IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 25, 2019, accepted August 9, 2019, date of publication August 22, 2019, date of current version September 18, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2936915

Salient Object Detection Integrating Both Background and Foreground Information Based on Manifold Preserving

BAOYAN WANG[1], TIE ZHANG[1], XINGANG WANG[2], AND HAIJUAN HU[3]

[1]School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China
[2]School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China
[3]School of Mathematics and Statistics, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China

Corresponding authors: Baoyan Wang (wangbaoyan2005@163.com) and Xingang Wang (wangxingang1217@126.com)

This work was supported by the National Natural Science Foundation of China under Grant 51475086.

ABSTRACT Graph-based two-stage algorithms have widely developed and achieved good performance to detect salient objects. For these algorithms, choosing the proper seeds using for saliency propagation is quite crucial and difficult. In this paper, we consider using background/foreground probability values of candidate background/foreground seeds as the estimation of the reliable seeds, not considering the refinement of candidate seeds. Moreover, these probability values are integrated into the proposed saliency models, which can avoid hard filtering for candidate seeds as well as simplify the procedure of the algorithm. In addition, considering the manifold structure of an image, we fuse the manifold-preserving term into the saliency models. Especially, reconstruction matrix A is determined based on the deep features extracted from FCN-32s, which can further improve detection performance of salient objects. The results of experiments in which the proposed SBFMP algorithm is applied to four datasets demonstrate SBFMP algorithm is prior to some existing state-of-the-art algorithms in terms of the different metrics.

INDEX TERMS Salient object detection, saliency map, background, foreground, manifold.

I. INTRODUCTION

Visual saliency aims at identifying the most visually distinctive parts in an image, and has been an important research topic in computer vision. In recent years, salient object detection, which can serve as a pre-processing step for a variety of computer vision, such as object detection and recognition [1], [2], image classification [3], gaze estimation [4], content-based image retrieval and editing [5], [6], visual tracking [7], and person re-identification [8].

Many researchers have studied this topic and proposed many effective algorithms on salient object detection [9]–[18]. Specially, graph-based propagation algorithms have developed and achieved good performance to detect salient objects [9], [12]–[14], [16]. General steps of these algorithms are summarized as follows, after constructing corresponding undirected and weighted graph of an image, reliable seeds are selected as labeled nodes firstly. Then the saliency information of these labeled nodes is propagated to all graph nodes by using the proposed saliency models, the saliency value of each node in the graph is thus obtained. Choosing the proper seeds using for propagation is quite crucial, which will have direct influences on the final detection results. Generally, seed candidates are firstly selected based on some prior methods, such as background or center prior. Then these candidates are further refined by defining some evaluation metrics, the reliable seeds thus obtained. Although there exist some refinement methods for candidate seeds [10]–[12], it is hard to guarantee the accuracy of selected seeds. Furthermore, candidates' refinement process is separate from the procedure of establishing propagation models, which is in fact a little trivial.

In this paper, we propose a graph-based two-stage salient object detection algorithm. For the selection of candidate seeds, we simply select image boundaries as candidate background seeds based on background prior [9]. In addition, candidate foreground seeds can be obtained according to the first-stage saliency map. Instead of refining candidate seeds, we consider using background/foreground probability values of candidate background/foreground seeds as the estimation

The associate editor coordinating the review of this article and approving it for publication was Wenbing Zhao.

of the reliable seeds. Moreover, these probability values are integrated into the proposed saliency models, which can avoid hard filtering for candidate seeds, meanwhile, simplify the procedure of the algorithm.

Similar to [9], [14], our proposed propagation model of each stage is also attributed to an energy minimization problem. However, different from [9], some background and foreground information is fused in the energy minimization model. Furthermore, manifold-preserving diffusion factor is also considered. Specially, reconstruction matrix A in manifold-preserving diffusion factor is obtained based on multi-level deep features learned by FCN-32s [19] different from [14]. As our saliency detection algorithm integrates both background and foreground information based on manifold preserving, we call it SBFMP algorithm. Our contributions in this paper are summarized as follows:

1) Instead of refining candidate seeds, we use probability values of candidate seeds for further estimation of reliable seeds.

2) We build a saliency propagation model in which background and foreground information is fused and manifold-preserving diffusion factor is also considered. Specially, reconstruction matrix A in manifold-preserving diffusion factor is obtained based on multi-level deep features.

3) The results of extensive comparisons on four datasets using different evaluation metrics verify the effectiveness of SBFMP algorithm.

The remainder of the paper is organized as follows. Sec. II reviews related work on salient object detection. Sec. III describes SBFMP algorithm in details, including graph construction, estimation of reconstruction matrix A and the two-stage detection algorithm of salient objects. Results of experiments and comparisons are given in Sec. IV. Finally, the conclusion is drawn in Sec. V.

II. RELATED WORK

Existing salient object detection methods can be roughly classified into two categories: hand-crafted features-based methods and deep learning-based methods.

A. HAND-CRAFTED FEATURES-BASED METHODS

Many existing salient object detection models are established based on hand-crafted features, including contrast-based methods, graph-based methods and machine learning-based methods. Considering the general properties of salient objects in an image, some prior methods are fused with the above mentioned hand-crafted feature-based models, such as center prior, spatial distribution prior, background prior, boundary prior, objectness prior and focus prior e.g. [14], [16].

Contrast-based methods include local and global contrast-based methods. Local contrast-based saliency is computed by the difference degree between a certain pixel or area and its neighbor components on some feature space, while global contrast-based saliency is computed by the difference degree between a certain pixel or area and other all components on some feature space. Cheng *et al.* [17] considered the global region contrast with respect to the entire image and spatial relationships across the regions to extract saliency map. Wang *et al.* [18] firstly obtained objectness areas using BING algorithm [20], background information thus determined. Furthermore, combining background contrast prior, background connectivity prior and spatial distribution prior with the obtained background information, saliency map related to backgrounds was formulated.

As our SBFMP algorithm is in fact proposed based on graph methods, the detailed crucial steps of graph-based methods will be expanded. (1) **For the selection of candidate background seeds**, image boundaries from four sides: top, down, left and right are selected as candidate background seeds based on background prior in [9]. However, for certain complex images in which salient objects touch the boundaries of the images, this selection method is unreliable. Therefore, some researchers have attempted to filter the error boundaries touched by salient objects [10]–[13]. Wang *et al.* [11] removed the foreground noise in the image borders by using image edge information. Xia *et al.* [10] extracted background seeds from four side boundaries via divergence information, which is based on edge weights and center prior. Zhai *et al.* [12] filtered out one of the four boundaries that most unlikely belong to the background based on color distribution of the four boundaries. Wang *et al.* [13] used probability of boundary superpixels belonging to background computed by Mahalanobis distance of boundary superpixels to analyze the properties of boundary. **For the selection of candidate foreground seeds**, different from [9], [11], [14] in which the first-stage saliency map using the adaptive threshold is simply binarized and candidate foreground seeds are extracted by selecting the superpixels whose saliency values were larger than the threshold, foreground regions were generated by using parametric maxflow based on the consideration of the spatial compactness of salient objects in [15]. Xia *et al.* [10] introduced a regularization rare term based on [15], thus the foreground regions can be obtained by solving an optimization problem. However, Zhai *et al.* [12] determined foreground seeds by using center prior, without referring to the first-stage saliency map. (2) **Methods for establishing saliency models** of each stage are various. Wang *et al.* [11] used Euclidean color and spatial distances between each superpixel and background or foreground seeds to obtain saliency of the superpixel. Xia *et al.* [10] utilized geodesic distance between each superpixel and background or foreground seeds to establish the two-stage detection models. Wei *et al.* [16] proposed the geodesic distance between an internal node and the given virtual background node viewed as the saliency of the internal node. Based on the different types of seeds, Wang *et al.* [13] used different similarity metric matrices to obtain the saliency of each superpixel. Zhai *et al.* [12] proposed a novel saliency detection framework via multiple random walkers (MRW) method based on background and foreground seeds. After constructing the graph of an image, Zhang *et al.* [21] proposed the travel time

