due to the initialization and update of the swarm particle states. To give fair performance evaluation of our method, we run our code on each test image for ten independent times, and all the results are averaged to get the final saliency map. Through this, we wish to avoid the performance evaluation bias caused by the possible instability of our detection results. In what follows, we will give the saliency map results as well as numerical evaluation indexes of the saliency methods on the benchmark datasets for comprehensive detection performance analysis and comparison.

B. PERFORMANCE EVALUATION RESULTS

The self-driven learning capability in open world scenarios is a significant difference for various saliency models. It is therefore necessary to verify the model performance in diversified complex scenes. The four saliency benchmark datasets used in this paper cover rich visual scenes with challenging situations and thus are well suited for this requirement. PASCAL-S dataset contains 850 images and their carefully labelled ground-truth maps with multi-level saliency annotations. Also, ECSSD dataset is composed of 1,000 low contrast images and their binary fore/back-ground masks. The 5,168 images in DUT-OMRON dataset cover wide range of natural scenes with complex spatial layouts. Also called as MSRA10K, THUS-10000 is a large salient object detection dataset (contains 10,000 images) with pixel-level labeling. Below we will show the testing results of the saliency methods on the above datasets for comprehensive modeling performance evaluation.

Fig. 4 shows some typical saliency maps of the top-ten performing methods (CGVS, DSR, GMR, HS, MBS+, MC, RC, ST, WF and DML) on the four benchmark datasets. As can be observed from the results, our method is able to produce high-quality saliency maps better than that of the other counterparts, when in face of multiple, variable-size, occluded and cropped objects as well as low image contrast and structural details. The saliency generation mechanism in open world scenarios is such complex that models based on specific rules or limited data will lack robustness in practice. For example, the sheep in the second image are not entirely found by most methods for their over-simplified assumptions like the focusness prior. Also, the camouflaged submarine in the fifth image is only partially detected from the ocean by some methods due to the low fore and back-ground contrast. Meanwhile, the cropped motorcyclist in the seventh image is not integrally highlighted in some saliency maps for their over-reliance on boundary prior. Free from external rules or data, our model can learn adaptive discriminative metric space from diversified test scenes for unsupervised salient object detection. The semantic separable condition and structured sparse property are tightly coupled together for concise and complete learning model and algorithm design. Our method provides a more general form for representing and integrating the heterogeneous saliency information and thus is more suitable for the problem.

Meanwhile, the average PR and F-measure curves of all the saliency methods on the four benchmark datasets are respectively drawn in Figs. 5 and 6 for objective performance evaluation. As can be seen from Fig. 5, the PR curves of DML keep lying in the most upper right corner of the plots, indicating the superiority of our saliency detection method. Up to now, the saliency modeling performance is becoming saturated on relatively simple datasets, but still needs great improvement in complex scene datasets. As a result of this, the modeling capability in open world scenarios is a major basis for detection performance evaluation. It can be observed from the PR curves that the advantage of our method over the other counterparts on the simple THUS-10000 dataset is not too obvious. But when it comes to the more complex PASCAL-S and ECSSD datasets, this performance advantage gets further enlarged. In terms of the challenging DUT-OMRON dataset, our method outperforms the other counterparts to a large extent owing to its stronger modeling capability in open world scenarios. Also, similar

