



A novel multi-graph framework for salient object detection

Ye Lu¹ · Kedong Zhou¹ · Xiyin Wu² · Penghan Gong³

Published online: 13 February 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Graph-based methods have been widely adopted for predicting the most attractive region in an image. Most of the existing graph-based methods only utilize single graph to describe the image information, and thus cannot adapt for complex scenes. In this paper, a novel multi-graph framework for salient object detection is proposed. The proposed method is divided into three steps. Firstly, an image is divided into superpixels and described as a multi-graph, where superpixels are represented as nodes and their information is computed by color space and location space. Secondly, the multiple graphs are combined into a novel multi-graph-based manifold ranking propagation framework to obtain a coarse map. Finally, a map refinement model is developed to improve the quality of the coarse map. Experimental results on four challenging datasets show that the proposed method performs favorably against the state-of-the-art salient object detection methods.

Keywords Salient object detection · Multi-graph framework · Map refinement · Manifold ranking

1 Introduction

Human visual system usually tends to pay more attention to some regions of an image and further processes rich visual contents. These regions are called salient since they stand out from their surroundings in the image [1]. Saliency detection is a task for simulating selective visual attention of humans to predict the most important parts where a human observer may look at. It can reveal the attentional mechanisms of biological visual systems. Numerous applications in computer vision, such as object recognition [2], image compression [3], image retrieval [4] and image segmentation [5], benefit from saliency detection as a preprocessing step to reduce computational cost.

Saliency detection has been studied by researchers in physiology, psychology, neural systems and computer vision, and lots of models have been proposed based on different assumptions. Generally speaking, these computational models can be divided into two categories: top-down models and

bottom-up models. Top-down models are goal-oriented and concerned with a specific object class, which is learned from a large number of images in a supervised way. These models are usually related to the subsequent applications. Maybank [6] utilizes a probabilistic model to describe salient regions for image matching. Yang et al. [7] obtain discriminative expression of target-specific objects by a combination of dictionary learning and conditional random fields (CRFs). Unlike top-down models, bottom-up models are stimuli-driven and adopted low-level cues, such as contrast, color, texture and boundary. Some representative models are presented in [8–10]. The focus of this paper is bottom-up saliency model.

Recently, graph-based methods have attracted much attention in bottom-up models and achieved state-of-the-art performance. For graph-based methods, an input image is described as a graph with superpixels as nodes, in which the adjacent nodes in the graph are connected by weighted edges. The most representative saliency nodes are served as seeds. Then, the saliency values are diffused along the weighted edges from seeds to other nodes. Up to now, a great many graph-based saliency detection methods have been proposed. Harel et al. [11] treat the saliency detection problem as a random walk scheme, in which salient value of a region is computed according to the frequency of node visits at equilibrium. Based on the work of [11], Wang et al. [12] evaluate saliency by entropy rate, which

✉ Ye Lu
luye_njust@163.com

¹ School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

² School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

³ Shijiazhuang Campus, AEU, Shijiazhuang 050003, China

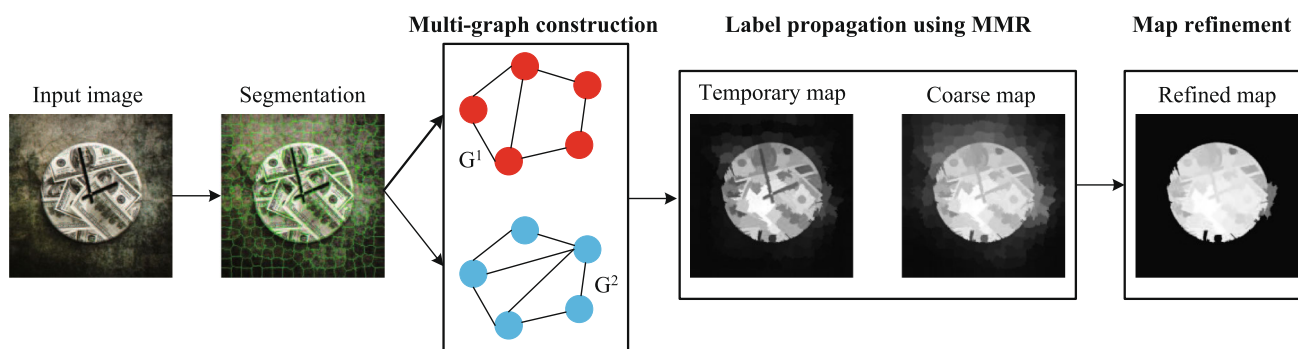


Fig. 1 Diagram of the proposed method

is the mean information transferred from a node to other nodes during a random walk. In the work of Mai et al. [13], a strategy called data aggregation is explored in saliency detection; more specifically, pixel-wise, conditional random field (CRF) and image-dependent aggregation are introduced to aggregate saliency map. The work of Gong et al. [14] employs the teaching-to-learn and learning-to-teach strategies to propagate saliency from related simple regions to difficult regions. Zhang et al. [15] combine objectness and compactness cues to estimate saliency. Qi et al. [16] exploit both Boolean-based and foreground-based models for saliency detection. Yang et al. [17] first use four boundaries of an image as background seeds to obtain background-based saliency value, then utilize the foreground seeds to formulate the final saliency map.

In the graph-based models employed by above methods, only one graph is constructed for saliency detection. However, constructing one graph for a picture may neglect some essential structure information in the complex scene. Since people often connect multiple types of relationships of a scene simultaneously, a corresponding multi-graph should be constructed to describe image information from different feature representations.

Multi-graph-based models have successfully utilized in many applications, such as social recommendations [18], object classification [19] and visual tracking [20]. To handle multi-relational social networks, Mao et al. [18] present a multi-graph ranking model to identify and recommend the nearest neighbors of particular users in high-order environments. Wu et al. [19] develop a positive and unlabeled multi-graph learning framework for complicated objects classification. As appearance always changing in object tracking, Yang et al. [20] combine graph-based ranking and multiple feature representation to multiple graph matrices. These applications of multi-graph-based models perform favorably against the state-of-the-art methods.

In this paper, a novel multi-graph framework is proposed to solve the problems of single-graph-based saliency detection methods. The diagram of our proposed algorithm is shown in Fig. 1. Firstly, an image is partitioned into

superpixels. A multi-graph is constructed by representing superpixels as nodes. The node information is described from two feature spaces, including color space and location space, to leverage the complementarity between them. Secondly, multiple cues are integrated in a new multi-graph-based propagation model to conquer the complex situation. Finally, a map refinement process is added to improve the propagation result. The major contributions of this paper are summarized as follows:

- (1) A new multi-graph construction way is proposed to express image information from two aspects.
- (2) A multi-graph-based manifold ranking framework is proposed to propagate information of nodes based on the new multi-graph. The propagation result is represented as a coarse map.
- (3) A map refinement model is developed to improve the quality of the coarse map by highlighting heterogeneous regions of a salient object and suppressing background noises simultaneously. The remainder of this paper is organized as follows. Section 2 briefly reviews the related work. In Sect. 3, we develop a label propagation framework for multiple graphs. Section 4 describes the process of proposed saliency detection algorithm in detail. Section 5 presents experimental results and discusses the performance of our framework. Finally, the paper is concluded in Sect. 6.