of the transient nodes being absorbed into absorbing nodes deemed as the saliency of the transient node. As the expanded work of [21], Zhang *et al.* [22] achieved a learnt transition probability matrix by the sparse-to-full method. In addition, angular embedding technique was used for refining the initial saliency maps to enhance the saliency of the whole salient object. Saliency detection was formulated as an energy function minimization problem in [9], [14]. Specially, manifold-preserving diffusion factor was added to the energy minimization model in [14] based on [9]. As the expanded work of [9], Zhang *et al.* [23] constructed a multi-scale graph to simultaneously capture local and global structure information of an image and proposed the three-stage cascade scheme for graph labeling. Similar to [23], our SBFMP algorithm is also proposed based on energy minimization problem. Different from [23], the graph constructed in SBFMP is single scale, not multi-scale. Qin *et al.* [24] introduced Hierarchical Cellular Automata (HCA) model consisting of two main components: Single-layer Cellular Automata (SCA) and Cuboid Cellular Automata (CCA). SCA can exploit the intrinsic relevance of similar regions on a graph. Low-level image features as well as high-level semantic information extracted from DNNs were incorporated into the SCA to measure the correlation between different image patches. Furthermore, CCA integrated multiple saliency maps generated by SCA at different scales in a Bayesian framework. (3) **The fusion of two-stage saliency maps**. The second-stage saliency map was regarded as the final saliency map in [9], [12]–[14], while the final saliency map was obtained by combining the first-stage saliency map with the second-stage saliency map via product operation in [11], [15].

Machine learning methods based on hand-crafted features are used to detect salient objects from two aspects, one is to learn weighted coefficients of different hand-crafted features using some machine learning methods, and saliency values of an image are then computed by fusing each feature with regarding corresponding coefficient as its weight; the other is to explore how to fuse various saliency maps from different methods by applying some machine learning methods, so that better detection results can be obtained. Liu *et al.* [25] proposed a set of features including the local multi-scale contrast, region center-surround histogram distance, and global color spatial distribution. The final saliency map was obtained by fusing these features according to CRF method. Meanwhile, this algorithm was also applied to salient objects detection of video sequence. Jiang *et al.* [26] firstly represented an image by a set of multi-level segmentations. For each segmentation, 93-dimensional feature vector of each area was obtained by extracting regional contrast, property and backgroundness descriptors. Furthermore, using random forest method to learn the regional saliency regressor based on the obtained regional feature vectors, saliency map for each segmentation was thus formulated. Lastly, learning weights of different saliency maps with multi-level segmentations using a least square estimator, the final saliency map was thus formed by a linear combinatory of saliency maps with multi-level segmentations. Mai *et al.* [27] proposed two methods for fusing different saliency maps. For the first method, the final saliency map was defined as a sigmoid function related to saliency maps obtained by different methods and some unknown coefficients. Herein, these unknown coefficients can be determined by using normal logistic regression method on training data. Inspired by [25], they proposed the second method by using CRF method to aggregate saliency maps obtained by multiple methods, which not only assigned the contribution of each saliency map to the final saliency map, but captured the relation between neighboring pixels. Yan *et al.* [28] integrated the saliency maps computed on the multi-level segmentations of an image into a tree-structure graph model, and the model can be solved by using belief propagation. Li *et al.* [29] proposed a salient object detection algorithm based on reconstruction errors. They firstly extracted boundary superpixels of an image as background templates. Dense and sparse reconstruction errors of each superpixel were then encoded based on these templates, respectively. Pixel-level saliency measures by dense and sparse reconstruction error propagation and refinement were thus obtained. Finally, a clean and uniform saliency map was formulated by combining the two saliency maps via dense and sparse reconstruction using the Bayesian integration method.

B. DEEP LEARNING-BSASED METHODS

Deep learning-based salient object detection models can be roughly divided into two categories: region-based approaches and fully convolutional networks (FCNs)-based approaches.

Region-based approaches use each image patch as the basic processing unit for detecting salient objects. Early deep learning-based research on salient object detection [30]–[32] is related to region-based models. Zhao *et al.* [30] presented a multi-context deep learning framework for salient object detection. They employed two different CNNs to extract global and local context information, respectively. Li and Yu [31] proposed a visual saliency model by utilizing multi-scale features extracted from deep CNNs. In addition, a complete saliency framework is developed by further integrating CNN-based saliency model with a spatial coherence model and multi-level image segmentations. Wang *et al.* [32] predicted saliency maps by integrating both local estimation and global search. Two different deep CNNs were trained to capture local information and global contrast. Region-based approaches are all operated at the patch level instead of the pixel level, and each pixel is simply assigned the saliency value of its enclosing patch. As a result, saliency maps are blurry without fine details, especially near the boundary of salient objects. Furthermore, all image patches are treated as independent data samples even when they are overlapping. Therefore, CNNs are run thousands of times to obtain the saliency values of every patch, which is very computationally expensive. To overcome the above mentioned problems, recent research on this field tends to use FCNs models trained end-to-end for detecting salient objects. Here, ''FCNs'' is

one of the most popular CNNs, while "end-to-end" means that deep network only needs to be run on the input image once to produce a complete saliency map with the same pixel resolution as the input image. Li and Yu [33] proposed an end-to-end deep contrast network which was composed of a pixel-level fully convolutional stream (FCS) and a segment-level spatial pooling stream (SPS). A fully connected CRF model can be optionally incorporated in the fused results from these two streams. Wang *et al.* [34] proposed a novel Localization-to-Refinement network for salient object detection. Global Recurrent Localization Network (RLN), which included inception-like contextual weighting module and recurrent module, focused on the location of salient objects under various scenarios. Inception-like contextual weighting module exploited contextual information by the weighted response map, while the recurrent module iteratively refined each convolutional block over time. Moreover, to effectively recover salient object boundaries, a local Boundary Refinement Network (BRN) and a refinement module were adopted to learn local context information by the propagation efficient. Li *et al.* [35] built internal semantic properties of salient objects by using multi-task networks based on FCNs, and fulfilled joint learning of the two tasks (salient object detection and semantic segmentation). Meanwhile, the graph Laplacian regularized nonlinear regression was used to refine the coarse saliency map obtained by FCNs model, which overcame the fuzzy object boundaries caused by the down-sampling operations of CNNs.