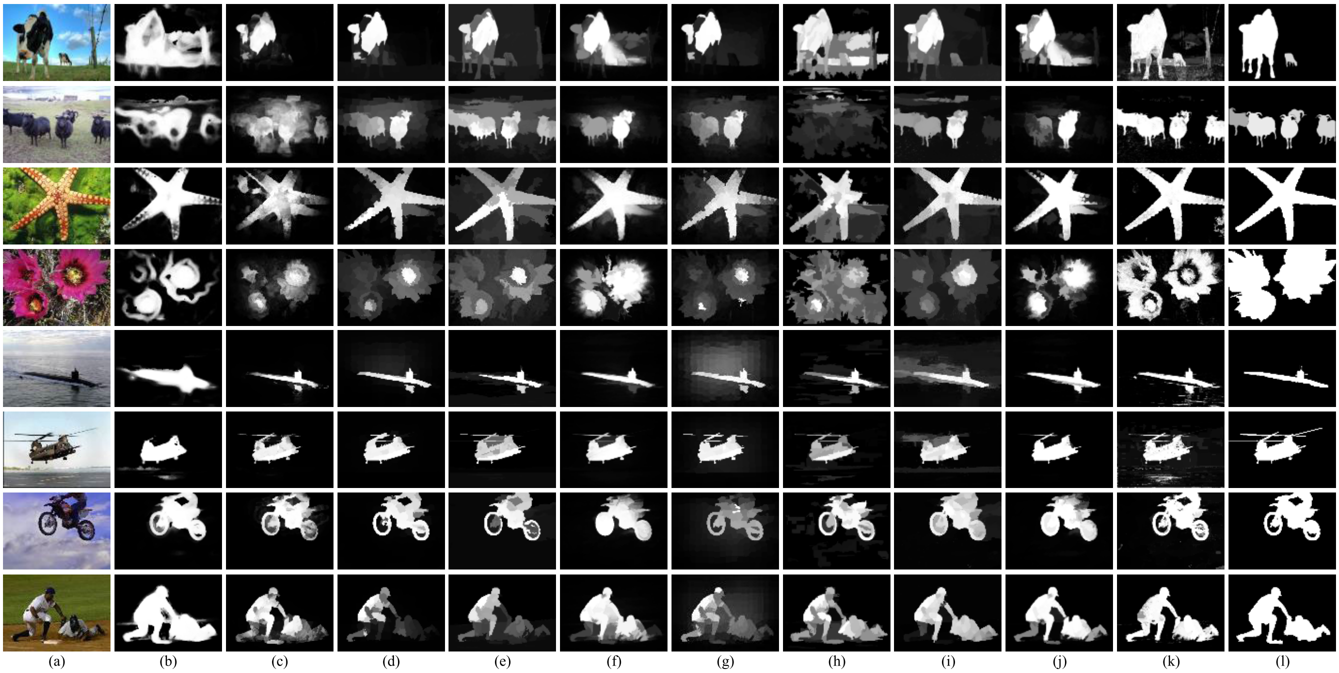
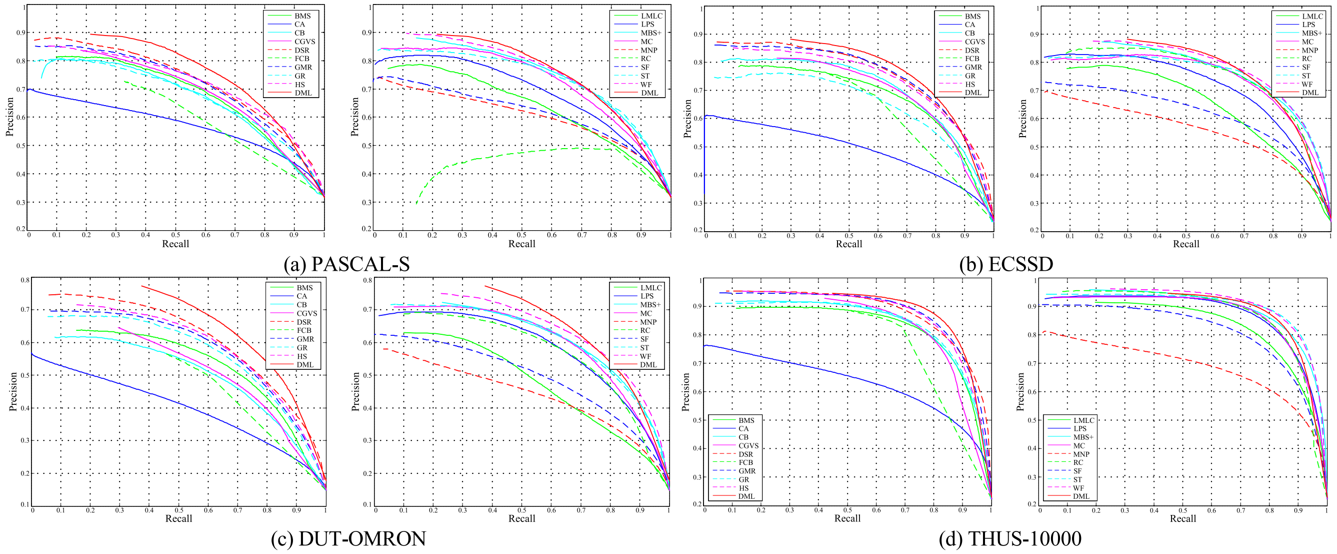


