

Received March 18, 2022, accepted April 8, 2022, date of publication April 12, 2022, date of current version April 20, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3166986

Learning Nonlinear Feature Mapping via Constrained Non-Convex Optimization 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, and in part by the Fundamental Research Funds for the Central Universities under Grant 31020180QD094.

ABSTRACT Mimicking the biological visual attention mechanism to detect salient objects in images has been widely studied in recent years. Most of the existing computational models rely on external learning for saliency prediction, which however lack robustness in diversified scenes. In this paper, an unsupervised learning model is proposed to detect salient objects by fully exploiting the internal information of the scene. Specifically, we formulate saliency detection as a mathematical programming problem with which to learn a nonlinear feature mapping from multi-view features to saliency scores. The optimization objective is to maximize the between-class variance of the attended and background regions in the resulting saliency maps, which is statistically optimal. Moreover, to solve the non-convex constrained mathematical programming problem, a hybrid external point method based particle swarm optimization algorithm is developed to find the optimal solution in a rapid manner. Finally, extensive experiments are conducted on four classical saliency benchmark datasets to test the effectiveness of the proposed method and it shows superior qualitative and quantitative performance than the other 16 state-of-the-art unsupervised saliency models.

INDEX TERMS Mathematical programming, external point method, particle swarm optimization (PSO), salient object detection, unsupervised learning.

I. INTRODUCTION

Visual saliency is a remarkable behavior of primates, which allows rapid scene analysis for novel information discovery. Over the past two decades, computational modeling of this intelligent behavior has been an emerging research topic, which has benefited a wide range of scientific and engineering fields [1]–[3]. In its early stage, the source of inspirations for visual saliency modeling mainly comes from biological cognition rules, such as local/global contrast, singularity/sparsity, shape/location prior, etc [4]–[7]. These models are artificial rule-based and generally have good theoretical interpretability. Later on, learning models with handcrafted features are widely studied for this problem, which try to build mappings from low-level visual features to saliency degrees [8]–[10]. Usually, external training images are needed for the learning of the detection model parameters.

The associate editor coordinating the review of this manuscript and approving it for publication was Ehab Elsayed Elattar .

In recent years, the end-to-end learning frameworks are designed to boost the detection performance without manual feature design [11]–[14]. In order to support exact mapping from raw data to saliency score, massive external labelled images are used to fit the model parameters. Concerning the basic model design process, existing saliency detection methods can be roughly classified into two categories, i.e., rule-based and learning-based.

In general, learning-based models are not easily restricted by rigid rules and thus more flexible and autonomous in handling this ill-posed problem. However, the knowledge transfer from external images to test image makes learning based models somewhat general to match well with a specific scene. In fact, biological visual system is more intelligent and can directly extract the regions of interest (ROIs) from the scene without external guidance [15]. Therefore, designing learning models that can make full use of the image internal information will be a promising direction towards more effective saliency detection. Different from traditional

rule-based or learning-based models, this learning from the image itself paradigm possesses great potentials in better formulating this problem. The challenge remains is how to develop self-taught mechanism to advance the saliency learning process towards more intelligent way. Till now, not too much work has been done on this direction and it is a meaningful research topic to develop unsupervised self-taught learning model (Fig. 1) for saliency detection.

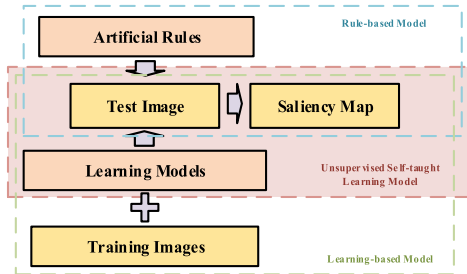


FIGURE 1. Schematic diagram of different kinds of saliency detection approaches.

As an important attempt towards this goal, unsupervised optimization based salient object detection methods are proposed in recent years [16], [17]. The sparse and low rank structures of salient object and background are characterized by different regularization norms to build an unsupervised decomposition model for salient object detection. From the learning perspective, these methods belong to discriminative models and are designed to learn background subspaces for sparse salient outlier detection. As is known, generative models are less dependent on specific assumptions and can provide a more universal expression for complex mapping. Thus, it is possible to design an unsupervised generative model to learn the complex relationships from image features to saliency scores without external supervision. The core is to design a reasonable self-taught mechanism so as to provide momentum to promote the learning towards a more intelligent way.

Motivated by the above observations, an unsupervised non-convex optimization model is proposed in this paper, which can learn nonlinear feature mappings directly from the internal scene for accurate saliency prediction. Different from previous methods, the saliency mapping function is designed to be nonlinear quadratic form, which contains both self-information and mutual-information of the features. This nonlinearity is more capable of describing the complex correlation relationship between feature spaces and saliency domain. Correspondingly, to learn the optimal mapping vector for each individual image, a constrained non-convex programming problem is proposed with the objective to maximize the between-class variance of the attended and background regions in the resulting saliency map. This special optimization objective can provide self-supervision information to guide the learning process to favor saliency modeling. Followed by this, a hybrid algorithm combining the idea of external point method and particle swarm optimization is deduced to solve the above programming problem

efficiently. Finally, with the learned mapping coefficient vector, high-quality saliency maps can be obtained through nonlinear feature transform and joint statistical inference.

The contributions of this paper are summarized as follows. First, we propose a novel unsupervised self-taught paradigm to learn from the image internal information for adaptive saliency detection from complex scenes. Secondly, a nonlinear mapping based constrained non-convex programming model is established to combine multi-view features for accurate saliency prediction. Thirdly, we design a hybrid iterative algorithm by combining the external point method and particle swarm optimization to learn the optimal mapping space from the above model efficiently. The rest of this paper is organized as follows. In Section II, we will introduce the proposed unsupervised nonlinear feature mapping model and hybrid iterative optimization algorithm in detail. Experimental results and performance evaluation on benchmark datasets as well as model ablation study will be given in Section III. Finally, we will briefly conclude this work in Section IV.

II. PROPOSED METHOD

Our method is mainly inspired by the following observations. First, previous methods mostly build linear feature mappings for saliency prediction, which ignore the coupling of different feature dimensions. To achieve stronger modeling capability, a nonlinear feature mapping model is proposed for saliency detection. Secondly, to avoid the bias of knowledge transfer from external images, an unsupervised internal learning scheme is developed for thorough scene analysis. The general pipeline of the proposed salient object detection method is shown in Fig. 2. As can be seen from Fig. 2, the general framework of the proposed method is composed of three components, i.e., nonlinear feature mapping (NFM) construction, unsupervised internal learning (UIL) model design, and hybrid iterative optimization (HIO) algorithm deduction. Detailed descriptions of these components will be given in the following subsections.