III. THE PROPOSED SBFMP ALGORITHM

As SBFMP algorithm is proposed based on graph method, the first step of SBFMP algorithm is to construct the corresponding undirected and weighted graph of an image. In addition, similar to algorithms in [10]–[16], SBFMP algorithm is also composed of two-stage salient object detection. In the first stage, we select four boundaries as candidate background seeds. Instead of further filtering candidates, background probability calculated by [36] is used for estimation of reliable background seeds, which is integrated into a saliency propagation model. Excepting for considering background information in the proposed model, the manifold-preserving factor is also fused in the propagation model. Through solving the first-stage propagation model, we obtain a closed solution which is defined as the first-stage saliency. In the second stage, partial candidate foreground seeds are generated by binary segmentation of the first-stage saliency map. The corresponding first-stage saliency is viewed as the foreground probability of each candidate foreground seed. Similar to the first-stage detection, the foreground probability is integrated into another saliency propagation model. Excepting for considering foreground information in the proposed model, the manifold-preserving factor is also added to the second-stage saliency propagation model. The second-stage saliency values are obtained by solving corresponding second-stage optimization model. Specially, reconstruction matrix A of manifold-preserving term in the propagation models is learnt by multi-level deep features extracted from FCN-32s [19]. Finally, the second-stage saliency map is viewed as the final saliency map.

The greater details of SBFMP algorithm are introduced in Sec. III A-D. The corresponding undirected and weighted graph construction for an image is described in Sec. III A. Reconstruction matrix A used for constructing manifold-preserving factor of propagation models is estimated in Sec. III B. The method of the first-stage detection for salient objects is proposed in Sec. III C. In addition, the second-stage detection for salient objects is described in Sec. III D.

A. GRAPH CONSTRUCTION

First, we segment an image into superpixels by SLIC method [37] and suppose that the number of superpixels is N. Each superpixel, denoted as v_i, $(i = 1, 2 \ldots, N)$, is viewed as a node. The node set V is composed of node v_i. As neighboring nodes are likely to have a similar appearance and saliency values, we use a k-regular graph to reveal the spatial relations of the nodes. In this paper, $k = 2$ is selected as the number of neighboring layers, that is, each node is connected with its neighboring nodes. Each node is also connected with the neighboring nodes of these nodes. In addition, the nodes of the outmost layer from four sides, namely, the top, bottom, left, and right, are selected as boundary nodes, which are connected with each other since they are very likely to belong to the same background regions. Suppose neighboring nodes v_i and v_j are connected with edge e_{ij}, the edge set E is thus composed of edges $e_{ij}(i,j = 1,2,\ldots, N)$. The weight of the edges e_{ij} can be defined as

$$w_{ij} = e^{-\frac{\|c_i - c_j\|}{\sigma^2}} \tag{1}$$

where c_i and c_j denote the mean color value of nodes v_i and v_j in the CIELAB space, respectively. σ is a constant used to control the strength of the weights. Matrix $W = [w_{ij}]_{N\times N}$ $(w_{ii} = 0)$ is defined as an affinity matrix. Thus, an undirected and weighted graph $G = (V, E)$ can be formed with node set V and edge set E.

For the undirected and weighted graph $G = (V, E)$, the degree of node v_i is defined as

$$d_i = \sum_j w_{ij} \tag{2}$$

Degree matrix D is a diagonal matrix, the diagonal entries of which are $d_i(i = 1, 2\ldots, N)$.

B. ESTIMATION OF RECONSTRUCTION MATRIX A

Similar to [14], supposing that manifold structure of an image on feature space (high-dimensional space) is similarly preserved on saliency space (low-dimensional space), we introduce reconstruction matrix A into the proposed models. Inspired by locally linear embedding (LLE) [38], matrix A

is formulated by minimizing the overall as follows,

$$\arg\min_{a_{ij}} \sum_{i=1}^{N} \left\{ \left\| f_i - \sum_{j,j\sim i} a_{ij} f_j \right\|_2^2 + \lambda \sum_{j,j\sim i} a_{ij}^2 \right\}$$
$$s.t \sum_{j,j\sim i} a_{ij} = 1, \quad a_{ii} = 0, \ \forall i \triangleq j, \ a_{ij} = 0 \tag{3}$$

where $j\sim i$ denotes the neighboring label of the node v_i and $i \triangleq j$ shows that v_i and v_j are not adjacent each other. λ is a small number for regularization which can guarantee unique solution of (3). N is the number of nodes. Condition $\sum_{j,j\sim i} a_{ij} = 1$ can ensure the reconstruction is linear and shift-invariant [38].

In (3), f_i is the feature vector of the node v_i, which is obtained by adopting different hierarchies of deep features extracted from FCN-32s [19]. As is well-known, deep features are highly versatile and have stronger presentational power than traditional handcrafted features [39]. Although the features in the last layers of CNNs encode semantic abstractions of objects, they cannot precisely capture the fine-grained low-level information, such as color, edge, and texture, due to their low spatial resolution. Therefore, we use low-level and high-level deep cues as the feature vector. Inspired by [24], f_i is determined based on features from the final pooling layers of Conv1 and Conv5 in FCN-32s. Specially, feature dimension of the final pooling layers of Conv1 and Conv5 are 64 and 512, respectively, by concatenating these features from multiple levels, the feature vector f_i of the superpixel i thus is obtained.

Supposing that the number of neighboring nodes for the node v_i is $\tau(i)$, $j\sim i$ denotes the neighboring label of the node v_i, $j = n_1, n_2, \ldots n_\tau$, $(n_1 < n_2 < \ldots < n_\tau)$. Through these definitions such as,

$$a_{ij} = \begin{cases} 0, i = j \\ 0, i \triangleq j, \\ b_{ij}, j = n_1, n_2, \cdots, n_\tau, j \sim i, \end{cases}, b_i = (b_{i,n_1}, b_{i,n_2}, \cdots, b_{i,n_\tau})^T,$$
$$H_i = (f_i - f_{n_1}, f_i - f_{n_2}, \cdots, f_i - f_{n_\tau}), \quad Z_i = H_i^T H_i,$$

Equation (3) is also described with vectors as,

$$\arg\min_{b_i} \sum_{i=1}^{N} \left\{ b_i^T Z_i b_i + \lambda b_i^T b_i \right\}$$
$$s.t \ b_i^T 1_\tau = 1, \quad i = 1, 2, \cdots, n \tag{4}$$

where 1_τ is an all-one vector. Equation (4) is solved by using Lagrange Multiplier method, and its solution is described as follows,

$$b_i = Norm\left[(Z_i + \lambda I)^{-1} \cdot 1_\tau \right], i = 1, 2, \cdots, N \tag{5}$$

where I denotes τ dimensional identity matrix, and $Norm(\cdot)$ represents the normalized function. It is worth noting that a graph is commonly constructed by using k-NN method in [38], the number of nonzero elements contained in each row vector of matrix A is identical. However, the graph obtained by using k-regular method in this paper determines that the number of nonzero elements for each row vector of matrix A is different.

C. THE FIRST-STAGE DETECTION FOR SALIENT OBJECTS

In the first-stage detection, candidate background seed information is used to establish salient object detection model. For SBFMP algorithm, selecting simply four boundaries of an image as the first-stage candidate background seeds without further considering the refinement of candidates is because background probability as the estimation of reliable background seeds would be included in the following propagation model.

Relative to candidate background seeds, $Y^b = diag(y_1^b, y_2^b, \cdots, y_N^b)$ is the corresponding index diagonal matrix of the node set V, and it is defined as following,

$$y_i^b = \begin{cases} 1, & v_i \in BG \\ 0, & v_i \notin BG \end{cases} \tag{6}$$

where BG denotes boundary superpixel set of an image. For the selected candidate background seeds, using boundary connectivity method proposed by [36] to calculate the background probability p_i^{bg} of the node v_i, and p_i^{bg} is defined as follows, where σ_{bndCon} is empirically set as 1.

$$p_i^{bg} = 1 - \exp(-BndCon^2(v_i) \Big/ 2\sigma_{bndCon}^2) \tag{7}$$

$BndCon(v_i)$ is the boundary connectivity of the node v_i and its detailed calculation is explained in [36].