FIGURE 4. Some saliency maps produced by the top-ten performing methods on the four benchmark datasets. Each column from (a) to (l) is respectively the input images, saliency maps of CGVS, DSR, GMR, HS, MBS+, MC, RC, ST, WF and DML, and the ground truths. Every two rows from top to bottom correspond to sample images from PASCAL-S, ECSSD, DUT-OMRON and THUS-10000 dataset.



(a) PASCAL-S

(b) ECSSD

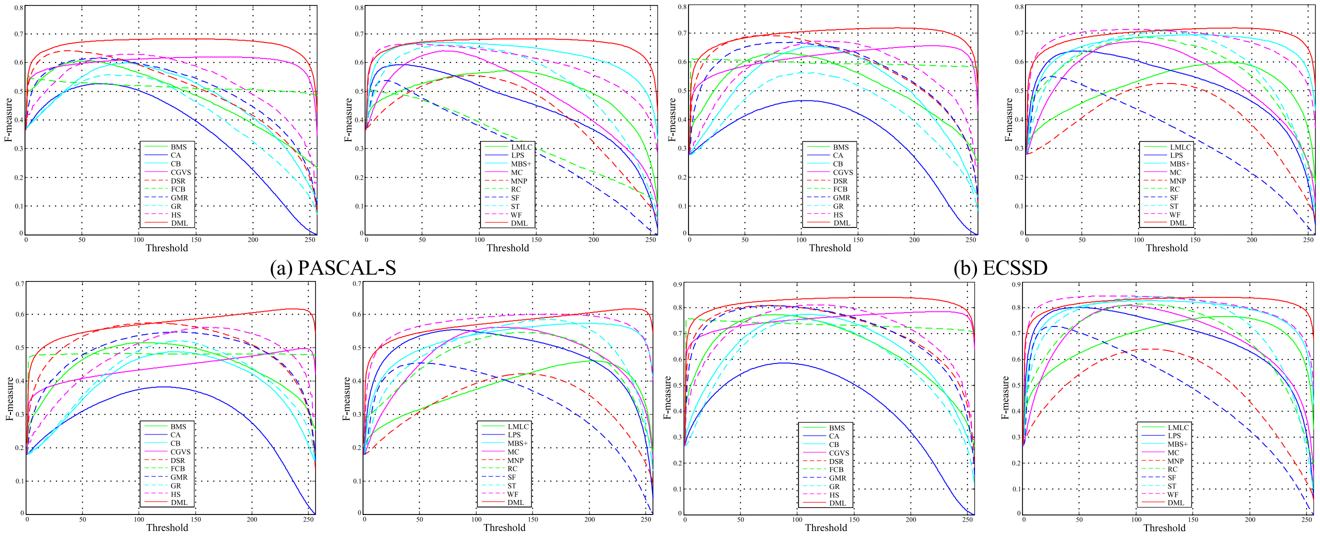
(c) DUT-OMRON

(d) THUS-10000

FIGURE 5. Average PR curves of all the saliency methods on the four benchmark datasets. For sake of observation convenience, PR curves of the 18 comparison methods are shown in two separate subplots and the PR curve of our method is shown in both subplots.

results can be observed from the F-measure curves in Fig. 6. Since our method works in a scene driven mode, it can learn a specific model for each test image by fully exploring the visual semantic information inside. Compared with external rule or data driven methods, it can provide a more general paradigm for adaptive saliency modeling in dynamic environment. Therefore, our method is better at capturing the complex saliency generation mechanism from diversified natural scenes.