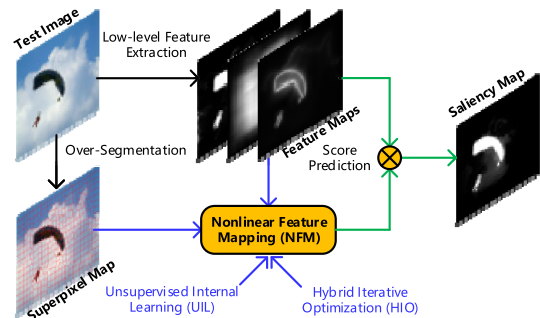


FIGURE 2. General pipeline of the proposed salient object detection method.

A. UNSUPERVISED NONLINEAR FEATURE MAPPING (NFM) MODEL

Building connections between low-level visual features and high-level perception results conforms to the basic biological

cognition rules [18]. Most of the existing methods make linear assumptions for saliency modeling, which however are not sufficient enough to describe this complex relationship. Concerning the low-level features, they are commonly considered as independent channels, but their coupling factor is seldomly studied. In this paper, we propose a nonlinear feature mapping model to better capture the information transferring process from low-level feature to saliency degree.

Given an image $I \in \mathfrak{R}^{m \times n \times 3}$ in RGB color space, we first extract its multi-scale contrast (f_1), center-surround histogram (f_2), and color spatial distribution (f_3) feature maps according to [19]. Meanwhile, the input image is over-segmented into several superpixel regions $\{s_1, s_2, \dots, s_k\}$ using LSC algorithm [20], each of which represents a uniform local image structure. Correspondingly, the feature values of the pixels reside in each superpixel are averaged and assigned as the superpixel feature value. In this way, we can create a feature matrix $FS = [fs_1, fs_2, \dots, fs_k] \in \mathfrak{R}^{3 \times k}$ with each column to represent the 3-dimensional feature vector of a superpixel.

The basic idea is that under a properly designed expression space, the saliency of a superpixel region can be precisely predicted from its feature vector. From a mathematical perspective, this problem can be expressed in the following form

$$sa_i = \psi(fs_i), \tag{1}$$

where, sa_i is the saliency value of the i -th superpixel region and $\psi(\cdot)$ is the functional projection space. Previous methods of this kind, which have limited modeling capability, adopt linear projection and assume no coupling among features. To improve the model adaptability, a quadratic mapping function considering both the feature self and mutual information is proposed in this paper, which has the following form

$$\begin{aligned} \psi(\lambda, f) &= \underbrace{\lambda_{11}f_1^2 + \lambda_{22}f_2^2 + \lambda_{33}f_3^2}_{\text{self information}} \\ &\quad + \underbrace{\lambda_{12}f_1f_2 + \lambda_{13}f_1f_3 + \lambda_{23}f_2f_3}_{\text{mutual information}} \\ &= \underbrace{\sum_{i=1}^3 \lambda_{ii}f_i^2}_{\text{self information}} + \underbrace{\sum_{i=1}^2 \sum_{j=i+1}^3 \lambda_{ij}f_i f_j}_{\text{mutual information}} = \sum_{i=1}^3 \sum_{j=i}^3 \lambda_{ij}f_i f_j \tag{2} \end{aligned}$$

where, $[f_1, f_2, f_3]^T$ is the raw feature vector, and $[\lambda_{11}, \lambda_{22}, \lambda_{33}]^T$ and $[\lambda_{12}, \lambda_{13}, \lambda_{23}]^T$ are respectively the mapping coefficient vectors of its corresponding self and mutual information terms. The quadratic and product terms can be considered as extended features and their linear weighted sum is used to predict saliency score. In practice, we need to learn an optimal mapping function such that foreground regions would have greater saliency values than background regions. To guarantee the existence and uniqueness of the solution, we enforce the following

constraints to the mapping coefficient vector.

$$\sum_{i=1}^3 \sum_{j=i}^3 \lambda_{ij} = 1, \text{ where } 0 \leq \lambda_{ij} \leq 1 \tag{3}$$

All the weighting coefficients are forced to reside in $[0, 1]$ and their sum is equal to 1. By this, the feature representation space is restricted to the surface of unit hyperplane. Till now, we can formulate the basic nonlinear feature mapping model as follows.

$$\begin{aligned} \psi(\lambda, f) &\sum_{i=1}^3 \sum_{j=i}^3 \lambda_{ij} f_i f_j \\ \text{s.t. } &\sum_{i=1}^3 \sum_{j=i}^3 \lambda_{ij} = 1, 0 \leq \lambda_{ij} \leq 1 \tag{4} \end{aligned}$$

The problem left is how to find an ideal mapping space for accurate saliency score prediction from the extended features. Inspired by the idea of maximizing the between-class variance, we propose an optimal criteria to evaluate the quality of the solution. The basic idea is to maximize the between-class variance of foreground and background saliency to produce high-contrast saliency map results. Denote the number of fore/back-ground superpixels as k_f and k_b , and the average saliency value of fore/back-ground region as s_f and s_b . The between-class variance σ_B^2 of a given saliency map is defined as

$$\sigma_B^2 = \frac{k_f}{k}(s_f - s_g)^2 + \frac{k_b}{k}(s_b - s_g)^2 \tag{5}$$

where, s_g is the average value of the saliency map. Through mathematical deduction, the above equation can also be written as

$$\sigma_B^2 = \frac{k_f k_b}{k^2} (s_f - s_b)^2 \tag{6}$$

By expanding the two terms s_f and s_b , we have

$$\begin{aligned} \sigma_B^2 &= \frac{k_f k_b}{k^2} \left[\frac{1}{k_f} \sum_{s_i \in S_f} sa_i - \frac{1}{k_b} \sum_{s_i \in S_b} sa_i \right]^2 \\ &= \frac{k_f k_b}{k^2} \left[\frac{1}{k_f} \sum_{s_i \in S_f} \sum_{i=1}^6 \hat{\lambda}_i \hat{f}_i s_i - \frac{1}{k_b} \sum_{s_i \in S_b} \sum_{i=1}^6 \hat{\lambda}_i \hat{f}_i s_i \right]^2 \\ &= \frac{k_f k_b}{k^2} \left\{ \sum_{i=1}^6 \hat{\lambda}_i \left[\frac{1}{k_f} \sum_{s_i \in S_f} \hat{f}_i s_i - \frac{1}{k_b} \sum_{s_i \in S_b} \hat{f}_i s_i \right] \right\}^2 \tag{7} \end{aligned}$$

where, s_f and s_b are the sets of fore/back-ground superpixels, $\hat{\lambda} = [\lambda_{11}, \lambda_{22}, \lambda_{33}, \lambda_{12}, \lambda_{13}, \lambda_{23}]^T$ is the joint mapping coefficient vector, and $\hat{f}s = [f_1^2, f_2^2, f_3^2, f_1f_2, f_1f_3, f_2f_3]^T$ is the extended feature vector.