As the first-stage detection is related to candidate background seeds, the proposed first-stage model contains some background information. Meanwhile, considering smoothness and manifold-preserving terms, the first-stage detection model is described as follows,

$$\min_s \sum_{i=1}^{N} \sum_{j,j\sim i} \frac{1}{2} w_{ij} (s_i - s_j)^2 + \sum_{i=1}^{N} (s_i - \sum_{j\sim i} a_{ij} s_j)^2$$
$$+ \sum_{i=1}^{N} p_i^{bg} y_i^b s_i^2 + \sum_{i=1}^{N} (1 - p_i^{bg})(1 - y_i^b)(s_i - 1)^2 \tag{8}$$

where p_i^{bg} is determined by (7). w_{ij} and a_{ij} are obtained by solving (1) and (3), respectively. y_i^b is assigned by (6). The solution s of (8) represents the first-stage saliency.

Now, we analyze the meanings of (8). In (8), the first term is the penalty to enforce the smoothness assumption so that nodes connected by large weights w_{ij} should have similar saliency values. The second term is the penalty to enforce the local reconstruction assumption where manifold structure of an image on feature space (high-dimensional space) is similarly preserved on saliency space (low-dimensional space). The third and fourth terms in (8) embody background and saliency information of graph nodes. If superpixel i is a candidate background seed, corresponding index value y_i^b is set as 1. As the value of the fourth term in (8) is zero, the fourth

term is ignored and the third term is remained in (8). If background probability p_i^{bg} is larger for the superpixel i, the energy function of (8) punishes the saliency of the superpixel i to be smaller, which is obviously reasonable. On the contrary, if the computed background probability p_i^{bg} is smaller for the superpixel i, that is, the selected superpixel i is likely to be a non-background superpixel, which implies the saliency value of the superpixel i can be a little larger. Nevertheless, the conclusion can exactly be drawn by analyzing the proposed saliency model of (8). Similarly, if superpixel i is a non-background seed, corresponding index value y_i^b is set as 0. As the value of the third term in (8) is zero, the third term is ignored and the fourth term is remained in (8). If background probability p_i^{bg} is smaller for the superpixel i, the energy function of (8) punishes the saliency of the superpixel i to be larger, which is obviously reasonable. On the contrary, if the computed background probability p_i^{bg} is larger for the superpixel i, that is, the selected superpixel i is likely to be a background superpixel, which implies the saliency value of the superpixel i can be a little smaller. Nevertheless, the conclusion can exactly be drawn by analyzing the proposed saliency model of (8). The above analysis results show that the first-stage saliency model has implied the idea of filtering candidate background seeds, which is different from some methods on filtering boundaries [10]–[12], [35]–[37].

Equation (8) can be rewritten equivalently in a matrix form as,

$$\min_s s^T \left[W^{bg}Y^b + B + D - W + (I-A)^T(I-A)\right] s - 2bs + \Phi^T B\Phi \quad (9)$$

where $W^{bg} = diag(p_1^{bg}, p_2^{bg}, \cdots, p_N^{bg})$, $B = (I - W^{bg})(I - Y^b)$, Φ is an N-dimension and all-one column vector, $b = \Phi^T B$, and I denotes N-dimension unit matrix. By taking the derivative on (9) and setting it equal to zero, the closed solution of (9) is obtained as,

$$s^b = \left[W^{bg}Y^b + B + D - W + (I-A)^T(I-A)\right]^{-1} b^T \quad (10)$$

where we denote s^b as the saliency values obtained by background seeds in order to distinguish the saliency values obtained by foreground seeds. Similar to [9], [14], Equation (10) can also be unified as the form $s = Ky$, yet there exist two different points from [9], [14]: firstly, some background information of an image is reflected in matrix K as (10), while no background information is reflected in K as [9], [14] when background seeds are selected as queries. Secondly, the model determined by (8) should ensure that the first-stage saliency values can directly be computed by (10). However, in [9], [14], the first-stage saliency values are indirectly obtained by computing $s = Ky$.

D. THE SECOND-STAGE DETECTION FOR SALIENT OBJECTS

As the final saliency map is the second-stage saliency map, we firstly introduce the second-stage saliency detection in this section. To capture the candidate foreground seeds, the first-stage saliency map is binary segmented using the mean value θ of the first-stage saliency map. The nodes that are larger than threshold value θ are viewed as candidate foreground seeds. We suppose that the set of candidate foreground seeds is FG. Relative to these foreground candidates, $Y^f = diag(y_1^f, y_2^f, \cdots, y_N^f)$ is the corresponding index diagonal matrix for the node set V, and it is defined as,

$$y_i^f = \begin{cases} 1, & v_i \in FG \\ 0, & v_i \notin FG \end{cases} \quad (11)$$

Based on the foreground seed candidates and their corresponding saliency values in the first-stage detection, foreground probability p_i^{fg} of the node v_i is defined as,

$$p_i^{fg} = s_i^b \quad (12)$$

Similar to the first-stage model, the second-stage model contains some foreground information. Meanwhile, considering smoothness and manifold-preserving terms, the second-stage detection model is described as follows,

$$\min_s \sum_{i=1}^N \sum_{j,j\sim i} \frac{1}{2} w_{ij}(s_i - s_j)^2 + \sum_{i=1}^N (s_i - \sum_{j\sim i} a_{ij}s_j)^2 + \sum_{i=1}^N p_i^{fg} y_i^f (s_i - 1)^2 + \sum_{i=1}^N (1 - p_i^{fg})(1 - y_i^f) s_i^2 \quad (13)$$

where w_{ij} and a_{ij} are obtained by solving (1) and (3), respectively. y_i^f is assigned by (11). p_i^{fg} is determined by (12). Now, we analyze the meanings of (13). As the meanings of the first and second terms (smoothness and manifold-preserving terms) in (13) are the same as (8), we only analyze the third and fourth terms in (13) embodying foreground information and saliency of graph nodes. If superpixel i is a foreground seed candidate, corresponding index value y_i^f is set as 1. As the value of the fourth term in (13) is zero, the fourth term is ignored and the third term is remained in (13). If foreground probability p_i^{fg} is larger for the superpixel i, the energy function of (13) punishes the saliency of the superpixel i to be larger, which is obviously reasonable. On the contrary, if the computed foreground probability p_i^{fg} is smaller for the superpixel i, that is, the selected superpixel i is likely to be a non-foreground superpixel, which implies that the saliency value of the superpixel i can be a little smaller. Nevertheless, the conclusion can be exactly drawn by analyzing the proposed saliency model of (13). Similarly, if superpixel i is a non-foreground seed, corresponding index value y_i^f is set as 0. As the value of the third term in (13) is zero, the third term is ignored and the fourth term is remained in (13). If foreground probability p_i^{fg} is smaller for the superpixel i, the energy function of (13) punishes the saliency of the superpixel i to be smaller, which is obviously reasonable. On the contrary, if the computed foreground probability p_i^{fg} is larger for the superpixel i, that is, the selected superpixel i is likely to be a foreground superpixel, which implies that the saliency

value of the superpixel i can be a little larger. Nevertheless, the conclusion can exactly be drawn by analyzing the proposed saliency model of (13). The above analysis results show that the second-stage saliency model has implied the idea of filtering candidate foreground seeds obtained by segmenting the first-stage saliency map.