To validate the performance stability of our method across different datasets, we run Friedman test on the four benchmark datasets with precision, recall and F-measure as the evaluation indexes. The obtained test statistic for our method is 5.8, which is smaller than the critical test value 7.4 looked up under significance level 0.05. This means that there is no obvious performance difference of our method on the four benchmark datasets, which confirms the robustness of our method across different types of datasets. Finally,



(a) PASCAL-S

(b) ECSSD

(c) DUT-OMRON

(d) THUS-10000

FIGURE 6. Average F-measure curves of all the saliency methods on the four benchmark datasets. For sake of observation convenience, F-measure curves of the 18 comparison methods are shown in two separate subplots and the F-measure curve of our method is shown in both subplots.

concerning the modeling efficiency, the proposed heuristic alternating optimization algorithm can converge to a stable solution after a few rounds of iterations. The time complexity of each individual iteration process is $O(2N + N\log_2 N)$, which mainly depends on the swarm particle number N and meanwhile shows approximately linear relationship with it. Also, it is worth noting that this dynamic search process can be further parallelized to greatly reduce the computational complexity. In this regard, our model has the potential to facilitate real-time saliency related applications in practice.

For some edge computing devices with less computing resources, our method still has potentials to guarantee the running efficiency on them. First of all, the search range of the optimization problem lies in the low-dimensional space spanned by the hyperplane variables, and thus the search complexity is relatively low. Secondly, since the best solution keeps moving towards better direction along with iterations, the stop criterion can be alternatively designed as the maximum iteration number so as to balance between the running efficiency and quality of solution. Finally, some algorithm hyperparameters, such as the particle number, termination error, and semantic level, can be further optimized and tailored to facilitate efficient application in edge computing devices. Since no extra training is needed, our model is especially suitable for the problems with limited or no supervision data available. Below, we will apply our saliency model to synthetic aperture radar (SAR) images for fast target detection from wide scenes.

C. APPLICATION TO SAR TARGET DETECTION

With the advancement of SAR imaging techniques, high-resolution SAR images are collected by carrying platforms for detailed earth observation. Different from natural images, SAR images generally have wide ground coverage and sparse target distribution. Traditional target detection methods follow the false alarm removal (FAR) idea to search for targets and are usually daunting and exhaustive. Thus, there is an urgent need to develop effective computer algorithms for rapid target detection from wide SAR scenes. As an intelligent perception mechanism, visual saliency can be modeled to filter out redundant scene information and direct the search towards regions of interest (ROIs). Its introduction into the SAR image analysis will hopefully improve the target detection performance. Inspired by this, we adapt our saliency model to fit with the SAR images and develop a saliency guided target detection approach. Specifically, the Pauli RGB composition mode is used to generate the color-coded SAR images from raw polarization data. SAR images covering sea and land areas are used to verify the model performance in real world applications.

Shown in Fig. 7(a) is a SAR image collected by RADARSAT-2 spaceborne imaging platform on May 2008 near Vancouver, Canada. It covers a sea area of 3.2 km by 6.4 km with 5 ships in the scene. Corner reflection from the ship can be clearly observed in this image. We perform saliency modeling to this scene and the corresponding result is shown in Fig. 7(b). As can be seen, ship regions are completely highlighted as ROIs along with their corner reflections and some sea clutters. This bottom-up modeling result is further filtered by top-down cognitive priors to produce more accurate ship detection result. Here, we use Otsu algorithm to binarize the saliency map and impose morphological processing to get the final detection result in Fig. 7(c). We can see that all the ships are correctly detected from the scene and meanwhile few false alarms exist, indicating a satisfactory ship detection performance of our model. It is worth mentioning that our saliency model