As a statistical criteria, σ_B^2 can reveal the divergence of foreground and background saliency value distributions. By maximizing σ_B^2 , we can find an optimal mapping coefficient vector with which to produce high-quality saliency

maps. Towards this end, we build an objective function to evaluate the quality of the solution.

$$F(\hat{\lambda}) = \frac{k_f k_b}{k^2} \left\{ \sum_{i=1}^6 \hat{\lambda}_i \left[\frac{1}{k_f} \sum_{s_i \in S_f} \hat{f} s_i - \frac{1}{k_b} \sum_{s_i \in S_b} \hat{f} s_i \right] \right\}^2 \quad (8)$$

Combing the constraint conditions in (3), we can obtain the standard form of our saliency optimization model.

$$\begin{aligned} \max_{\hat{\lambda}^*} F(\hat{\lambda}) &= \frac{k_f k_b}{k^2} \left\{ \sum_{i=1}^6 \hat{\lambda}_i \left[\frac{1}{k_f} \sum_{s_i \in S_f} \hat{f} s_i - \frac{1}{k_b} \sum_{s_i \in S_b} \hat{f} s_i \right] \right\}^2 \\ \text{s.t. } \sum_{i=1}^6 \hat{\lambda}_i - 1 &= 0, \quad -\hat{\lambda}_i \leq 0 (i = 1, 2, \dots, 6) \end{aligned} \quad (9)$$

Since s_f and s_b are not known in practice, they need to be iteratively learned with $\hat{\lambda}$ to maximize the objective function. This is a compound constrained optimization problem with dynamic solution search space. Below we will develop a feasible way to solve the above problem in an efficient manner.

B. EFFICIENT HYBRID ITERATIVE OPTIMIZATION (HIO) ALGORITHM

In the above hybrid saliency optimization model, the objective function depends both on the immediate variable $\hat{\lambda}$ and two hidden variables s_f and s_b . The division of s_f and s_b is based on the saliency map generated under the mapping of immediate variable $\hat{\lambda}$. Meanwhile, the quality of the immediate variable $\hat{\lambda}$ is evaluated based on the divided fore/back-ground superpixel sets s_f and s_b . The whole learning process works in a self-taught manner with no external information involved. To solve this optimization problem, we first adopt idea from the external point method (EPM) to merge the constraint terms into the optimization objective. Here, we use the quadratic penalty function and the corresponding penalty term can be constructed as follows.

$$H(\hat{\lambda}) = \sum_{i=1}^6 \left(\max(0, -\hat{\lambda}_i) \right)^2 + \left(\sum_{i=1}^6 \hat{\lambda}_i - 1 \right)^2 \quad (10)$$

By adding the penalty term to the objective function, we can obtain the overall fitness function in the following.

$$\begin{aligned} \text{Fitness}(\hat{\lambda}, \eta) &= F(\hat{\lambda}) - \eta H(\hat{\lambda}) \\ &= \frac{k_f k_b}{k^2} \left\{ \sum_{i=1}^6 \hat{\lambda}_i \left[\frac{1}{k_f} \sum_{s_i \in S_f} \hat{f} s_i - \frac{1}{k_b} \sum_{s_i \in S_b} \hat{f} s_i \right] \right\}^2 \\ &\quad - \eta \left\{ \sum_{i=1}^6 \left(\max(0, -\hat{\lambda}_i) \right)^2 + \left(\sum_{i=1}^6 \hat{\lambda}_i - 1 \right)^2 \right\} \end{aligned} \quad (11)$$

where, η is the penalty coefficient used to control the penalty strength for any illegal solutions. In practice, η will gradually

increase along with the iteration process to guarantee that the solution moves towards feasible space.

The above fitness function contains both hidden variables and hyper-parameter, which has no explicit analytical solution in practice. Here, we propose a particle swarm optimization (PSO) based hybrid algorithm for efficient solution search. The basic idea is to first randomly sample a group of candidate solutions in the feasible space and then iteratively update their positions towards greater overall fitness value. Given a candidate solution, the saliency value of each superpixel region can be determined according to (2). Followed by this, we can organize a coarse saliency map with the predicted saliency of superpixel regions. Using Otsu algorithm, we binarize the coarse saliency map to get the fore/back-ground superpixel sets. In turn, based on the divided fore/back-ground sets, the fitness value of the candidate solution can be calculated with (11). After each iteration, the position of each candidate solution is updated according to the heuristic rules of PSO and the penalty coefficient is increased to force the candidate solution to reside in feasible space. This iteration process will proceed until the optimal solution of the swarm reaches a stable state. The hidden variables and hyper-parameter in (11) change constantly during iterations and meanwhile the fitness function value is an approximate estimation result due to the coarse saliency map binarization. Therefore, (11) is a dynamic objective function with noisy optimization environment, which can be better coped with by PSO in a stable manner [21].

Pseudo code of the proposed hybrid iterative optimization algorithm is summarized in Algorithm 1. The general process combines both the idea of the external point method and particle swarm optimization. The penalty coefficient η is magnified along with the iteration number so as to guarantee the solution to move towards legal space. Meanwhile, a group of particles collaborate intelligently to search for the optimal solution in order to maximize the fitness function. By simultaneously considering the legality and optimality of the solution, satisfactory results can be obtained in a rapid manner for saliency mapping. It is also worth mentioning that the division of s_f and s_b can add new momentum to this problem and the use of Otsu algorithm can direct it to the right direction.

With the above EPM-PSO algorithm, the optimal mapping coefficient vector $\hat{\lambda}^*$ can be obtained, with which we can predict the saliency values of superpixel regions according to (2). After filling the predicted saliency values into the corresponding image space and performing pixel-level statistical inference, we will get the final saliency map of the image for succeeding analysis. Since both low and mid-level information of the image are exploited, high-quality saliency map can be obtained as will be seen later.

C. IMPLEMENTATION DETAILS

In the experiments, the expected number of superpixels in LSC algorithm and the number of particles in the swarm

Algorithm 1 Pseudo Code of the Hybrid Iterative Optimization Algorithm

Input: Extended superpixel feature matrix $\hat{F}S = [\hat{f}_{s_1}, \hat{f}_{s_2}, \dots, \hat{f}_{s_k}] \in \mathbb{R}^{6 \times k}$, penalty coefficient η in EPM, velocity updating parameters $\{\omega, c_1, c_2\}$ in PSO.

Output: Optimal joint mapping coefficient vector $\hat{\lambda}^*$.

Initialization: Iteration number $t=1$; Swarm size sn , position $Swarm.pos \in \mathbb{R}^{6 \times sn}$, velocity $Swarm.vel \in \mathbb{R}^{6 \times sn}$, personal best $Swarm.pbest \in \mathbb{R}^{6 \times sn}$, global best $Swarm.gbest(t) \in \mathbb{R}^{6 \times 1}$, and fitness vector $Swarm.fit \in \mathbb{R}^{sn \times 1}$.

1: do

2: for $i = 1:1:sn$

3: $sa = Swarm.pos(:, i)^T \times \hat{F}S$ % compute superpixel saliency

4: Organize coarse saliency map $Sa \in \mathbb{R}^{m \times n}$ and use Otsu algorithm to divide the superpixels into fore/background sets S_f and S_b .