2 Related work

In the last few years, numerous saliency methods have been proposed. A comprehensive survey of salient object detection methods is presented in [1]. The following gives a review of saliency detection methods that are related to our method.

The first computational model of saliency is proposed by Itti et al. [21]. This model is a bottom-up model deriving from psychological theories. A large number of approaches are proposed to detect salient object in images afterward.

Harel et al. [11] define a graph on image and utilize random walks to compute saliency. Liu et al. [22] treat salient object detection as a picture segmentation problem by using contrast and center-surround mechanism. Achanta et al. [23] develop a frequency-tuned method which detects saliency using color deviation from the mean image color. Shen et al. [24] propose a low-rank matrix recovery model which decomposes the feature matrix extracted in an image into low-rank part and sparse part. Cheng et al. [25] utilize the color contrast information between foreground and background to assign saliency values. Deep learning-based methods have attracted much attention to researchers recently. Li et al. [26] develop a multi-task FCNN to model the intrinsic semantic saliency properties in a data-driven manner. Kuen et al. [27] adapt convolutional-deconvolutional network and provide contextual information in attended subregions to detect salient objects of multiple scales. Li et al. [28] evaluate saliency of a query region by utilizing both high-level and low-level features generated in deep networks. The framework proposed by [29] introduces a series of short connections between shallower and deeper side-output layers, which can highlight the entire salient object and locate its boundary in the meantime. Hu et al. [30] model the semantic features of salient objects and tackle the blurred saliency boundaries by their proposed deep level set network. Although deep learning-based methods have shown their effectiveness for saliency detection, they are usually hard to be fine-tuned or trained.

In recent years, more and more researchers pay their attention to graph-based methods for their simplicity and efficiency. Among these graph-based methods, it is proved that manifold ranking is a potential model. Yang et al. [17] bring manifold ranking in saliency detection firstly. An input image is represented as a single-layer graph, and four boundary nodes are picked out as seeds based on background prior. Then, saliency information is diffused by manifold ranking and the optimal ranking result is regarded as saliency value. This method achieves good performance in terms of both accuracy and speed, so that a great many of the improvements are proposed. These improvements are developed by three aspects: graph construction, seed and propagation model. Although a well-constructed graph is essential for final saliency result, few works focus on graph construction. Wang et al. [31] construct an effective single graph to capture local/global contrast. Then, the information of seed is propagated similar to [17]. Sun et al. [32] represent the image as a sparse single graph which not only reflects the local neighborhood information by linking nodes and their neighbors, but also takes the symmetry of images into consideration by connecting the opposite border nodes. The boundary seeds adopted in [17] may be occupied by foreground object. Li et al. [33] relieve this problem by eliminating erroneous boundaries which have distinctive color distribution compared to the remaining three boundaries. Wang et al. [34] select back-

ground seeds by removing the foreground noises which have low boundary probability. Lots of works pay their attention on label propagation model. Tao et al. [35] model the ranking problem in a matrix factorization framework, which can enforce similar saliency values on neighboring nodes by combining spatial information and embedding labels. Li et al. [33] propose a regularized random walks ranking by integrating random walks and GMR. The regularized random walks define a new constraint to consider local image data and prior estimation. Xia et al. [36] use locally linear embedding to conduct manifold propagation. The locally linear embedding algorithm can discover the relationship between each node and its neighbors in the feature subspace.

Previous methods mentioned above only construct a single graph in their models, so that they are not robust to complex scenes. To address this problem, we propose a novel multi-graph construction model and propagate saliency information under it. Furthermore, a corresponding multi-graph framework is developed to combine multiple cues. The details of proposed method are given in the following section.

3 Manifold ranking based on multi-graph

In this section, we first review the standard MR, and then we extend MR model to a novel multi-graph-based MR framework.

3.1 Single-graph-based manifold ranking

MR can be viewed as an extreme case of semi-supervised learning algorithm, in which only positive or negative labeled data are available [37,38]. The goal of GMR is to learn a ranking function, which ranks the unlabeled data according to their relevances to the given seeds. An description of GMR is as follows. A graph is constructed according to the relevance among data. A positive ranking score is assigned to each seed, and zero is assigned to each remaining node. All nodes diffuse their ranking score to their neighbors based on a propagation function. Then, their final ranking score can be got after the propagation process.

For a natural image, we can treat each pixel or super-pixel of this image as a node. Then, a graph $G = (V, E)$ can be constructed. This graph is an undirected single graph in general. All of the nodes are constituted to a set $V = \{v_1, v_2, \dots, v_n\}$; some of these nodes are labeled as seeds and the rest need to be ranked based on their relevance to the seeds. A vector $\mathbf{y} = [y_1, y_2, \dots, y_n]$ is defined as an indication vector for seeds, in which $y_i = 1$ if v_i is a seed, and $y_i = 0$ otherwise. The edges E are weighted by an affinity matrix $\mathbf{W} = [w_{ij}]_{n \times n}$. Let vector $\mathbf{f} = [f_1, \dots, f_n]$ denote

the ranking result, the optimal ranking result of nodes is calculated by solving the following optimization problem:

$$\mathbf{f} = \arg \min_{\mathbf{f}} \frac{1}{2} \sum_{i,j=1}^n w_{ij} (f_i - f_j)^2 + \mu \sum_{i=1}^n d_{ii} \left(f_i - \frac{y_i}{d_{ii}} \right)^2 \quad (1)$$

where d_{ii} is an entry of diagonal degree matrix $\mathbf{D} = \text{diag}(d_{11}, \dots, d_{nn})$ with $d_{ii} = \sum_j w_{ij}$. μ is a parameter controlling the balance of the first term and the second term. The first term is the smoothness constraint, which denotes that a good ranking function should not change too much between neighbor nodes. The second term is the label fitness constraint, which represents that a good ranking function should not change too much from the initial seed assignment. Equation (1) can be rewritten as the matrix form:

$$\mathbf{f} = \arg \min_{\mathbf{f}} \mathbf{f}^T \mathbf{L} \mathbf{f} + \mu (\mathbf{f} - \mathbf{D}^{-1} \mathbf{y})^T \mathbf{D} (\mathbf{f} - \mathbf{D}^{-1} \mathbf{y}) \quad (2)$$

where \mathbf{L} is the normalized graph Laplacian matrix with $\mathbf{L} = \mathbf{D} - \mathbf{W}$. Deviating Eq. (2) to be zero, the optimized closed-form solution can be depicted as:

$$\mathbf{f} = (\mathbf{D} - \gamma \mathbf{W})^{-1} \mathbf{y} \quad (3)$$

where $\gamma = 1/(1 + \mu)$.