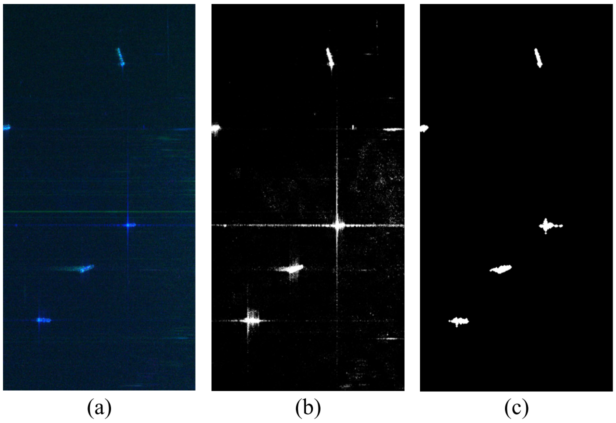


FIGURE 7. SAR ship detection results of our saliency model. (a) is the SAR image collected by RADARSAT-2, (b) is the saliency modeling result, and (c) is the detected ship targets.

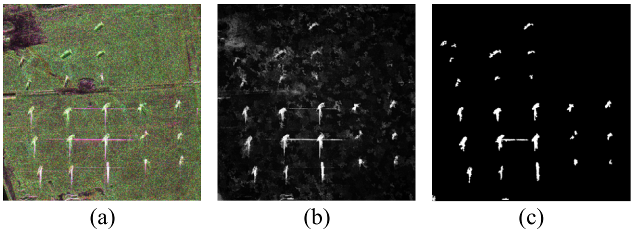


FIGURE 8. SAR tank detection results of our saliency model. (a) is the SAR image collected by X-SAR, (b) is the saliency modeling result, and (c) is the detected tank targets.

can adapt to different data forms and possesses application potentials in diversified scenarios.

Also, the SAR image in Fig. 8(a) is collected by X-SAR spaceborne imaging platform on Oct. 2014. It covers a wide range of land areas with 22 tanks in the scene. Corner reflections from tanks can also be observed in this image and the background appears more cluttered. The saliency modeling result of our method on this scene is shown in Fig. 8(b). As can be seen, despite the corner reflections and land clutters, the tank regions are more obviously highlighted in the result. Similarly, we successively impose Otsu binarization and morphological processing to the saliency map to get the final detection result in Fig. 8(c). We can see that all the tanks are correctly detected from the scene and meanwhile only a few false alarms occur in the detection result. This further confirms the robust target detection performance of our model in cluttered environment. Developing intelligent algorithms for the interpretation of SAR images is a promising research direction [47] and the saliency model in this paper is a meaningful attempt towards this goal.

IV. CONCLUSION

In this paper, we propose an unsupervised discriminative metric learning model to jointly explore the multi-context visual semantic separability and structured sparsity property for robust salient object detection from complex open world scenes. A novel discrete multi-objective constrained optimization problem with hidden variables is established for adaptive scene driven saliency modeling. Meanwhile, we develop a hybrid intelligent optimization algorithm by combining the idea of EPM and PSO for efficient optimal feasible solution search. Extensive experiments on saliency benchmark datasets demonstrate the superior performance of our method to other classical saliency detection approaches, especially in face of complex open world scenarios. Also, the proposed saliency model is applied to wide SAR image analysis for rapid target detection from remotely sensed data, and promising results are obtained in typical ground areas. Different from existing methods, the proposed model is not restricted by specific rules or external data and possesses unsupervised scene driven saliency learning capability. In summary, we provide a more general and flexible learning model for boosting the saliency detection performance in complex open world problems. In the future, it is a valuable research topic to develop customized unsupervised saliency learning models to meet the demands of diversified scene understanding.

ACKNOWLEDGMENT

The author would like to thank the editors and anonymous reviewers for their valuable comments and suggestions that help to greatly improve this paper.

REFERENCES

[1] R. Cong, J. Lei, H. Fu, M.-M. Cheng, W. Lin, and Q. Huang, "Review of visual saliency detection with comprehensive information," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 10, pp. 2941–2959, Oct. 2019.

[2] J. Li, Z. Wang, and Z. Pan, "Double structured nuclear norm-based matrix decomposition for saliency detection," *IEEE Access*, vol. 8, pp. 159816–159827, 2020.