5: Calculate fitness function $Fitness(Swarm.pos(:, i), \eta)$ according to (11).

6: if ($Fitness(Swarm.pos(:, i), \eta) > Swarm.fit(i)$) % find pbest

7: $Swarm.pbest(:, i) = Swarm.pos(:, i); Swarm.fit(i) = Fitness(Swarm.pos(:, i), \eta)$

8: end

9: end

10: $i^* = \min_{i \in [1, sn]} Swarm.fit(i); Swarm.gbest(t + 1) = Swarm.pbest(:, i^*)$ % find gbest

11: for $i = 1:1:sn$ % update particle velocity and position

12: $Swarm.vel(:, i) = \omega \times Swarm.vel(:, i) + \dots$
 $c_1 \times rand(1) \times [Swarm.pbest(:, i) - Swarm.pos(:, i)] + \dots$
 $c_2 \times rand(1) \times [Swarm.gbest(t + 1) - Swarm.pos(:, i)]$

13: $Swarm.pos(:, i) = Swarm.pos(:, i) + Swarm.vel(:, i)$

14: end

15: $t = t + 1; \eta = \sqrt{t} \times \eta$ % increase penalty coefficient

16: while ($\|Swarm.gbest(t+1) - Swarm.gbest(t)\|_2 > \varepsilon$)

are respectively set to 500 and 36. Also, the three velocity updating parameters ω , c_1 and c_2 in PSO are fixed to be 1, 2 and 2 during the iteration process. For reasonable running of the algorithm, the initial value for η and the termination error ε are chosen to be 1 and 0.001. Concerning the initialization of the swarm particles, we use random function to generate legal vectors for the particle positions, velocities, personal best and global best. Finally, the fitness vector $Swarm.fit$ is initialized to be an sn -dimensional negative infinite vector for ease of succeeding update of solutions. During implementation, our code is written in MATLAB and run on a HP Z8 G4 workstation with 8 core CPU of 1.70 GHz and 64 GB RAM, and 64 bits Windows 10 operating system. It is also worth mentioning that each image will be assigned a specific optimal mapping function during its unsupervised self-taught learning process.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS**A. EXPERIMENTAL SETUP**

In this section, our model (referred to as NFM) is tested on four classical saliency benchmark datasets (PASCAL-S [22], ECSSD [23], DUT-OMRON [24] and THUS-10000 [25]) along with 19 state-of-the-art saliency detection methods, which are respectively denoted as BMS [26], CGVS [27],

CRF [19], DFI [28], DSR [29], FCB [30], GR [31], HS [23], JCS [32], LDF [33], LPS [34], MBS+ [35], MC [36], MST [37], PSO [38], RC [25], ST [39], WF [40] and WSC [41]. For fair comparison, we use the source codes released by the original authors for performance assessment except CRF and PSO whose codes are not yet publicly available. In practice, we use the simulation code provided in [42] for CRF and implement PSO on our own with MATLAB by following up [38]. During experiments, we use the default parameters suggested by the original authors for saliency map generation.

Both qualitative and quantitative results are obtained for comprehensive analysis of the modeling capabilities. As a common practice, saliency map is used as a qualitative way to evaluate the quality of the detection results. For observation convenience, only saliency maps of the ten top-performing unsupervised methods (which are chosen based on quantitative comparison results) as well as the three deep learning methods are displayed for visual comparison. Meanwhile, precision-recall (PR) and F-measure curves of the saliency detectors are drawn to quantitatively evaluate the detection performance. Precision is the ratio between true positive (TP) and the sum of true positive (TP) and false positive (FP), while recall is the ratio between true positive (TP) and the sum of true positive (TP) and false negative (FN). The two indexes are complementary to each other and usually need to be balanced in practice. To give a unified evaluation standard, F-measure is proposed which is the harmonic mean of precision and recall.

$$F_\beta = \frac{(1 + \beta^2) \times \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}} \quad (12)$$

where, $\beta^2 = 0.3$ is used to give more emphasis on precision than recall as suggested in [43]. For each 8 bits grey-scale saliency map $S \in \mathbb{R}^{m \times n}$, we use a dynamic threshold ranging from 0 to 255 to binarize it, and 256 precision-recall point pairs can be calculated by comparing it with the binary ground-truth map $G \in \mathbb{R}^{m \times n}$. After plotting all the 256 point pairs on a 2-D plane, we can draw the PR curve of the saliency map. Correspondingly, 256 F-measure values can be derived according to (12) to form the F-measure curve.

Also, we use four numerical indexes, including area under curve (AUC), mean absolute error (MAE), S-measure (SM) [44] and weighted F-measure (WF) [45], to evaluate the salient object detection performance in a more comprehensive way. AUC is the area covered by the receiver operator characteristic (ROC) curve, which is drawn based on 256 detection and false alarm rate point pairs similar to that of PR curve. MAE is the average pixel-wise absolute difference between the saliency map and its corresponding ground-truth map on a benchmark dataset. As two newly proposed indexes, SM and WF take image structure and dependency relationship into consideration and thus is able to provide more reliable and objective performance evaluation results.

Below we will give the saliency maps and numerical indexes of the various saliency detectors on the benchmark