3.2 Multi-graph-based manifold ranking

In this subsection, a new framework is proposed combining multiple graphs. Given K graphs and their corresponding edge weights $\{W^k = (w_{ij}^k)\}_{k=1}^K$, we integrate these K graphs into a new regularized framework. Thus, Eq. (1) is extended to a multi-graph-based manifold ranking (MMR) framework:

$$\begin{aligned} \mathbf{f} = \arg \min_{\mathbf{f}} & \frac{1}{2} \sum_{k=1}^K \alpha^k \sum_{i,j=1}^n w_{ij}^k (f_i - f_j)^2 \\ & + \mu \sum_{k=1}^K \sum_{i=1}^n d_{ii}^k \left(f_i - \frac{y_i}{d_{ii}^k} \right)^2 \quad (4) \\ \text{s.t.} & \sum_{k=1}^K \alpha^k = 1, \alpha^k \in [0, 1] \end{aligned}$$

where d_{ii}^k is the (i, j) -element of degree matrix of k -th graph. We integrate different graphs by a weighted combination. α^k denotes the weight of k -th graph. Equation (4) can be rewritten as follows for similarity:

$$Q(\mathbf{f}) = \sum_{k=1}^K \{ \alpha^k \mathbf{f}^T \mathbf{L}^k \mathbf{f} + \mu [\mathbf{f} - (\mathbf{D}^k)^{-1} \mathbf{y}]^T \mathbf{D}^k [\mathbf{f} - (\mathbf{D}^k)^{-1} \mathbf{y}] \} \quad (5)$$

$$\mathbf{f} = \arg \min_{\mathbf{f}} Q(\mathbf{f}) \quad \text{s.t.} \sum_{k=1}^K \alpha^k = 1, \alpha^k \in [0, 1]$$

where $\mathbf{L}^k = \mathbf{D}^k - \mathbf{W}^k$ is the normalized graph Laplacian matrix of k -th graph.

Thus, the task changes to minimize the regularization framework above with respect to the ranking result vector \mathbf{f} . By setting the derivative of Eq. (5) to zero, we have

$$\frac{\partial Q(\mathbf{f})}{\partial \mathbf{f}} = \sum_{k=1}^K \{ \alpha^k \mathbf{L}^k \mathbf{f} + \mu \mathbf{D}^k [\mathbf{f} - (\mathbf{D}^k)^{-1} \mathbf{y}] \} = 0 \quad (6)$$

which can be transformed into

$$\sum_{k=1}^K \{ \alpha^k \mathbf{L}^k \mathbf{f} + \mu \mathbf{D}^k [\mathbf{f} - (\mathbf{D}^k)^{-1} \mathbf{y}] \} = 0 \quad (7)$$

Then, we obtain:

$$\mathbf{f} = \sum_{k=1}^K \alpha^k (\mathbf{D}^k - \gamma \mathbf{W}^k)^{-1} \mathbf{y} \quad (8)$$

By combining different graphs, a picture can be analyzed through multiple properties. We illustrate the visual comparison of MR and MMR in Fig. 2. Since MMR integrates more visual cues by multiple graph, it can obtain more accurate saliency maps than MR.

4 Saliency detection via multi-graph-based manifold ranking

Our method includes three main parts: construct multi-graph, generate a coarse map via MMR and refine the coarse map. In this section, we state these three parts sequentially.

4.1 Multi-graph construction

As most of saliency detection methods suffer from complex scenes consisting of numerous perceptual heterogeneous regions, a multi-graph is constructed in this paper to relieve the problem. To analyze an image from multiple aspects, the multi-graph describes image information from different image feature spaces. We construct two graphs $G^1 = (V, E^1, W^1)$ and $G^2 = (V, E^2, W^2)$ which correspond to location and color features, respectively. For an image, the two graphs share common nodes but have their individual

Fig. 2 Visual comparison of MR and MMR. **a** Input image. **b** Saliency map generated by using MR. **c** Saliency map generated by using MMR. **d** Ground truth

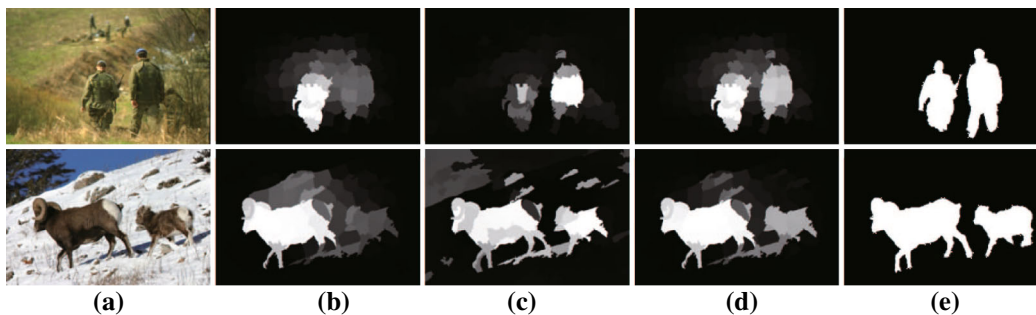
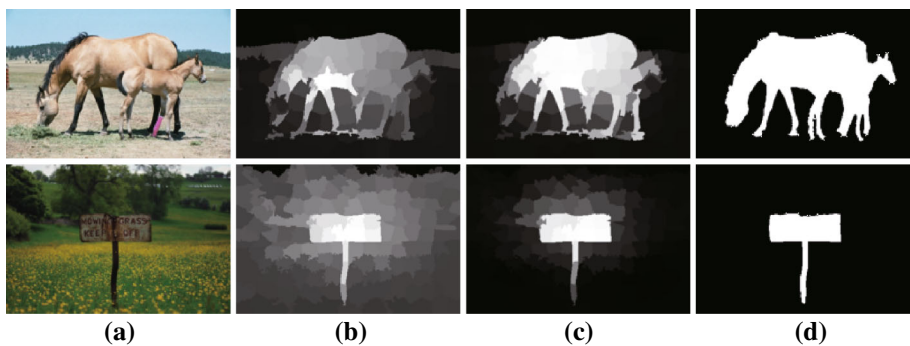


Fig. 3 The effectiveness of multi-graph structure. **a** Input image. **b** Single graph owing same structure with G^1 . **c** Single graph owing same structure with G^2 . **d** Multi-graph structure. **e** Ground truth

edges and weights. The visual comparisons of constructing single graph and multi-graph are shown in Fig. 3. The multi-graph structure can take advantage of each single graph. Thus, more reliable results are generated.

4.1.1 Nodes

The multi-graph constructed in this paper treats each superpixel as a node. As demonstrated in the work of [39], segmenting an image into homogeneous superpixels is a good means to capture intrinsic structure information and improve computational efficiency of saliency detection. In this paper, simple linear iterative clustering (SLIC) algorithm [40] is applied to generate n superpixels for an image. n is set to 200 in our experiments as previous graph-based methods [41].

4.1.2 Edges

Since nearby nodes are likely to have similar appearance, previous graph-based saliency detection methods connect nodes based on their spatial relationship [41,42]. Each edge connects two neighbor nodes by an undirected link. Similar with previous work, we also connect location nearby nodes in the first graph G^1 . A pair of nodes v_i and v_j are linked submitting to following rules:

- (1) As boundary prior proposed by assuming that the background regions of a picture are always close to the

boundaries, we enforce arbitrary boundary nodes to be connected to each other. In this case, the graph is enforced to be a closed-loop graph, which tends to reduce the geodesic distance of similar nodes.

- (2) v_i is not only connected to its immediate neighbors, but also the nodes sharing same boundaries with immediate neighbors. Thus, nodes in G^1 are connected with a two-ring topology and local smoothness can be utilized in this way.

Based on the above two rules, the edges set $E^1 = e^1 \cup e^2$ of G^1 can be obtained as follows:

$$\begin{aligned}
 e^1 &= \{(v_i, v_j) : v_i, v_j \in B\} \\
 e^2 &= \{(v_i, v_j) : v_j \in N_i^s \vee \exists v_k \in N_i^s | v_j \in N_k^s\}
 \end{aligned}
 \tag{9}$$

where B denotes the set of boundary nodes consisted by superpixels at the image border. N_i^s is the set of spatial adjacent nodes of v_i .