[3] C. Yao, Y. Kong, L. Feng, B. Jin, and H. Si, "Contour-aware recurrent cross constraint network for salient object detection," *IEEE Access*, vol. 8, pp. 218739–218751, 2020.

[4] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[5] K. Yan, X. Wang, J. Kim, W. Zuo, and D. Feng, "Deep cognitive gate: Resembling human cognition for saliency detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 4776–4792, Sep. 2022.

[6] S. Huo, Y. Zhou, W. Xiang, and S. Y. Kung, "Semisupervised learning based on a novel iterative optimization model for saliency detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 225–241, Jan. 2019.

[7] S. Wang, "Learning nonlinear feature mapping via constrained non-convex optimization for unsupervised salient object detection," *IEEE Access*, vol. 10, pp. 40743–40752, 2022.

[8] Y. Qiu, L. Tang, B. Li, S. Niu, and T. Niu, "Uneven illumination surface defects inspection based on saliency detection and intrinsic image decomposition," *IEEE Access*, vol. 8, pp. 190663–190676, 2020.

[9] G. Zhang, Z. Li, X. Li, C. Yin, and Z. Shi, "A novel salient feature fusion method for ship detection in synthetic aperture radar images," *IEEE Access*, vol. 8, pp. 215904–215914, 2020.

[10] S. Wang, M. Wang, S. Yang, and K. Zhang, "Salient region detection via discriminative dictionary learning and joint Bayesian inference," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 5, pp. 1116–1129, Jan. 2018.

[11] W. Wang, J. Shen, X. Dong, A. Borji, and R. Yang, "Inferring salient objects from human fixations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 1913–1927, Aug. 2020.

[12] L. Wang, L. Wang, H. Lu, P. Zhang, and X. Ruan, "Salient object detection with recurrent fully convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1734–1746, Jul. 2019.

[13] D. Kang, S. Park, and J. Paik, ''SdBAN: Salient object detection using bilateral attention network with dice coefficient loss,'' *IEEE Access*, vol. 8, pp. 104357–104370, 2020.

[14] X. Sun, X. Zhang, C. Xu, M. Xiao, and Y. Tang, ''Tensorial multiview representation for saliency detection via nonconvex approach,'' *IEEE Trans. Cybern.*, pp. 1–14, 2022, doi: 10.1109/TCYB.2021.3139037.

[15] L. Zhang, J. Sun, T. Wang, Y. Min, and H. Lu, ''Visual saliency detection via kernelized subspace ranking with active learning,'' *IEEE Trans. Image Process.*, vol. 29, pp. 2258–2270, 2020.

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

[17] Z. Tu, Y. Ma, C. Li, C. Li, J. Tang, and B. Luo, ''Edge-guided non-local fully convolutional network for salient object detection,'' *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 2, pp. 582–593, Feb. 2021.

[18] X. Huang, Y. Zheng, J. Huang, and Y.-J. Zhang, ''50 FPS object-level saliency detection via maximally stable region,'' *IEEE Trans. Image Process.*, vol. 29, pp. 1384–1396, 2020.

[19] S. Wang, S. Yang, M. Wang, and L. Jiao, ''New contour cue-based hybrid sparse learning for salient object detection,'' *IEEE Trans. Cybern.*, vol. 51, no. 8, pp. 4212–4226, Aug. 2021.

[20] W. J. Zhu, S. Liang, Y. C. Wei, and J. Sun, ''Saliency optimization from robust background detection,'' in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2814–2821.

[21] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, ''Salient object detection: A discriminative regional feature integration approach,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2083–2090.

[22] Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, and P. H. S. Torr, ''Deeply supervised salient object detection with short connections,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 4, pp. 815–828, Apr. 2019.

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

[24] L. Wang, R. Chen, L. Zhu, H. Xie, and X. Li, ''Deep sub-region network for salient object detection,'' *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 2, pp. 728–741, Feb. 2021.