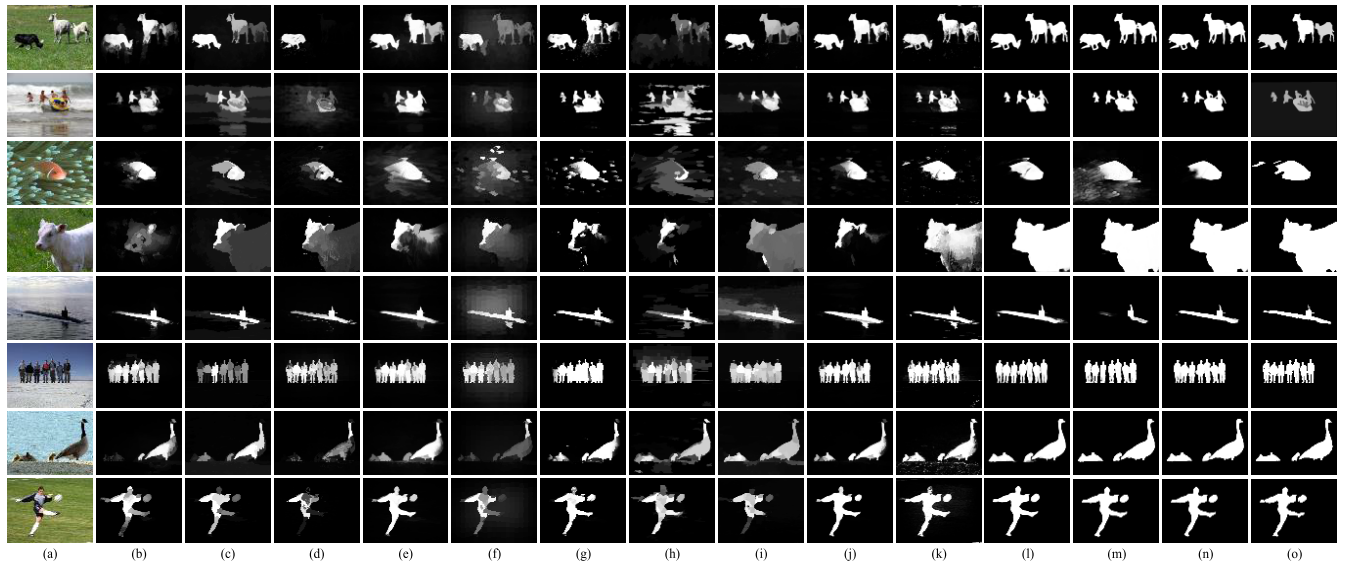


FIGURE 3. Some saliency map results of the top-ten performing unsupervised methods as well as the three deep learning methods on the four benchmark datasets. From (a) to (o) are respectively the input images, saliency maps of DSR, HS, LPS, MBS+, MC, MST, RC, ST, WF, NFM, DFI, JCS and LDF, and the ground truths. Every two rows from top to down correspond to test images from PASCAL-S, ECSSD, DUT-OMRON and THUS-10000 dataset.

datasets for comprehensive analysis and comparison of the model performance.

B. PERFORMANCE EVALUATION RESULTS

The ability to accurately detect salient objects from complex scenes is a major difference for various saliency models. It is therefore necessary to observe the model behavior in typical challenging situations. Fig. 3 shows some representative saliency maps of the top-ten performing unsupervised methods (DSR, HS, LPS, MBS+, MC, MST, RC, ST, WF and NFM) as well as the three deep learning methods (DFI, JCS and LDF) on the four benchmark datasets. As can be seen from the results, our method can produce high-quality saliency maps comparable to or better than the other counterparts, when in face of multiple, hidden, and cropped objects as well as low image contrast and motion blur. Due to the complexity of saliency generation mechanism, models relying heavily on certain aspect of cognitive cues will lack robustness in face of diversified scenes. For example, the cow in the fourth image is not integrally detected by most methods for their excessive dependence on boundary prior. Also, the football and its player are not given full attention by some methods for their overemphasis on focussness prior. Free from specific cognitive cues, our model can build representative mapping space for complex scene saliency modeling. It provides a more general way to handle the different forms of saliency information and thus is better suited to this problem.

Meanwhile, PR and F-measure curves of all the saliency models are respectively shown in Figs. 4 and 5 for objective performance evaluation. As can be observed from Fig. 4, among all the unsupervised methods, the PR curves of NFM keep lying in the most up-right corner of the plot, indicating the superiority of our model in unsupervised saliency

detection. Till now, the saliency modeling performance is going to be saturated on some simple datasets, but still needs improvement in challenging datasets. PR indexes of the saliency detectors on the simple THUS-10000 dataset are relatively high and the advantage of our method over the others is not too obvious. When it comes to more challenging datasets like PASCAL-S and ECSSD, the performance gap between our method and the others becomes enlarged. And our method improves the PR index on the complex DUT-OMRON dataset to a large extent, implying its strong modeling capability in complex scenes. Similar results can also be observed from the F-measure index in Fig. 5. Since our model adopts nonlinear feature representation and self-taught learning mechanism, it is more suitable for detecting salient objects from diversified scenes than the other unsupervised methods. Concerning the three deep learning methods, they have favorable PR and F-measure indexes via the usage of massive external supervision information.

Besides, we show the numerical indexes, including area under ROC curve (AUC), mean absolute error (MAE), S-measure (SM) and weighted F-measure (WF), of the saliency models on the four datasets in Tables 1, 2, 3 and 4 respectively for further performance comparison. The up/down arrow means the larger/smaller the value is, the better the detection performance will be. For observation convenience, the best three non-deep learning results on each index are marked with red, green and blue. Since SM and WF both consider the dependency relationship among pixels, they can provide more objective evaluation standards than that of AUC and MAE. As can be seen, our method has the best SM and WF indexes among all the unsupervised methods on the four datasets, implying its superior performance under diversified scenarios. According to the three indexes (AUC,

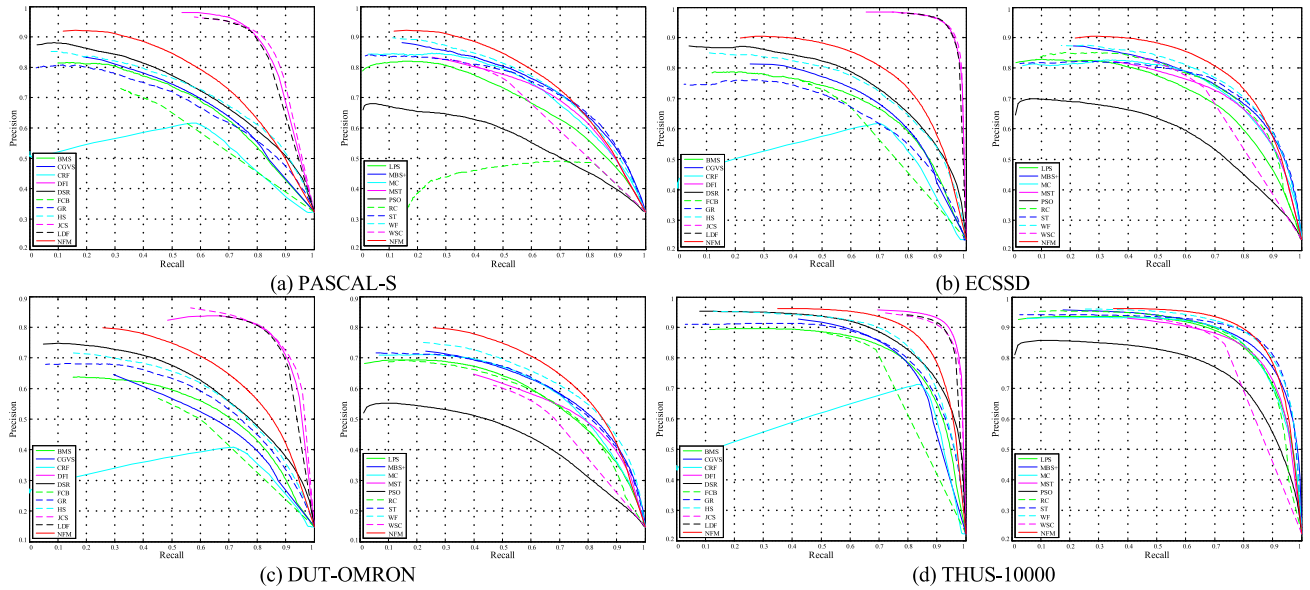


FIGURE 4. PR curves of all the saliency methods on the four benchmark datasets. For observation convenience, curves of the 19 comparison methods are shown in two separate subplots and the curve of our method is shown in both subplots.