Previous graph-based methods exploited one graph so that they ignore the color similar relationship between nodes. These nodes which are sharing similar labels are probably far away from each other. Therefore, G^1 with the two-ring topology is difficult to link such nodes. To address this issue, we construct another graph G^2 based on following rules:

- (1) Same to the first rule of E^1 , all boundary nodes are connected with each other. According to the observation

of [43], background regions closed to boundaries are usually large and homogeneous. In other words, most of the boundary nodes may have similar color.

- (2) v_i is connected to its proximity nodes in Lab color space. Since affinity propagation algorithm can find clusters with less error and require less time, we use this algorithm to acquire the color neighbors of v_i .

We conclude the rules of edges set $E^2 = e^1 \cup e^3$ in the formulation way:

$$\begin{aligned} e^1 &= \{(v_i, v_j) : v_i, v_j \in B\} \\ e^3 &= \{(v_i, v_j) : v_j \in N_i^c\} \end{aligned} \quad (10)$$

where N_i^c is the cluster of v_i in Lab color space computed by affinity propagation algorithm. The way to choose edges of E^2 connects each node with its color nearby nodes. It can make up the lack of E^1 and further help to detect multiple objects or a large single object with heterogeneous regions.

4.1.3 Weights

In previous work, the weight W_{ij} between two connected nodes v_i and v_j is usually evaluated by color distance between nodes [17,41] as follows:

$$W_{ij} = \begin{cases} \exp\left(-\frac{\|\mathbf{c}_i - \mathbf{c}_j\|}{\sigma^2}\right), & \text{if } v_i \text{ and } v_j \text{ are connected} \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

where σ is a constant parameter that controls the affinity scale of weight. \mathbf{c}_i is the mean value of the superpixel corresponding to v_i in the CIELab color space. The weight definition in Eq. (11) may not appropriate when background regions have similar color to salient regions. Therefore, a term represented location relationship between nodes is added. We adopt the sine location distance described in our previous work [44] to define the location relationship. Specifically, since the nodes belonging to opposite borders of an image are always visually similar, these nodes can be considered spatially close in the sine location distance. The weight matrix W^1 of G^1 is defined as follows:

$$W_{ij}^1 = \begin{cases} \exp\left(-\frac{\|\mathbf{c}_i - \mathbf{c}_j\| + \|\sin(\pi|\mathbf{d}_i - \mathbf{d}_j|\|)}{\sigma^2}\right), & \text{if } v_{ij} \in E^1 \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

where $\sin(\cdot)$ is sine location distance which calculates the sine function of a vector element wisely. We should note that W^1 is a symmetrical matrix which satisfies $w_{ij} = w_{ji}$.

For the second graph G^2 , the edge set E^2 links color neighbors in the same cluster of the whole image. It is beneficial to find out similar nodes which have long distance with each

other. Therefore, we further design another weight matrix W^2 to consider the color distance of clusters:

$$W_{ij}^2 = \begin{cases} \exp\left(-\frac{\|\mathbf{C}_i - \mathbf{C}_j\| + \|\sin(\pi|\mathbf{d}_i - \mathbf{d}_j|\|)}{\sigma^2}\right), & \text{if } v_{ij} \in E^2 \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

where \mathbf{C}_i is the mean Lab color value of all nodes which belong to the cluster N_i^c .

4.2 Coarse map generation via multi-graph-based manifold ranking

In this section, a coarse saliency map is built by propagating saliency values on the multi-graph constructed in Sect. 4.1. Specifically, we establish the coarse map from the perspective of background to obtain that how each superpixel is likely to be the background.

Given the indication vector \mathbf{y} , where the i -th element of \mathbf{y} is set to 1 if v_i is a background seed, and zero otherwise. The saliency values of nodes are calculated by their ranking scores based on Eq. (8). For easier presentation, we rewrite Eq. (8) similar to [45]:

$$\mathbf{f} = \mathbf{A}\mathbf{y}, \quad \mathbf{A} = \sum_{k=1}^2 \alpha^k (\mathbf{D}^k - \gamma \mathbf{W}^k)^{-1} \quad (14)$$

where \mathbf{A} is a learned optimal affinity matrix. For the reason that \mathbf{y} represents an indicator vector, f_i can be treated as the sum of the relevances of the i -th node to all seeds. The saliency scores of nodes are assigned by the normalized ranking value $1 - \bar{\mathbf{f}}$ for the situation of background seeds, and $\bar{\mathbf{f}}$ for the situation of foreground seeds. Moreover, the diagonal elements of \mathbf{A} should be set to zero to avoid its strong self-correlation problem.

Nodes which are probably background can be set to seeds in the label propagation model. In this paper, boundary prior is employed to pick out background seeds. Boundary prior states that the salient object seldom touches picture border, so that the nodes along the image's four boundaries can be treated as seeds (as shown in Fig. 1). Since this simple prior is probably not correct in all images, the separation/combination (SC) strategy [17] is utilized to alleviate this problem. The SC strategy obtains four propagation results $\mathbf{f}(\text{T})$, $\mathbf{f}(\text{D})$, $\mathbf{f}(\text{L})$ and $\mathbf{f}(\text{R})$ by selecting the top, down, left and right boundary nodes as seeds, respectively. These four propagation results are integrated to generate a temporary map $\mathbf{S}^t = (s_i^t)_n$:

$$s_i^t = \prod_{p \in \{\text{T}, \text{D}, \text{L}, \text{R}\}} [1 - \bar{\mathbf{f}}_i(p)] \quad i = 1, 2, \dots, n \quad (15)$$

where $\bar{\mathbf{f}}$ is the normalized vector of \mathbf{f} . In other words, \mathbf{f} is normalized to the range between 0 and 1 before using it. Each element in \mathbf{f} denotes the relevance of a node to the background seeds, while its complement can be viewed as the saliency measure. The four sub-maps of the coarse map are calculated with different \mathbf{y} and same weight matrix $\mathbf{W}^1, \mathbf{W}^2$. So the learned optimal affinity matrix \mathbf{A} computes only once for each image. For the reason that the number of superpixels is small, \mathbf{A} can be calculated efficiently, and the computation time of the temporary map is low.

As show in Fig. 4, some foreground regions cannot be highlighted completely. Therefore, an addition process is added to alleviate this problem as [42]. The saliency values of nodes in the temporary map denote the confidence of each nodes being salient, so they are treated as the value of indication vector y . We can obtain another map based on the new indication vector and Eq. (14). This map is named as coarse map $\mathbf{S}^c = (s_i^c)_n$ and defined as follows:

$$s_i^c = \bar{\mathbf{f}}_i \quad i = 1, 2, \dots, n \tag{16}$$

where $\bar{\mathbf{f}}_i$ is a normalized vector which denotes normalizing all of the values of \mathbf{f} between the range of 0 and 1. The salient object is usually compact and homogeneous than background region. Hence, a seed belongs to object can propagate its information to other nodes of the object easily. Therefore, the coarse map usually shows better results than the temporary map, as shown in Fig. 4.

4.3 Map refinement

While most parts of salient objects can be detected in the coarse map, some background regions are still contained (see Fig. 4). To get more convincing results, we should diffuse the saliency information from the candidate foreground regions to further improve the coarse map \mathbf{S}^c .