[25] Y. Ji, H. Zhang, Z. Zhang, and M. Liu, ''CNN-based encoder–decoder networks for salient object detection: A comprehensive review and recent advances,'' *Inf. Sci.*, vol. 546, pp. 835–857, Feb. 2021.

[26] J. Chen, Z. Li, and B. Huang, ''Linear spectral clustering superpixel,'' *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3317–3330, Jul. 2017.

[27] K. E. Parsopoulos and M. N. Vrahatis, ''On the computation of all global minimizers through particle swarm optimization,'' *IEEE Trans. Evol. Comput.*, vol. 8, no. 3, pp. 211–224, Jun. 2004.

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

[29] J. Shi, Q. Yan, L. Xu, and J. Jia, ''Hierarchical image saliency detection on extended CSSD,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 4, pp. 717–729, Apr. 2016.

[30] L. Zhang, C. Yang, H. Lu, R. Xiang, and M.-H. Yang, ''Ranking saliency,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 9, pp. 1892–1904, Sep. 2017.

[31] J. Zhang and S. Sclaroff, ''Exploiting surroundedness for saliency detection: A Boolean map approach,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 5, pp. 889–902, Aug. 2016.

[32] S. Goferman, L. Zelnik-Manor, and A. Tal, ''Context-aware saliency detection,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915–1926, Oct. 2012.

[33] H. Z. Jiang, J. D. Wang, Z. J. Yuan, T. Liu, N. N. Zheng, and S. P. Li, ''Automatic salient object segmentation based on context and shape prior,'' in *Proc. Brit. Mach. Vis. Conf.*, 2011, pp. 1–12.

[34] K.-F. Yang, H. Li, C.-Y. Li, and Y.-J. Li, ''A unified framework for salient structure detection by contour-guided visual search,'' *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3475–3488, Aug. 2016.

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

[36] G.-H. Liu and J.-Y. Yang, ''Exploiting color volume and color difference for salient region detection,'' *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 6–16, Jan. 2019.

[37] C. Yang, L. Zhang, and H. Lu, ''Graph-regularized saliency detection with convex-hull-based center prior,'' *IEEE Signal Process. Lett.*, vol. 20, no. 7, pp. 637–640, Jul. 2013.

[38] Y. Xie, H. Lu, and M.-H. Yang, ''Bayesian saliency via low and mid level cues,'' *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1689–1698, May 2013.

[39] H. Li, H. Lu, Z. Lin, X. Shen, and B. Price, ''Inner and inter label propagation: Salient object detection in the wild,'' *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3176–3186, Oct. 2015.

[40] J. M. Zhang, S. Sclaroff, Z. Lin, X. H. Shen, B. Price, and R. Měch, ''Minimum barrier salient object detection at 80 FPS,'' in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1404–1412.

[41] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang, ''Saliency detection via absorbing Markov chain,'' in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1665–1672.

[42] R. Margolin, L. Zelnik-Manor, and A. Tal, ''Saliency for image manipulation,'' *Vis. Comput.*, vol. 29, no. 5, pp. 381–392, 2013.

[43] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, ''Saliency filters: Contrast based filtering for salient region detection,'' in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 733–740.

[44] Z. Liu, W. Zou, and O. L. Meur, ''Saliency tree: A novel saliency detection framework,'' *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 1937–1952, May 2014.

[45] X. Huang and Y. Zhang, ''Water flow driven salient object detection at 180 FPS,'' *Pattern Recognit.*, vol. 76, pp. 95–107, Apr. 2018.

[46] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, ''Frequency-tuned salient region detection,'' in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1597–1604.

[47] S. Wang, M. Wang, S. Yang, and L. Jiao, ''New hierarchical saliency filtering for fast ship detection in high-resolution SAR images,'' *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 351–362, Jan. 2017.

SHIGANG WANG received the bachelor's degree in automation and English from Lanzhou Jiaotong University, Lanzhou, China, in 2012, and the Ph.D. degree in information and communication engineering from Xidian University, Xi'an, China, in 2018. He is currently an Assistant Professor with the School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an. His current research interests include computer vision, machine learning, and remote sensing.