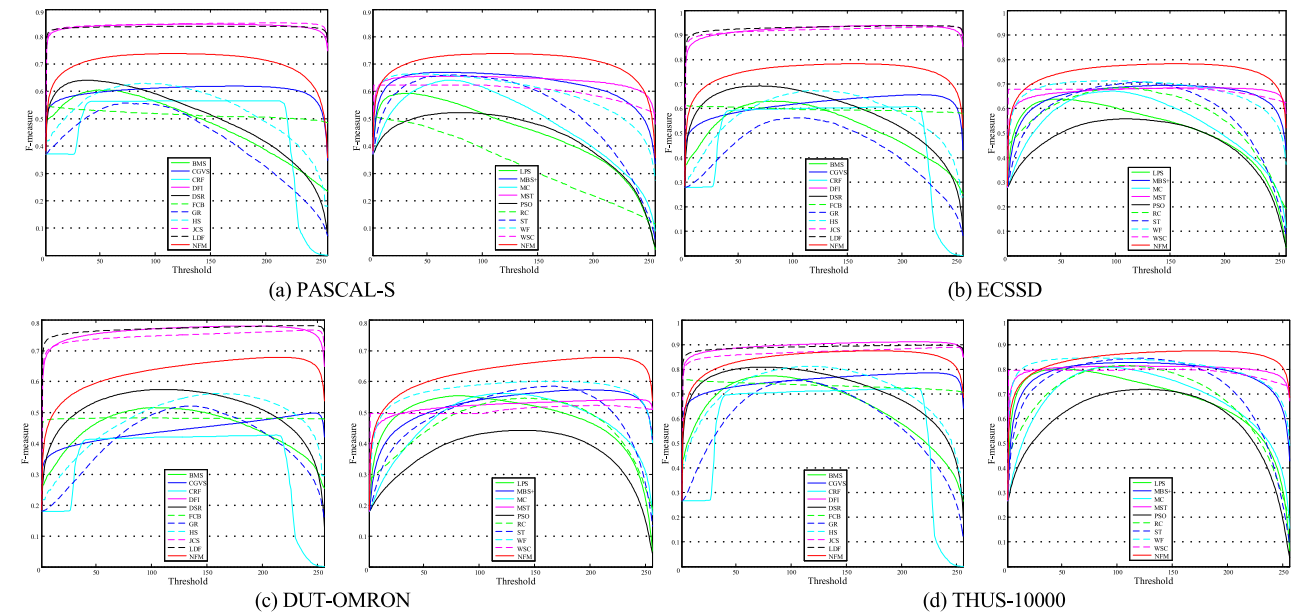


FIGURE 5. F-measure curves of all the saliency methods on the four benchmark datasets. For observation convenience, curves of the 19 comparison methods are shown in two separate subplots and the curve of our method is shown in both subplots.

TABLE 1. Numerical indexes of the saliency models on PASCAL-S dataset.

Indexes	BMS	CGVS	CRF	DFI	DSR	FCB	GR	HS	JCS	LDF
AUC \uparrow	0.8221	0.6403	0.7216	0.9213	0.8529	0.6791	0.8204	0.8489	0.9227	0.9048
MAE \downarrow	0.2184	0.2272	0.2861	0.0704	0.1992	0.0726	0.2834	0.2474	0.0771	0.0674
SM \uparrow	0.5655	0.6161	0.5599	0.8067	0.5829	0.515	0.5644	0.6163	0.8147	0.7985
WF \uparrow	0.4053	0.5308	0.4142	0.7927	0.4344	0.4156	0.376	0.4523	0.8112	0.7916
Indexes	LPS	MBS+	MC	MST	PSO	RC	ST	WF	WSC	NFM
AUC \uparrow	0.83	0.8741	0.8644	0.8447	0.7559	0.6849	0.856	0.8644	0.7304	0.7918
MAE \downarrow	0.2111	0.1908	0.2181	0.1909	0.2695	0.2944	0.2117	0.182	0.193	0.2176
SM \uparrow	0.5369	0.6496	0.6106	0.6354	0.5376	0.474	0.6349	0.6252	0.581	0.6753
WF \uparrow	0.3778	0.535	0.4244	0.5523	0.3776	0.3415	0.4767	0.5073	0.4966	0.5571

MAE and WF) provided in [17] (on PASCAL-S) and [16] (on ECSSD and THUS-10000), our method shows better

numerical results than the above two recent unsupervised optimization based approaches.

TABLE 2. Numerical indexes of the saliency models on ECSSD dataset.

Indexes	BMS	CGVS	CRF	DFI	DSR	FCB	GR	HS	JCS	LDF
AUC ↑	0.8781	0.8648	0.7923	0.988	0.9142	0.7523	0.8754	0.9062	0.9866	0.9814
MAE ↓	0.2142	0.1934	0.2567	0.0349	0.1713	0.0577	0.2835	0.2275	0.0338	0.0335
SM ↑	0.6485	0.679	0.6351	0.9266	0.685	0.6201	0.6169	0.6851	0.9235	0.9244
WF ↑	0.437	0.5535	0.4198	0.9086	0.5106	0.5101	0.359	0.4572	0.9116	0.9162
Indexes	LPS	MBS+	MC	MST	PSO	RC	ST	WF	WSC	NFM
AUC ↑	0.8947	0.9253	0.9254	0.9035	0.8209	0.9017	0.9196	0.9236	0.8074	0.854
MAE ↓	0.1849	0.1707	0.2025	0.1568	0.2455	0.186	0.1924	0.1513	0.1519	0.1759
SM ↑	0.6366	0.7238	0.6924	0.7124	0.6193	0.7054	0.7246	0.728	0.6877	0.758
WF ↑	0.4482	0.5695	0.4583	0.6055	0.3939	0.5121	0.5136	0.5942	0.5942	0.6359

TABLE 3. Numerical indexes of the saliency models on DUT-OMRON dataset.

Indexes	BMS	CGVS	CRF	DFI	DSR	FCB	GR	HS	JCS	LDF
AUC ↑	0.8582	0.8249	0.7734	0.9505	0.8927	0.7418	0.868	0.8694	0.9502	0.9344
MAE ↓	0.1754	0.222	0.2748	0.055	0.1388	0.0496	0.2591	0.2274	0.0688	0.0517
SM ↑	0.6287	0.5925	0.5578	0.8395	0.6725	0.6051	0.6006	0.6326	0.8247	0.839
WF ↑	0.3562	0.3996	0.2919	0.7449	0.4188	0.4358	0.2809	0.3523	0.7327	0.7674
Indexes	LPS	MBS+	MC	MST	PSO	RC	ST	WF	WSC	NFM
AUC ↑	0.8717	0.9027	0.8878	0.878	0.8087	0.8576	0.8918	0.9024	0.7958	0.8417
MAE ↓	0.1448	0.1679	0.1863	0.1609	0.2301	0.1883	0.1839	0.13	0.1361	0.1565
SM ↑	0.6521	0.669	0.6491	0.6529	0.5843	0.645	0.6677	0.7008	0.638	0.7047
WF ↑	0.3975	0.4394	0.3508	0.471	0.2916	0.3741	0.3877	0.4973	0.4615	0.5138

TABLE 4. Numerical indexes of the saliency models on THUS-10000 dataset.