In the map refinement process, we first construct a new multi-graph by treating the result of \mathbf{S}^c as the mid-level feature. Mid-level features usually can be estimated by the likelihood of a pixel or region belonging to a generic object. The definition of the node and edge remain the same as Sects. 4.1.1 and 4.1.2, while the definition of weights changes in the new multi-graph. We should note that N_i^c in Eq. (10) changes to the cluster of V_i in the coarse map by affinity propagation algorithm. The weights $\mathbf{W}^{r1} = (w_{ij}^{r1})_{n \times n}$ and $\mathbf{W}^{r2} = (w_{ij}^{r2})_{n \times n}$ are computed as follows:

$$w_{ij}^{r1} = \begin{cases} \exp\left(-\frac{\|s_i^c - s_j^c\|}{\sigma^2}\right), & \text{if } v_{ij} \in E^1 \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

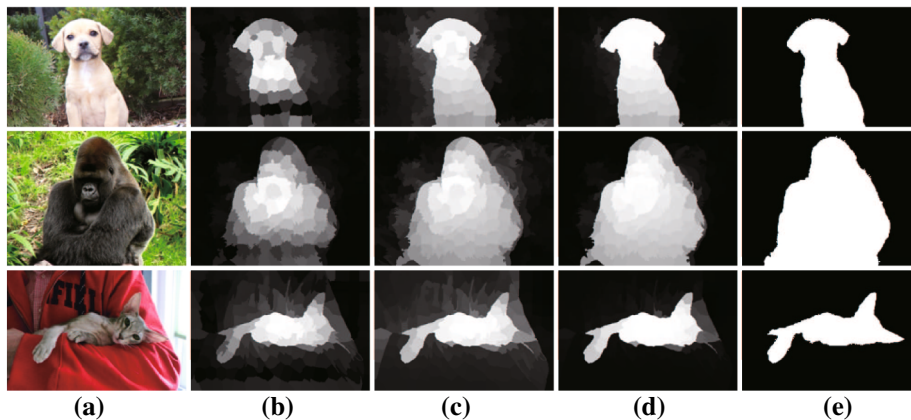
$$w_{ij}^{r2} = \begin{cases} \exp\left(-\frac{\|C_i^c - C_j^c\|}{\sigma^2}\right), & \text{if } v_{ij} \in E^2 \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

where C_i^c is the mean value of all nodes which belong to the new cluster N_i^c . The computation way of new weights is similar to the original weights in Sect. 4.1.3, while the difference is that the distance between these nodes is computed by the mid-level feature rather than low-level features.

The next step is to choose a new indication vector. The direct way is to pick out the nodes with large saliency value in \mathbf{S}^c to be foreground seeds. The foreground seeds usually are chosen by a threshold [14,17]. However, this way mixes with the background noises which will be confused in the map refinement process easily. Since the coarse map indicates the confidence degree of each nodes to be salient, here we adopt the saliency values of \mathbf{S}^c to be the values of indication vector.

To refine the coarse map, we propose a new map refinement model with a new multi-graph and a new indication vector. The refined result $\mathbf{g} = (g_i)_n$ can be generated by solving the following optimization problem:

Fig. 4 Examples of three maps generated in the proposed method. **a** Input image. **b** Temporary map. **c** Coarse map. **d** Refined map. **e** Ground truth



$$\mathbf{g} = \arg \min_{\mathbf{g}} \frac{1}{2} \sum_{k=1}^2 \alpha^k \sum_{i,j=1}^n w_{ij}^{r_k} (s_i^c - s_j^c)^2 + \mu \sum_{i=1}^n (g_i - s_i^c)^2 \quad (19)$$

$$\text{s.t. } \sum_{k=1}^2 a^k = 1, a^k \in [0, 1]$$

Equation (19) can be rewritten as matrix form:

$$\mathbf{g} = \arg \min_{\mathbf{g}} \sum_{k=1}^2 \alpha^k \mathbf{g}^T (\mathbf{D}^{r_k} - \mathbf{W}^{r_k}) \mathbf{g} + \mu \|\mathbf{g} - \mathbf{S}^c\|^2 \quad (20)$$

$$\text{s.t. } \sum_{k=1}^2 a^k = 1, a^k \in [0, 1]$$

where \mathbf{D}^{r_k} is the diagonal matrix corresponding to \mathbf{W}^{r_k} . The two terms of Eq. (20) define costs from different constraints. The first term is the smoothness constraint which encourages continuous saliency values. This term can help to highlight salient object uniformly and achieve better results. The second term is a fitness constraint, which restricts the refined result \mathbf{g} should not change too much from the coarse map \mathbf{S}^c . The optimized solution of Eq. (20) can be obtained by deriving with respect to \mathbf{g} and equal to zero. The result is given as follows:

$$\mathbf{g} = \sum_{k=1}^2 \alpha^k (\mathbf{D}^k - \beta \mathbf{W}^k + \mathbf{I})^{-1} \mathbf{S}^c \quad (21)$$

All the elements of \mathbf{g} are normalized to $[0, 1]$ to generate the refined map \mathbf{S}^r . The refined map can obtain a better result by highlighting the salient regions and restraining the background regions, as shown in Fig. 4. The process of our proposed method in this paper is summarized in Algorithm 1.

Algorithm 1 Saliency detection via multi-graph-based manifold ranking

Require: A natural image and related parameters.

Ensure: A saliency map \mathbf{S}^r with the same size as the input image.

- 1: Use SLIC [40] algorithm to segment the input image into n superpixels.
 - 2: Construct multi-graph G^1 and G^2 with superpixels as nodes, link nodes of G^1 and G^2 by Eqs. (9) and (10), respectively, compute the edge weight matrices of the two graph by Eqs. (12) and (13), respectively.
 - 3: Obtain an indicator vector \mathbf{y} by selecting nodes on each border of the image as seeds, compute their corresponding saliency maps based on Eq. (14). Combine the four saliency maps to generate the temporary map \mathbf{S}^t by Eq. (15).
 - 4: Acquire an indicator vector \mathbf{y} by the saliency values of \mathbf{S}^t , calculate the coarse map \mathbf{S}^c by Eq. (16).
 - 5: Use \mathbf{S}^c to form a new multi-graph, the weights of the new multi-graph are defined in Eqs. (17) and (18).
 - 6: Refine the coarse map to get a better result \mathbf{g} by Eq. (21). \mathbf{g} is normalized to obtain the refined map \mathbf{S}^r .
-

5 Experimental results

5.1 Experimental setup

5.1.1 Datasets

The salient object detection performance of the proposed method is evaluated on five public benchmark datasets: ASD [23], MSRA10K [46], THUR15K [47], PASCALS [9] and ECSSD [48]. ASD is widely used and contains 1000 images with accurate pixel-wise masks. Images in ASD usually have only one object and simple background. MSRA10K conquers the drawback of ASD with 10,000 images. The huge images make this dataset more challenging and provide more varieties. THUR15K owns 15,000 images and divides them into five categories. Only 6223 pictures of them equipped with human-labeled masks. PASCALS contains 850 images with multiple objects. The ECSSD dataset contains 1000 images with many semantically meaningful but structurally complex scenes.