Equation (13) can be rewritten equivalently in a matrix form as,

$$\min_s s^T\Big[(I-W^{fg})(I-Y^f)+F+D-W+(I-A)^T(I-A)\Big]s - 2fs+\Phi^T F\Phi \quad (14)$$

where $W^{fg}=diag(p_1^{fg},p_2^{fg},\cdots,p_N^{fg})$, $F=W^{fg}Y^f$, Φ is an N-dimension and all-one column vector,$f=\Phi^T F$ and I denotes N-dimension unit matrix. By taking the derivative on (14) and setting it equal to zero, the closed solution of (14) is obtained as,

$$s^f=\Big[(I-W^{fg})(I-Y^f)+F+D-W+(I-A)^T(I-A)\Big]^{-1}f^T \quad (15)$$

where we denote s^f, the second-stage saliency, as the saliency values obtained by foreground seeds in order to distinguish the saliency values obtained by background seeds. Herein, the second-stage saliency map formulated by the second-stage saliency is in fact the final saliency map.

IV. EXPERIMENTAL RESULTS

The proposed SBFMP algorithm is evaluated on four datasets. The first dataset is the MSRA-10K dataset [43], which contains 10,000 images. The second one is the DUT-OMRON dataset [17], which contains 5,168 images. The third dataset is the ECSSD dataset [44], which contains 1,000 images. The fourth dataset is the PASCAL-S dataset [45], which contains 850 natural images. The accurate salient objects of the four datasets are marked manually. We compares the performance of the proposed SBFMP algorithm with that of the 19 state-of-the-art methods on the four datasets, namely, CA [46], DSR [29], FES [47], MC [21], RBD [36], COV [48], SEG [49], SIM [50], SR [51], SS [52], SWD [53], MR [9], SF [54], GS [16], HS [44], MPDS [14], Neuro [55], the algorithm proposed by [56] and MILPS [57]. Now, we briefly introduce these comparison methods. Goferman *et al*. [46] proposed a new type of saliency-context-aware saliency which was based on four principles observed in the psychological literature. Tavakoli *et al*. [47] proposed a new saliency detection algorithm based on estimating saliency of local feature contrast in a Bayesian framework. Zhu *et al*. [36] proposed a robust background measure called boundary connectivity and combined this metric with the contrast to build a contrast function weighted by the background coefficients. Erdem and Erdem [48] used covariance matrices of simple image features as meta-features for saliency estimation. Rahtu *et al*. [49] proposed a saliency measure formulated by using a probabilistic framework, where different features were fused together in joint distributions. A saliency model was obtained by generalizing an efficient model of color appearance including a principled selection of parameters as well as an innate spatial pooling mechanism in [50]. By analyzing the log-spectrum of an input image, Hou and Zhang [51] proposed a fast saliency detection method based on the spectral residual of image. Hou *et al*. [52] used the image signature as a descriptor of natural scenes to highlight sparse salient regions. Duan *et al*. [53] defined the saliency of each patch drawn from an image through three elements: dissimilarity, spatial distance and central bias. Yang *et al*. [9] proposed a two-stage saliency detection algorithm by using graph-based manifold ranking. Perazzi *et al*. [54] firstly decomposed an image into perceptually homogeneous elements and then derived saliency estimations from two well-defined contrast measures based on the uniqueness and spatial distribution of those elements. Shi *et al*. [44] constructed multi-scale saliency maps based on local contrast location heuristic and obtained the final results by feeding these saliency maps into a hierarchical inference model. Fu *et al*. [14] built a saliency propagation model based on the manifold smoothness and the local reconstruction assumptions and proposed two-stage detection scheme in which boundary prior, Harris convex hull and foci convex hull were integrated. Fu *et al*. [55] applied Ncut to a superpixel graph to obtain hierarchical spectral segments. For these segments, three effective cues were incorporated to estimate the regional saliency of each segment. Multi-layer saliency maps were formed by propagating the saliency of foreground seeds obtained using contrast prior and center prior and background seeds obtained using boundary prior in [56]. Huang *et al*. [57] took the proposals as the bags of instances of multiple instances learning (MIL) and formulated saliency detection problem as an MIL task. In addition, DSR [29], MC [21] and GS [16] algorithms have been briefly introduced in Sec. II.

A. PARAMETER SETTINGS AND EVALUATION METRICS

The number N of superpixels using SLIC method in all the experiments is 300. Parameter σ^2 determined using (1) is equal to 0.1 to construct affinity matrix W. While estimating reconstruction matrix A, we set $\lambda=1\times10^{-4}$ in (4).

The above mentioned salient object detection algorithms are evaluated with the help of the precision, recall, F-measure, and the mean absolute error (MAE) metrics. Here, binary mask Q is formed based on the predicted saliency map S normalized between 0 and 255. G means the binary mask of the ground-truth of the salient objects. In terms of the binary mask, $|\cdot|$ indicates the number of non-zero entries in the mask.

$$precision=\frac{|Q\cap G|}{|Q|},\quad recall=\frac{|Q\cap G|}{|G|} \quad (16)$$

As quality of a saliency map can not be well evaluated only using the precision and the recall values, the F-measure is assumed as another evaluation metric according to a weighted

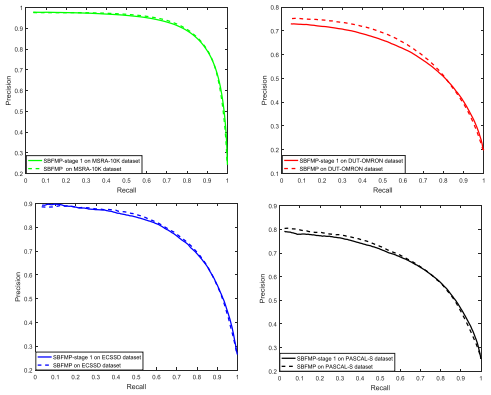


FIGURE 1. P-R curves generated by each stage detection on four datasets.

harmonic of the precision and recall,

$$F_\beta = \frac{(1+\beta^2)precision \times recall}{\beta^2 precision + recall} \tag{17}$$

where $\beta^2 = 0.3$ can emphasize the importance of the precision [58].

In addition, the MAE metric is also adopted to evaluate the algorithm proposed in this paper. Here, S and G are normalized within the range of $[0, 1]$ as $\bar{S}$ and $\bar{G}$, respectively. The MAE metric can be defined as,

$$MAE = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} \left| \bar{S}(x, y) - \bar{G}(x, y) \right| \tag{18}$$

where W and H are the width and height of the saliency map.

B. VALIDATION OF THE PROPOSED SBFMP ALGORITHM

To verify the effectiveness of the proposed two-stage detection, we compare the performance of the first-stage detection and the second-stage detection on the mentioned four datasets. For convenience, the algorithm only containing the first-stage detection is remarked as SBFMP-stage1. Results in Fig.1 show that the performance of the second stage is better than that of the first stage on four datasets by analyzing the P-R curves.