Indexes	BMS	CGVS	CRF	DFI	DSR	FCB	GR	HS	JCS	LDF
AUC ↑	0.9321	0.9055	0.8755	0.9759	0.9537	0.8213	0.9543	0.9476	0.9696	0.9637
MAE ↓	0.1507	0.1154	0.2089	0.0454	0.1207	0.0402	0.1976	0.1486	0.0623	0.0475
SM ↑	0.7461	0.7803	0.7241	0.9093	0.7809	0.7189	0.7439	0.7866	0.8853	0.8979
WF ↑	0.5675	0.7001	0.4958	0.8841	0.6519	0.6467	0.4907	0.6035	0.8557	0.8763
Indexes	LPS	MBS+	MC	MST	PSO	RC	ST	WF	WSC	NFM
AUC ↑	0.956	0.9677	0.9551	0.9416	0.9177	0.9413	0.9658	0.9649	0.8556	0.9373
MAE ↓	0.1251	0.1073	0.1451	0.0974	0.1786	0.1372	0.1224	0.0906	0.1051	0.1061
SM ↑	0.7472	0.8228	0.7852	0.8098	0.741	0.7893	0.8284	0.8292	0.7751	0.8481
WF ↑	0.6159	0.7005	0.578	0.7364	0.5202	0.6094	0.662	0.7439	0.7175	0.765

Finally, concerning the modeling efficiency, the EPM-PSO algorithm can converge rapidly to an ideal solution in a few iterations for saliency prediction. The average running time of our method on a test image is around 3 seconds using the aforementioned computing platform. It is worth mentioning that the EPM-PSO algorithm can be further parallelized to greatly speed up the modeling process. In this regard, our method can obtain comparable or even favorable running efficiency than traditional salient object detection approaches. Since models based on deep learning rely on high-performance GPU platform, they are generally fast in speed during the testing stage. Below we will conduct ablation study to verify the effectiveness of each component in our model for saliency detection.

C. ABLATION STUDY AND MODEL VERIFICATION

From the view of information fusion, our saliency model combines multi-view feature maps under a nonlinear mapping space for comprehensive saliency prediction. Through joint optimization learning, individual feature maps are efficiently fused to get enhanced saliency map results. To further verify the effectiveness of the proposed saliency fusion model, we conduct ablation study on THUS-10000 dataset with PR curve as the evaluation criteria. Specifically,

the three feature maps used in our model, including multi-scale contrast (MSC), center-surround histogram (CSH), and color spatial distribution (CSD), are compared with the ground truth map to derive their PR curves on THUS-10000. Fig. 6 shows some example feature maps and their correspondingly fused saliency maps by the proposed method on THUS-10000 dataset. Meanwhile, the PR curves of the three feature maps on THUS-10000 dataset are drawn in Fig. 7 for performance improvement evaluation.

As can be seen from Figs. 6 and 7, both the qualitative and quantitative results are boosted to a large extent after the proposed nonlinear optimization modeling process. The multi-view feature maps contain rich contour, location and color information, which are fully exploited by the representative and discriminative learning model for high-quality saliency generation. As a result of this, the fused saliency map shows better visual and numerical effects than each individual feature map. After the fusion process, the maximum precision value is increased by 0.1, which is a significant improvement on this dataset.

In Table 5, we summarize the learned optimal mapping coefficient vectors by the proposed method for the 6 test images displayed in Fig. 6. As can be observed, each image is assigned a non-negative projection vector in the feature

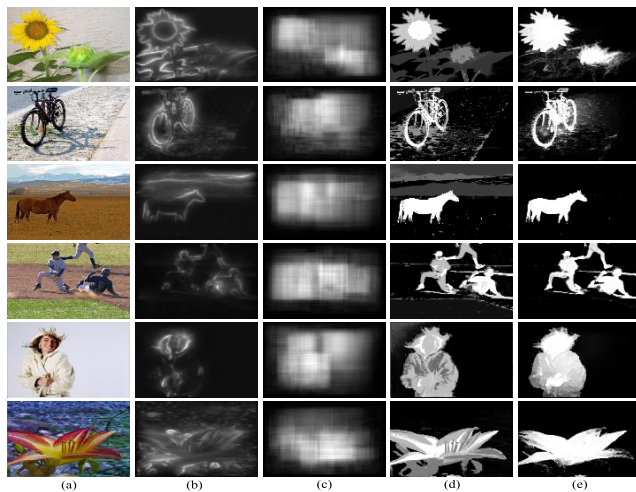


FIGURE 6. Some feature maps and their correspondingly fused saliency maps by the proposed method on THUS-10000 dataset. From (a) to (e) are respectively the input images, three feature maps (MSC, CSH, and CSD), and the final saliency maps. Each row from top to down corresponds to a test image from THUS-10000 dataset.

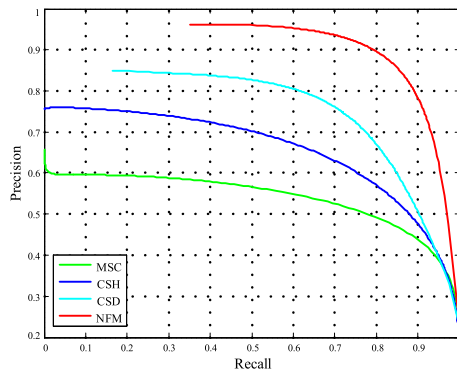


FIGURE 7. PR curves of the three feature maps (MSC, CSH, and CSD) as well as NFM on THUS-10000 dataset.

TABLE 5. Learned mapping coefficient vectors for images in Fig. 6.

Test image	Learned optimal mapping coefficient vector
1 st image	[0.3351, 0.3368, 0.0125, 0.0177, 0.2930, 0.0048] ^T
2 nd image	[0.1197, 0.3330, 0.2505, 0.1753, 0.0499, 0.0716] ^T
3 rd image	[0.1960, 0.0063, 0.3484, 0.0284, 0.0595, 0.3614] ^T
4 th image	[0.0255, 0.0040, 0.3396, 0.1110, 0.1990, 0.3209] ^T
5 th image	[0.0666, 0.3483, 0.1317, 0.0436, 0.1084, 0.3013] ^T
6 th image	[0.0192, 0.2828, 0.0829, 0.0546, 0.3933, 0.1671] ^T

representation space for accurate saliency estimation. Take the first image as an example. The self-information from MSC and CSH, as well as the mutual-information between MSC and CSD are mainly used to constitute the saliency map. The two self-information components provide contour and location estimation results, and the mutual-information component provides compound contour and color analysis result. They together contribute to the formation of the final fine-grained saliency map. Due to the designed intelligent learning and optimization mechanism, our model can sense the importance of each feature map and meanwhile build correlation relationships among them to compose the ideal

saliency map. Through unsupervised self-taught learning process, a scene-specific saliency mapping function can be constructed for each individual image without external guidance. Thus, it is quite suitable for the robust modeling of saliency from complex open scenes.