5.1.2 Evaluation metrics

We use five metrics for comprehensive evaluation. The five metrics include the precision-recall (PR) curve, F-measure curve, mean absolute error (MAE), overlapping ratio (OR) and weighted F-measure (WF) score.

(1) **PR curve:** It is a popular metric to evaluate saliency map. Segmenting a saliency map into a binary map with vary thresholds in range $[0, 255]$, the precision (P) and recall (R) of this saliency map can be computed by comparing the binary map M_B to the ground-truth mask M_G .

$$P = \frac{|M_B \cap M_G|}{|M_B|}, \quad R = \frac{|M_B \cap M_G|}{|M_G|} \quad (22)$$

where $|\cdot|$ is the number of nonzero entries in a binary mask. The precision value can be viewed as the accuracy of a saliency detection method, while the recall value depicts the detection consistency.

(2) **F-measure curve:** In general, both precision and recall cannot estimate the quality of a saliency map completely. Therefore, F-measure curve is applied to evaluate the weighted harmonic mean between precision and recall values. Same to PR curve, we exploit vary thresholds from 0 to 255 to generate binary salient object mask. The definition of F-measure F_β is as follows:

$$F_\beta = \frac{(1 + \beta^2)P \cdot R}{\beta^2 P + R} \quad (23)$$

where β^2 is set to 0.3 to stress precision than recall [1].

(3) **MAE**: It is used to evaluate the true negative saliency assignments of non-salient pixels. MAE is calculated between a saliency map M and a binary ground-truth mask M_G for all pixels [39]:

$$MAE = \frac{|M - M_G|}{N} \tag{24}$$

where N is the number of pixels.

(4) **OR**: It is a metric to measure the true negative. A mask M_T is obtained by binarizing saliency map M [49]. OR is defined as follows:

$$OR = \frac{|M_T \cap M_G|}{|M \cup M_G|} \tag{25}$$

(5) **WF**: It is a new measure proposed by Margolin et al. [50] recently. WF amends the interpolation, dependency and equal importance flaws of other measures. Similar with F-measure, WF is calculated based on a weighted harmonic mean of weighted precision P^w and recall R^w :

$$F_\beta^w = \frac{(1 + \beta^2)P^w \cdot R^w}{\beta^2 P^w + R^w} \tag{26}$$

5.1.3 Parameter setting

The parameters of the proposed method are set as follows. Same as previous MR-based works [17,31], the constant parameter σ^2 in Sect. 4.1.3 is set to 0.1 and the balancing parameter γ in Sect. 4.2 is set to 0.99. The parameters α_1 and α_2 in Sect. 3.2 are the weights of G^1 and G^2 , respectively. For the reason that $\alpha_1, \alpha_2 \in [0, 1]$ and $\alpha_2 = 1 - \alpha_1$, the parameter α_2 can be ensured based on α_1 . Now we consider about the extreme cases. The MMR is converted to GMR when the value of α^1 is either 0 or 1. The propagation result of MMR depends on G^1 when α^1 is set to 1, whereas G^2 when α_1 is set to 0. We test the influences on different values of α_1 on ECSSD dataset using the average value of F-measure curve (AF) as the evaluation metric. According to the experimental result shown in Table 1, the performance achieves best when $\alpha_1 = 0.8$. Therefore, α_1 and α_2 are set to 0.8 and 0.2, respectively, for all experiments.

Table 1 Performance comparisons of parameter α^1

α^1	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
AF	0.6226	0.6354	0.6476	0.6604	0.6730	0.6843	0.6916	0.6975	0.6999	0.6967	0.6841

The best result is highlighted in bold

5.2 Validation of the proposed method

To verify the effectiveness of different components in the proposed method, three experiments are tested on ECSSD dataset using PR curve as the evaluation metric. The three experiments include the necessity of using multi-graph, the effectiveness weight computation and the results of three maps generated in our method. Firstly, the necessity of using multi-graph is tested and the results are shown in Fig. 5a. The blue curve and green curve provide the performance of only using G^1 and G^2 , respectively, and their corresponding propagation models can be viewed as MR. The red curve presents the performance of using multi-graph structure, and its corresponding propagation model is MMR. The results show that using multi-graph structure can achieve better result than the single graph. That also reflects the multi-graph structure can acquire the supplementary information of G^1 and G^2 .

Then, the design of weight computation is examined in Fig. 5b. The blue curve and green curve denote only using color feature and spatial feature in Eqs. (12) and (13). The red curve is the result of using both color and spatial features. From the result shown in Fig. 5b, we can find out that the situation of combining two features can achieve the better performance with large margin than the other two situations. So both of the two features are necessity in the weight computation.

Finally, the result of temporary map, coarse map and refined map computed by Eqs. (15), (16) and (21) is compared in Fig. 5c. For the reason that more foreground regions can highlighted in the coarse map, the performance of the coarse map is better than the temporary map. The refined map can suppress background noises in the coarse map, so the map refinement process favorably improves the accuracy of the coarse map.

5.3 Comparison with the state-of-the-art

In this section, we evaluate our proposed method and compare to 17 state-of-the-art methods, including IT [21], LC [51], GB [11], SR [8], FT [23], HC [25], GS [43], LR, MR [17], RC [25], PCA [9], RBD [52], TLLT [14], RRW [33], WLR [53], DS [26] and DeepMC [54]. The 17 methods are chosen based on two aspects. On the one hand, we select some recent methods, such as TLLT, RRW, DSR, WLR, DS and DeepMC. On the other hand, the proposed method is compared with variety methods: IT is biologically moti-

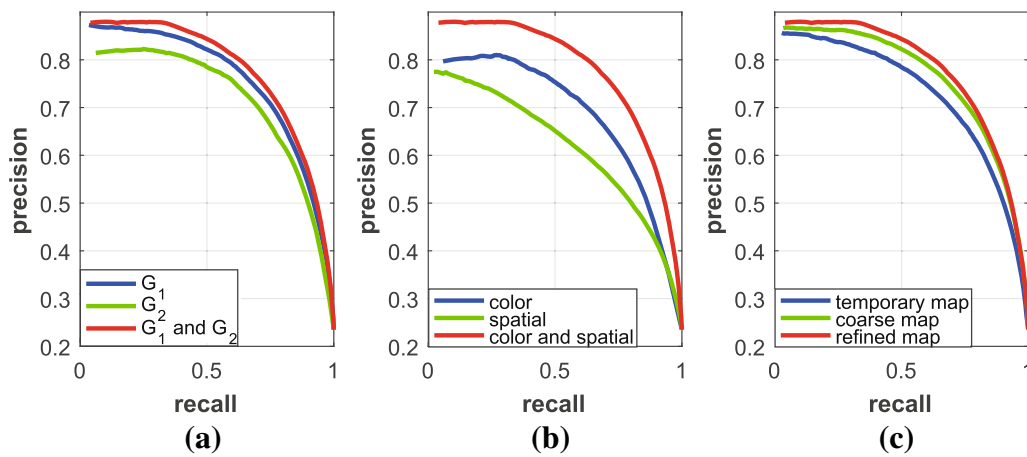


Fig. 5 Validation of the proposed method on ECSSD dataset. **a** Necessity of using multi-graph. **b** Effectiveness weight computation. **c** Results of three maps generated in our method