Except for considering background and foreground information in the propagation models, manifold-preserving term is also fused into our proposed models. To directly show the effects of the manifold-preserving term on saliency models, we compare the performance with and without manifold-preserving term on four datasets. For convenience, the algorithm removing manifold-preserving terms of (8) and (13) is remarked as SBFMP-wm. Results in Fig.2 show that SBFMP algorithm achieves better performance than SBFMP-wm algorithm on four datasets by analyzing the P-R curves, which demonstrates that the fusion of manifold-preserving term into the saliency models is efficient.

C. COMPARISONS WITH OTHER STATE-OF-THE-ART ALGORITHM

The proposed SBFMP algorithm is compared with the 19 state-of-the-art algorithms mentioned above, on the four

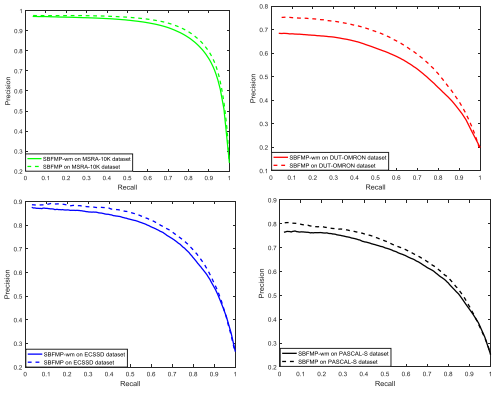


FIGURE 2. P-R curves generated by without and with manifold-preserving term on four datasets.

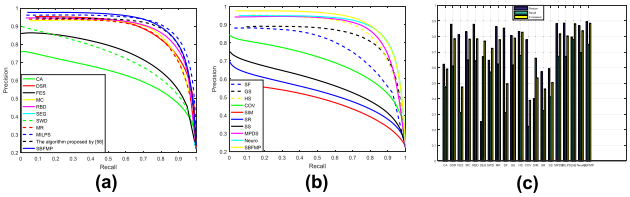


FIGURE 3. Comparison results of SBFMP algorithm and other 19 algorithms on MSRA-10K dataset. (a) P-R curves of 11 algorithms, (b) P-R curves of 10 algorithms, and (c) precision, recall and F-measure metrics of 20 algorithms.

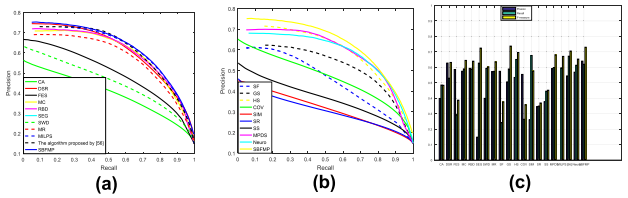


FIGURE 4. Comparison results of SBFMP algorithm and other 19 algorithms on DUT-OMRON dataset. (a) P-R curves of 11 algorithms, (b) P-R curves of 10 algorithms, and (c) precision, recall and F-measure metrics of 20 algorithms.

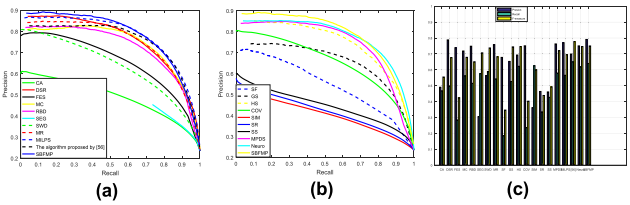


FIGURE 5. Comparison results of SBFMP algorithm and other 19 algorithms on ECSSD dataset. (a) P-R curves of 11 algorithms, (b) P-R curves of 10 algorithms, and (c) precision, recall and F-measure metrics of 20 algorithms.

datasets in terms of various evaluation measures. It can be seen from the P-R curves in Fig.3. (a)-(b), Fig.4. (a)-(b), Fig.5. (a)-(b), Fig.6. (a)-(b) that the SBFMP algorithm is prior to the state-of-the-art algorithms on the four datasets. In addition, the precision, recall, and F-measure metrics on the four datasets are shown in Fig.3. (c), Fig.4. (c), Fig.5. (c), Fig.6. (c). It is worth to noting that F-measure score still

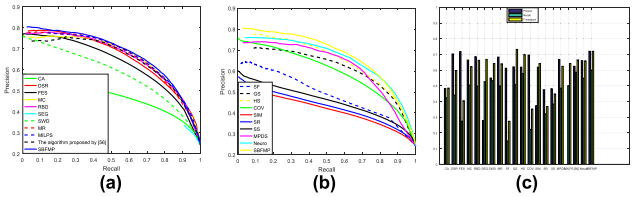


FIGURE 6. Comparison results of SBFMP algorithm and other 19 algorithms on PASCAL-S dataset. (a) P-R curves of 11 algorithms, (b) P-R curves of 10 algorithms, and (c) precision, recall and F-measure metrics of 20 algorithms.

performs commensurate or even worse than some methods on DUT-OMRON and ECSSD datasets. On DUT-OMRON dataset, even though F-measure score of SBFMP algorithm is commensurate with the SEG algorithm and is worse than GS algorithm, precision score of SEG algorithm is too poor and precision and recall scores of GS perform worse than SBFMP. On ECSSD dataset, even though F-measure score of SBFMP algorithm is commensurate with the GS, HS, Neuro algorithms and is worse than the algorithm in [56], precision score of the algorithm in [56] is a bit low and precision and recall scores of GS, HS and Neuro algorithms perform worse than SBFMP. Therefore, we think SBFMP is prior to other methods through overall evaluations on four datasets. The MAE values of these algorithms on the four datasets are listed in Table 1. Although SBFMP algorithm does not achieve the top rank by analyzing these MAE values, it achieves a high level of performance compared to the other algorithms. A few saliency maps on the four datasets are illustrated in Fig.7.

TABLE 1. Mean absolute error results of 20 algorithms on the four datasets.

Models	MSRA-10K	DUT-OMRON	ECSSD	PASCAL-S
CA[46]	0.237	0.255	0.310	0.300
DSR[29]	0.121	**0.139**	0.171	0.205
FES[47]	0.185	0.156	0.212	0.219
MC[21]	0.145	0.186	0.202	0.229
RBD[36]	0.111	0.146	0.171	**0.200**
COV[48]	0.197	0.156	0.215	0.227
SEG[49]	0.314	0.336	0.340	0.349
SIM[50]	0.388	0.429	0.433	0.414
SR[51]	0.249	0.246	0.311	0.291
SS[52]	0.317	0.277	0.343	0.315
SWD[53]	0.267	0.311	0.318	0.315
MR[9]	0.135	0.191	0.190	0.221
SF[54]	0.171	0.147	0.219	0.237
GS[16]	0.140	0.175	**0.164**	0.222
Neuro[55]	0.110	0.176	0.171	0.213
[56]	0.136	0.193	0.204	0.241
MILPS[57]	0.110	0.168	0.177	0.218
HS[44]	0.149	0.227	0.228	0.263
MPDS[14]	**0.107**	0.165	0.171	0.215
SBFMP	**0.108**	**0.139**	**0.168**	**0.194**

We find that SBFMP algorithm can suppress backgrounds as well as highlight the salient objects well.