Since the PSO has some randomness due to the initialization of the particle velocity and position as well as the update of the particle velocity, it is necessary to observe the stability of the optimization algorithm during the saliency modeling process. We run the EPM-PSO algorithm on PASCAL-S dataset for 10 independent times to check the influence of its randomness on the detection performance. The experimental results indicate that PSO can converge stably to the optimal solution in each individual test, leading to exactly the same saliency map result. In terms of convergent speed, variance of the average running time (in seconds) on a test image among the 10 independent runs is 0.02, which further proves the stability of the convergent process. In general, the essence of the optimization objective is to find a nonlinear unit mapping so that the between-class variance of the feature points after projection is maximized. The collaborative search mechanism of PSO makes the optimization algorithm avoid being trapped into local optimum and converge rapidly to the optimal solution.

IV. CONCLUSION

In this paper, we propose an unsupervised internal learning model to construct nonlinear feature mapping for robust salient object detection from complex natural images. A novel optimization objective based on between-class variance maximization is deduced to form a constrained non-convex programming problem. Based on this, we develop an external point method based particle swarm optimization algorithm to search for the ideal solution in real time. Extensive experiments on benchmark datasets verify the superiority of the proposed method than other classic saliency detection models, especially in face of complex imaging conditions and scenes. Also, the effectiveness of the nonlinear feature integration scheme is confirmed by model ablation study. Different from previous methods, the proposed model is free from specific rules or generalization bias, and possesses self-driven saliency learning capability. In general, we provide a more representative and focused model for boosting the feature mapping based saliency detection performance. In the future, it is a meaningful research topic to develop unsupervised scene-specific saliency learning model to meet the requirement for open world scene understanding.

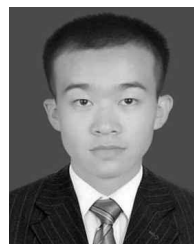
ACKNOWLEDGMENT

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] 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.
- [3] L. Han, X. Li, and Y. Dong, "Convolutional edge constraint-based U-Net for salient object detection," *IEEE Access*, vol. 7, pp. 48890–48900, 2019.
- [4] Y. Zhou, T. Zhang, S. Huo, C. Hou, and S.-Y. Kung, "Adaptive irregular graph construction-based salient object detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 6, pp. 1569–1582, Jun. 2020.
- [5] 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, May 2018.
- [6] J. Zhang, S. Fang, K. A. Ehinger, H. Wei, W. Yang, K. Zhang, and J. Yang, "Hypergraph optimization for salient region detection based on foreground and background queries," *IEEE Access*, vol. 6, pp. 26729–26741, 2018.
- [7] S. Yang, Q. Gao, and S. Wang, "Learning a deep representative saliency map with sparse tensors," *IEEE Access*, vol. 7, pp. 117861–117870, 2019.
- [8] 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.
- [9] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 2106–2113.
- [10] 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.
- [11] W. Guan, T. Wang, J. Qi, L. Zhang, and H. Lu, "Edge-aware convolution neural network based salient object detection," *IEEE Signal Process. Lett.*, vol. 26, no. 1, pp. 114–118, Jan. 2019.
- [12] 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.
- [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] 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.
- [15] C. Chen, Y. Li, S. Li, H. Qin, and A. Hao, "A novel bottom-up saliency detection method for video with dynamic background," *IEEE Signal Process. Lett.*, vol. 25, no. 2, pp. 154–158, Feb. 2018.
- [16] M. Li, Y. Zhang, M. Q. Xiao, W. Q. Zhang, and X. L. Sun, "Unsupervised learning for salient object detection via minimization of bilinear factor matrix norm," *IEEE Trans. Neural Netw. Learn. Syst.*, Early Access, Aug. 30, 2021, doi: 10.1109/TNNLS.2021.3105276.
- [17] M. Li, Y. Yang, L. Xu, C. Xu, and X. L. Sun, "Salient object detection based on unified convex surrogate for non-convex Schatten-p norm," *IEEE Access*, vol. 8, pp. 20171–20180, 2020.
- [18] 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.
- [19] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Y. Shum, "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 353–367, Feb. 2011.
- [20] J. Chen, Z. Li, and B. Huang, "Linear spectral clustering superpixel," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3317–3330, Jul. 2017.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] J. Zhang and S. Sclaroff, "Saliency detection: A Boolean map approach," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 153–160.
- [27] 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.
- [28] J.-J. Liu, Q. Hou, and M.-M. Cheng, "Dynamic feature integration for simultaneous detection of salient object, edge, and skeleton," *IEEE Trans. Image Process.*, vol. 29, pp. 8652–8667, 2020.
- [29] 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.
- [30] 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.
- [31] 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.
- [32] A. Li, J. Zhang, Y. Lv, B. Liu, T. Zhang, and Y. Dai, "Uncertainty-aware joint salient object and camouflaged object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 10066–10076.
- [33] J. Wei, S. Wang, Z. Wu, C. Su, Q. Huang, and Q. Tian, "Label decoupling framework for salient object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 13022–13031.
- [34] 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.
- [35] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, and R. Mech, "Minimum barrier salient object detection at 80 FPS," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1404–1412.
- [36] 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.
- [37] W.-C. Tu, S. He, Q. Yang, and S.-Y. Chien, "Real-time salient object detection with a minimum spanning tree," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2334–2342.
- [38] N. Singh, R. Arya, and R. K. Agrawal, "A novel approach to combine features for salient object detection using constrained particle swarm optimization," *Pattern Recognit.*, vol. 47, no. 4, pp. 1731–1739, Apr. 2014.
- [39] Z. Liu, W. Zou, and O. Le Meur, "Saliency tree: A novel saliency detection framework," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 1937–1952, May 2014.
- [40] X. Huang and Y. Zhang, "Water flow driven salient object detection at 180 FPS," *Pattern Recognit.*, vol. 76, pp. 95–107, Apr. 2018.
- [41] N. Li, B. Sun, and J. Yu, "A weighted sparse coding framework for saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5216–5223.
- [42] S. Dhar, V. Ordonez, and T. L. Berg, "High level describable attributes for predicting aesthetics and interestingness," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 1657–1664.
- [43] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1597–1604.
- [44] D.-P. Fan, M.-M. Cheng, Y. Liu, T. Li, and A. Borji, "Structure-measure: A new way to evaluate foreground maps," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 4548–4557.
- [45] R. Margolin, L. Zelnik-Manor, and A. Tal, "How to evaluate foreground maps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 248–255.



SHIGANG WANG received the bachelor's degrees 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.

• • •