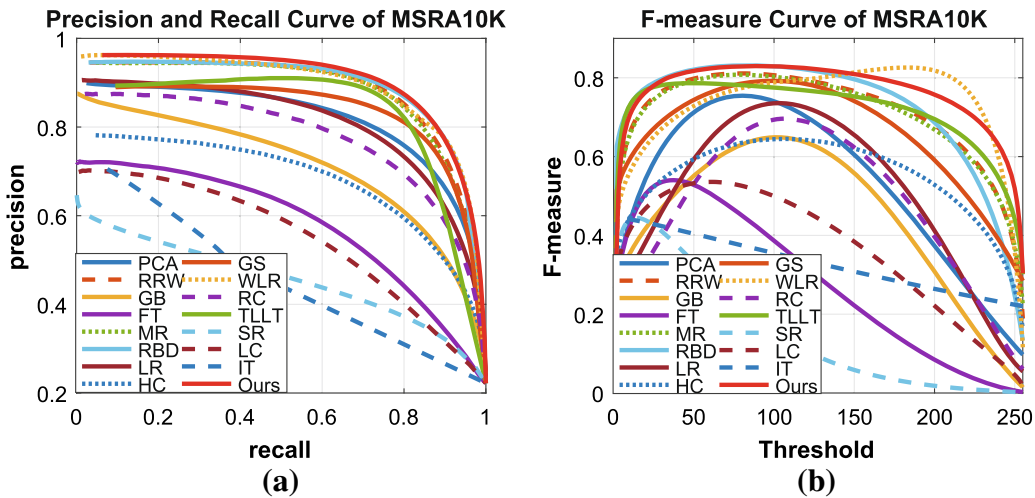


Fig. 6 Quantitative comparisons on MSRA10K dataset in terms of PR curve and F-measure curve. **a** PR curve. **b** F-measure curve

vated, LC is a purely computational model, SR utilizes the information in frequency domain, FT is based on frequency-tuned technology, RC utilizes contrast information between regions, PCA combines fixation techniques, HC is based on color histogram, LR and WLR are based on low-rank matrix recovery model, DS and DeepMC are deep learning methods, GB, GS, MR, TLLT, RRW and RBD are graph-based method. Saliency maps of these methods are obtained by running the source codes or executables provided by the authors.

5.3.1 Quantitative comparisons with unsupervised methods

Since the proposed method is unsupervised, the proposed method is compared to the unsupervised methods firstly. The PR curve, F-measure curve of four datasets are shown

in Figs. 6, 7, 8 and 9, respectively. The PR curves of proposed method are higher than all of the other algorithms on these four datasets. This results denote that our method can achieve a relatively high accuracy than the state-of-the-art algorithms. For the F-measure curves, our method outperforms on the most thresholds when segmenting saliency maps, whereas it is lower than WLR in the range [150, 240]. However, WLR obtains relatively low F-measure value by varying threshold in the range [0, 100]. Our method can provide more stable performance than WLR. The MAE, OR and WF values are displayed in Tables 2, 3, 4, respectively. In terms of MAE value, the proposed method gains the best performance on MSRA10K dataset and ECSSD dataset, ranks second on PASCALS and ranks third on THUR15K. WLR and TLLT achieves best MAE on PASCALS and THUR15K, respectively. For OR and WF value, our method outperforms state-of-the-art methods on all of four dataset. It means that

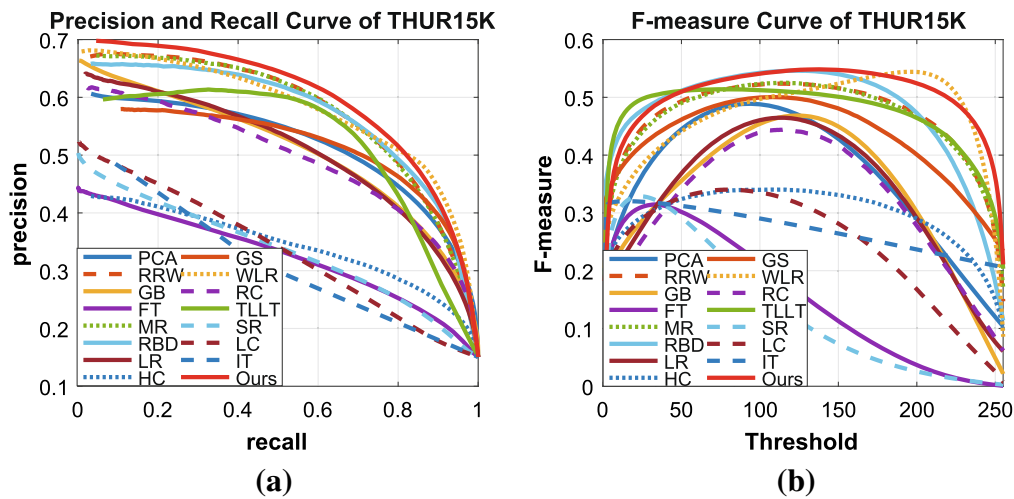


Fig. 7 Quantitative comparisons on THUR15K dataset in terms of PR curve and F-measure curve. **a** PR curve. **b** F-measure curve

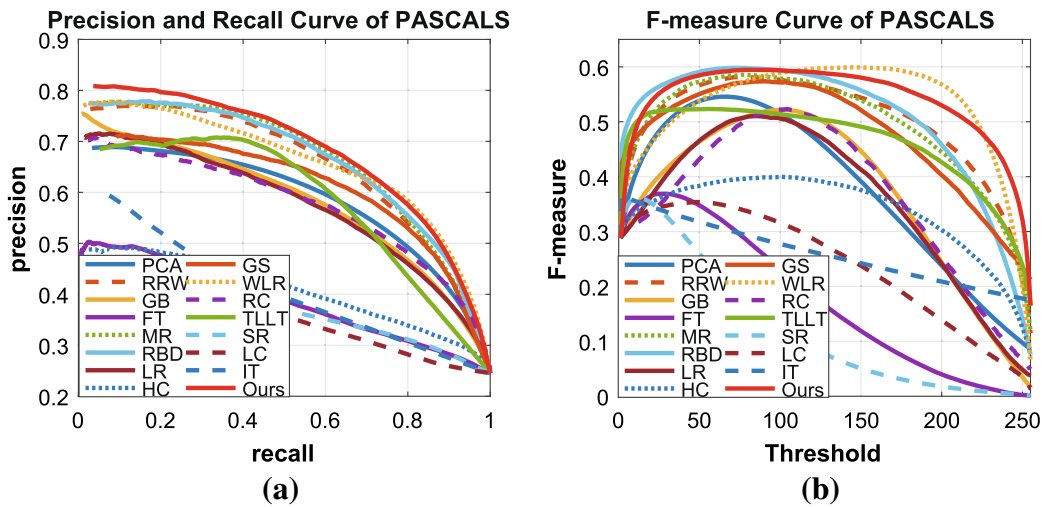


Fig. 8 Quantitative comparisons on PASCALS dataset in terms of PR curve and F-measure curve. **a** PR curve. **b** F-measure curve