V. CONCLUSION

Graph-based two-stage salient object detection algorithm is proposed in this paper. In the first stage, as not all boundary superpixels coming from four sides for an image are reliable background seeds, background information is integrated into the first-stage saliency model to improve the robustness of

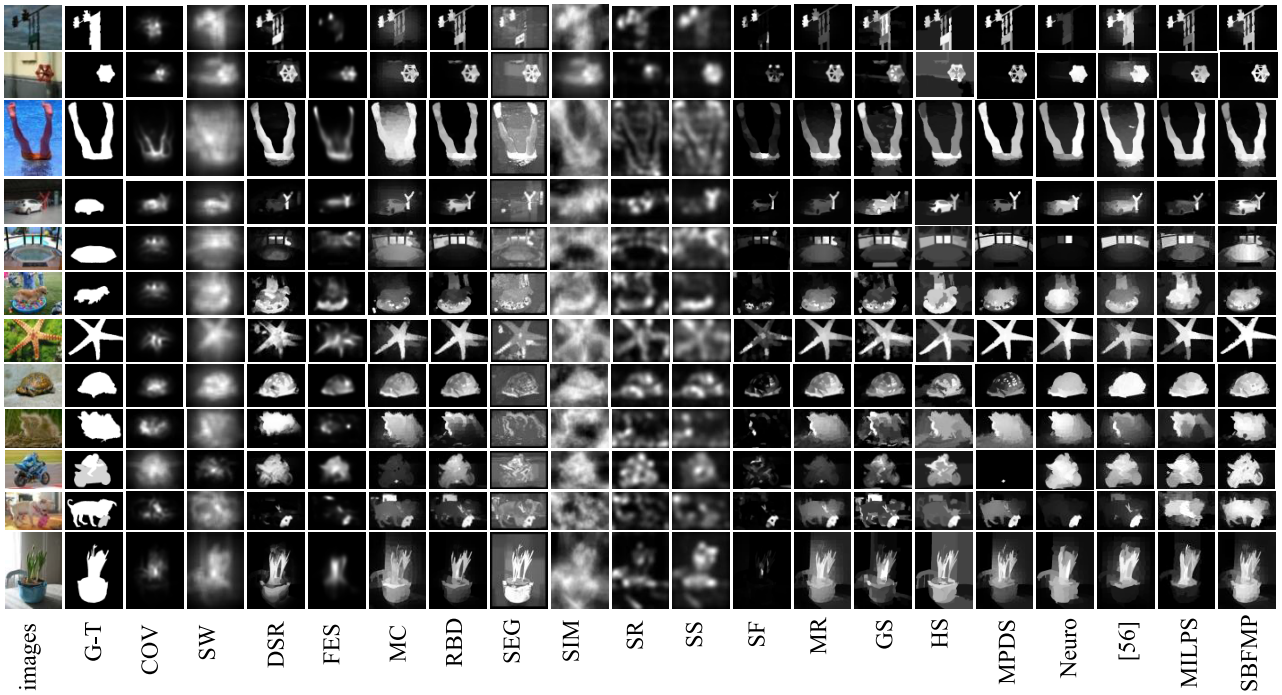


FIGURE 7. Saliency detection results of different methods. The images in rows 1–3, 4–6, 7–9, and 10–12 are from the MSRA-10K, DUT-OMRON, ECSSD, and PASCAL-S datasets, respectively.

the proposed algorithm. In the second stage, as not all foreground seeds obtained by segmenting the first-stage saliency map are flawless, foreground information is fused into the second-stage saliency model. Meanwhile, considering the manifold structure of an image, manifold-preserving term is also added to the two-stage saliency models, respectively. Specially, reconstruction matrix A is learnt by using multi-level features extracted from FCN-32s, which is more reasonable than only using color features for the estimation of A. Therefore, the two-stage saliency models for SBFMP algorithm include background and foreground information as well as fully consider manifold-preserving property of an image. The experimental results demonstrate that SBFMP algorithm can achieve satisfactory detection results compared with the state-of-the-art algorithms in terms of different metrics.

REFERENCES

[1] H. Shen, S. Li, C. Zhu, H. Chang, and J. Zhang, "Moving object detection in aerial video based on spatiotemporal saliency," *Chin. J. Aeronaut.*, vol. 26, no. 5, pp. 1211–1217, Oct. 2013.

[2] Z. Ren, S. Gao, L.-T. Chia, and I. W.-H. Tsang, "Region-based saliency detection and its application in object recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 5, pp. 769–779, May 2013.

[3] R. Wu, Y. Yu, and W. Wang, "SCaLE: Supervised and cascaded Laplacian eigenmaps for visual object recognition based on nearest neighbors," in *Proc. IEEE CVPR*, Portland, OR, USA, Jun. 2013, pp. 868–874.

[4] Y. Sugano, Y. Matsushita, and Y. Sato, "Appearance-based gaze estimation using visual saliency," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 329–341, Feb. 2013.

[5] L. Li, S. Jiang, Z.-J. Zha, Z. Wu, and Q. Huang, "Partial-duplicate image retrieval via saliency-guided visual matching," *IEEE MultimediaMag.*, vol. 20, no. 3, pp. 13–23, Jul. 2013.

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

[7] D. A. Klein, D. Schulz, S. Frintrop, and A. B. Cremers, "Adaptive real-time video-tracking for arbitrary objects," in *Proc. IEEE/RSJ IROS*, Taipei, Taiwan, Oct. 2010, pp. 772–777.

[8] S. Bi, G. Li, and Y. Yu, "Person re-identification using multiple experts with random subspaces," *J. Image Graph.*, vol. 2, no. 2, pp. 381–392, 2014.

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

[10] C. Xia, H. Zhang, and K. Li, "A novel graph structure for salient object detection based on divergence background and compact foreground," *J. Vis. Commun. Image Represent.*, vol. 48, pp. 432–441, Nov. 2017.

[11] J. Wang, H. Lu, X. Li, N. Tong, and W. Liu, "Saliency detection via background and foreground seed selection," *Neurocomputing*, vol. 152, pp. 359–368, Mar. 2015.

[12] J. Zhai, J. Zhou, Y. Ren, and Z. Wang, "Salient object detection via multiple random walks," *KSII Trans. Internet Inf. Syst.*, vol. 10, no. 4, pp. 1712–1731, Apr. 2016.

[13] B. Wang, T. Zhang, and X. Wang, "Salient object detection based on Laplacian similarity metrics," *Vis. Comput.*, vol. 34, no. 5, pp. 645–658, May 2018.

[14] K. Fu, I. Y. H. Gu, C. Gong, and J. Yang, "Robust manifold-preserving diffusion-based saliency detection by adaptive weight construction," *Neurocomputing*, vol. 175, pp. 336–347, Jan. 2016.

[15] Z. Wang, G. Xu, Z. Wang, and C. Zhu, "Saliency detection integrating both background and foreground information," *Neurocomputing*, vol. 216, pp. 468–477, Dec. 2016.

[16] W. Yichen, W. Fang, Z. Wangjiang, and S. Jian, "Geodesic saliency using background priors," in *Proc. 12th Eur. Conf. Comput. Vis.*, Florence, Italy, 2012, pp. 29–42.

[17] 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.

[18] G. Wang, Y. Zhang, and J. Li, "High-level background prior based salient object detection," *J. Vis. Commun. Image Represent.*, vol. 48, pp. 432–441, Oct. 2017.

[19] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE CVPR*, Boston, MA, USA, Jun. 2015, pp. 3431–3440.

[20] M.-M. Cheng, Z. Zhang, W.-Y. Lin, and P. Torr, "BING: Binarized normed gradients for objectness estimation at 300 fps," in *Proc. IEEE CVPR*, Columbus, OH, USA, Jun. 2014, pp. 3286–3293.

[21] L. H. Zhang, H. C. Lu, C. Yang, and M.-H. Yang, "Saliency detection via absorbing Markov chain," in *Proc. IEEE CVPR*, Sydney, NSW, Australia, Dec. 2013, pp. 1665–1672.

[22] L. Zhang, J. Ai, B. Jiang, H. Lu, and X. Li, "Saliency detection via absorbing Markov chain with learnt transition probability," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 987–998, Feb. 2018.

[23] L. Zhang, C. Yang, H. Lu, and M. Yang, "Ranking saliency," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 9, pp. 1892–1903, Dec. 2017.

[24] Y. Qin, M. Feng, H. Lu, and G. W. Cottrell, "Hierarchical cellular automata for visual saliency," *Int. J. Comput. Vis.*, vol. 126, no. 7, pp. 751–770, 2018.

[25] T. Liu, J. Sun, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," in *Proc. IEEE CVPR*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8.

[26] 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 CVPR*, Portland, OR, USA, Jun. 2013, pp. 2083–2090.

[27] L. Mai, Y. Niu, and F. Liu, "Saliency aggregation: A data-driven approach," in *Proc. IEEE CVPR*, Portland, OR, USA, Jun. 2013, pp. 1131–1138.

[28] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proc. IEEE CVPR*, Portland, OR, USA, Jun. 2013, pp. 1155–1162.

[29] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *Proc. ICCV*, Sydney, NSW, Australia, Dec. 2013, pp. 2976–2983.

[30] R. Zhao, W. Ouyang, H. Li, and X. Wang, "Saliency detection by multi-context deep learning," in *Proc. IEEE CVPR*, Boston, MA, USA, Jun. 2015, pp. 1265–1274.

[31] G. Li and Y. Yu, "Visual saliency based on multiscale deep features," in *Proc. IEEE CVPR*, Boston, MA, USA, Jun. 2015, pp. 5455–5463.

[32] L. Wang, H. Lu, X. Ruan, and M.-H. Yang, "Deep networks for saliency detection via local estimation and global search," in *Proc. IEEE CVPR*, Boston, MA, USA, Jun. 2015, pp. 3183–3192.

[33] G. Li and Y. Yu, "Deep contrast learning for salient object detection," in *Proc. IEEE CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 478–487.

[34] T. Wang, L. Zhang, S. Wang, H. Lu, G. Yang, X. Ruan, and A. Borji, "Detect globally, refine locally: A novel approach to saliency detection," in *Proc. IEEE CVPR*, Salt Lake City, UT, USA, Jun. 2018, pp. 3127–3135.

[35] X. Li, L. Wei, M.-H. Yang, F. Wu, Y. Zhuang, H. Ling, and J. Wang, "DeepSaliency: Multi-task deep neural network model for salient object detection," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3919–3930, Aug. 2016.

[36] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proc. IEEE CVPR*, Columbus, OH, USA, Jun. 2014, pp. 2814–2821.

[37] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[38] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[39] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013.

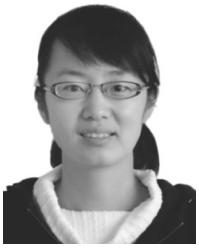
[40] Y. Qin, H. Lu, and Y. Xu, "Saliency detection via cellular automata," in *Proc. IEEE CVPR*, Boston, MA, USA, Jun. 2015, pp. 110–119.

[41] C. Li, Y. Yuan, W. Cai, Y. Xia, and D. D. Feng, "Robust saliency detection via regularized random walks ranking," in *Proc. IEEE CVPR*, Boston, MA, USA, Jun. 2015, pp. 2710–2717.

[42] Y.-W. Jiang, L.-Y. Tan, and S.-J. Wang, "Saliency detection model based on selective edges prior," *J. Electron. Inf. Technol.*, vol. 37, no. 1, pp. 130–136, Jan. 2015.

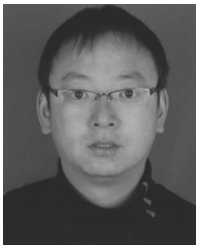
[43] Mingming. *MSRA10K Salient Object Database*. Accessed: Jul. 28, 2014. [Online]. Available: http://mmcheng.net/msra10k/

[44] 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.
[45] Y. Li, X. Hou, C. Koch, J. Rehg, and A. Yuille, "The secrets of salient object segmentation," in *Proc. IEEE CVPR*, Columbus, OH, USA, Jun. 2014, pp. 4321–4328.
[46] 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.
[47] H. R. Tavakoli, E. Rahtu, and J. Heikkilä, "Fast and efficient saliency detection using sparse sampling and kernel density estimation," in *Proc. SCIA*, Ystad, Sweden, vol. 6688. 2011, pp. 666–675.
[48] E. Erdem and A. Erdem, "Visual saliency estimation by nonlinearly integrating features using region covariances," *J. Vis.*, vol. 13, no. 4, p. 11, Mar. 2013.
[49] E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä, "Segmenting salient objects from images and videos," in *Proc. ECCV*, Crete, Greece, 2010, pp. 366–379.
[50] N. Murray, M. Vanrell, and X. Otazu, "Saliency estimation using a non-parametric low-level vision model," in *Proc. IEEE CVPR*, Colorado Springs, CO, USA, Jun. 2011, pp. 433–440.
[51] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE CVPR*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8.
[52] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 194–201, Jan. 2012.
[53] L. Duan and C. J. Wu Miao, "Visual saliency detection by spatially weighted dissimilarity," in *Proc. IEEE CVPR*, Colorado Springs, CO, USA, Jun. 2011, pp. 473–480.
[54] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE CVPR*, Providence, RI, USA, Jun. 2012, pp. 733–740.
[55] K. Fu, I. Y.-H. Gu, and J. Yang, "Spectral salient object detection," *Neurocomputing*, pp. 1–6, 2017.
[56] I. Hwang, S. H. Lee, J. S. Park, and N. I. Cho, "Saliency detection based on seed propagation in a multilayer graph," *Multimedia Tools Appl.*, vol. 76, no. 2, pp. 2111–2129, Jan. 2017.
[57] F. Huang, J. Qi, H. Lu, L. Zhang, and X. Ruan, "Salient object detection via multiple instance learning," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1911–1922, Apr. 2017.
[58] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE CVPR*, Miami, FL, USA, Jun. 2009, pp. 1597–1604.



BAOYAN WANG received the master's degree from Northeastern University, China, in 2008, where she is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, under the supervision of Prof. T. Zhang. Her research interests include image processing, computer vision, and machine learning.

TIE ZHANG received the Ph.D. degree from the Mathematics School, Jilin University, China, in 1995. He is currently a Professor with the College of Sciences, Northeastern University. His major research interests include numerical solution of partial differential equations and image processing.



XINGANG WANG received the Ph.D. degree from the School of Mechanical Engineering and Automation, Northeastern University, China, in 2010. He is currently an Associate Professor with the School of Control Engineering, Northeastern University at Qinhuangdao. His major research interests include reliability and sensitivity modeling and analysis.

HAIJUAN HU received the Ph.D. degree from LMBA, Universite de Bretagne-Sud, France, in 2012. She is currently with the School of Mathematics and Statistics, Northeastern University at Qinhuangdao. Her major research interest includes image processing.