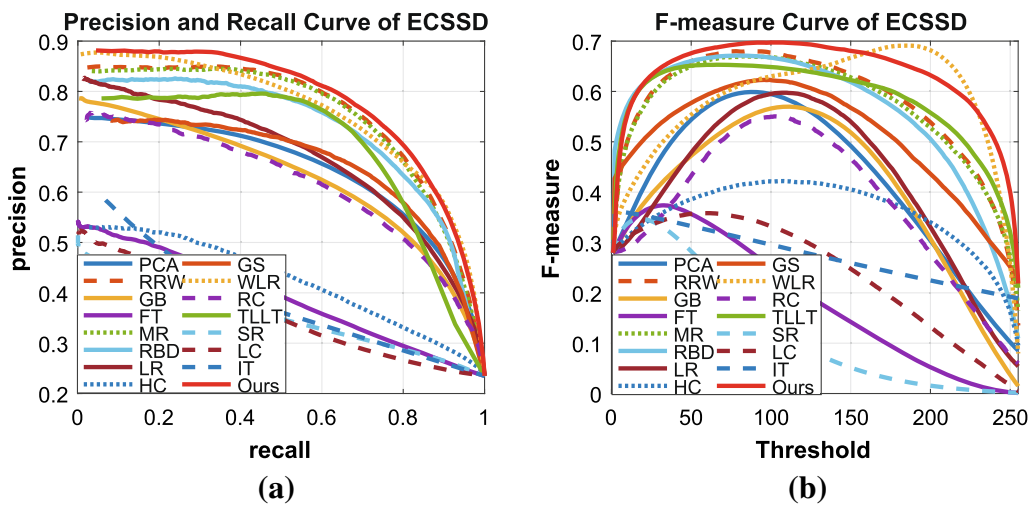


Fig. 9 Quantitative comparisons on ECSSD dataset in terms of PR curve and F-measure curve. **a** PR curve. **b** F-measure curve

Table 2 Quantitative comparisons on four datasets in terms of MAE

	Ours	PCA	RRW	GB	FT	MR	RBD	LR	HC	GS	WLR	RC	TLLT	SR	LC	IT
MSRA10K	0.1079	0.1854	0.1234	0.2225	0.2306	0.1265	0.1080	0.2241	0.2156	0.1385	0.1264	0.2517	<i>0.1141</i>	0.2319	0.2335	0.2130
THUR15K	<i>0.1600</i>	0.1982	0.1786	0.2305	0.2074	0.1791	0.1505	0.2480	0.2839	0.1755	0.2093	0.2781	0.1473	0.1751	0.2280	0.1611
PASCALS	0.2100	0.2480	0.2287	0.2730	0.2855	0.2233	0.2012	0.2869	0.3411	0.2235	0.2344	0.3129	0.2101	0.2634	0.3140	0.2441
ECSSD	0.1693	0.2469	0.1838	0.2631	0.2698	0.1893	0.1714	0.2741	0.3258	0.2059	0.2104	0.3004	<i>0.1723</i>	0.2637	0.2940	0.2368

The lower value achieved, the better performance is. The best three methods are highlighted in bold, italic and bold and italic

Table 3 Quantitative comparisons on four datasets in terms of OR

	Ours	PCA	RRW	GB	FT	MR	RBD	LR	HC	GS	WLR	RC	TLLT	SR	LC	IT
MSRA10K	0.7162	0.5764	0.6987	0.4580	0.3788	0.6918	0.7158	0.5240	0.4848	0.6638	<i>0.7014</i>	0.4345	0.6494	0.2559	0.3801	0.2198
THUR15K	0.4359	0.3613	<i>0.4290</i>	0.3268	0.2087	0.4271	0.4305	0.3252	0.2407	0.3876	0.4123	0.2870	0.4118	0.2116	0.2324	0.1840
PASCALS	0.2181	0.4031	0.4192	0.2399	0.3585	0.1783	0.1891	0.1685	0.4373	0.3464	0.4119	0.3151	0.1972	<i>0.4306</i>	0.4356	0.2975
ECSSD	0.5443	0.3950	0.5312	0.3631	0.2162	0.5204	<i>0.5253</i>	0.3695	0.2475	0.4606	0.4978	0.2816	0.4985	0.1762	0.2095	0.1777

The higher value achieved, the better performance is. The best three methods are highlighted in bold, italic and bold and italic

Table 4 Quantitative comparisons on four datasets in terms of WF value

	Ours	PCA	RRW	GB	FT	MR	RBD	LR	HC	GS	WLR	RC	TLLT	SR	LC	IT
MSRA10K	0.6920	0.4731	0.6464	0.3937	0.2772	0.6406	0.6855	0.4247	0.4728	0.6062	0.6650	0.3839	0.6560	0.1545	0.3449	0.2302
THUR15K	0.4280	0.2971	0.3826	0.2696	0.1580	0.3797	0.4216	0.2603	0.2483	0.3710	0.3978	0.2411	<i>0.4141</i>	0.1336	0.2203	0.1982
PASCALS	0.4665	0.3338	0.4236	0.3254	0.1910	0.4235	<i>0.4564</i>	0.3125	0.2981	0.4224	0.4655	0.3167	0.3988	0.1303	0.2357	0.1843
ECSSD	0.5474	0.3643	0.5001	0.3506	0.1950	0.4929	0.5128	0.3509	0.3035	0.4501	0.5003	0.3142	<i>0.5101</i>	0.1371	0.2361	0.1925

The higher value achieved, the better performance is. The best three methods are highlighted in bold, italic and bold and italic

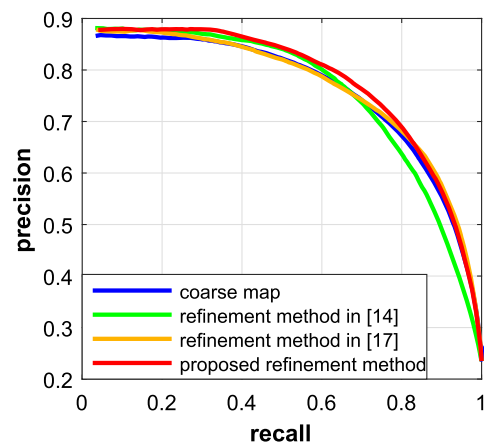


Fig. 10 Controlled validation for the proposed refinement model on ECSSD dataset

our method can obtain high accuracy and relatively low error rate in the meantime. It is worth mention that our method performs best on ECSSD dataset in terms of all evaluation metrics. Thus, the proposed method is validated strong potential when dealing with complex scenes. Moreover, the controlled experiments are added to verify the effectiveness of our proposed refinement model on ECSSD dataset. Two existing refinement methods [14] and [17] are exploited to refine the coarse saliency map, respectively. The experimental results are shown in Fig. 10. The refinement method in [14] improves the result of coarse map in the range [0,0.7] of recall, while reduces in the range [0.7,1] of recall. The refinement method in [17] can improve the result of coarse map slightly. The result of our proposed refinement model is better than [14] and [17] which both exploit heuristic thresholding method.

5.3.2 Quantitative comparisons with deep learning methods

Deep learning-based saliency detection methods have achieved good performance recently. We compare two of them in this section in terms of PR curves. The experimental results of THUR15K and ECSSD are shown in Fig. 11, respectively. For the THUR15K dataset with 6233 images, both of two deep learning methods get much better performance than the proposed method. For the ECSSD dataset, the proposed method can get comparable results with the deep learning methods. That proves our method has good performance in the complex scenes. Although the deep learning methods perform better than our method, it needs a long time to train and fine-tuned. Our method is unsupervised without training process and has lower requirement of computer hardware than the deep learning method.

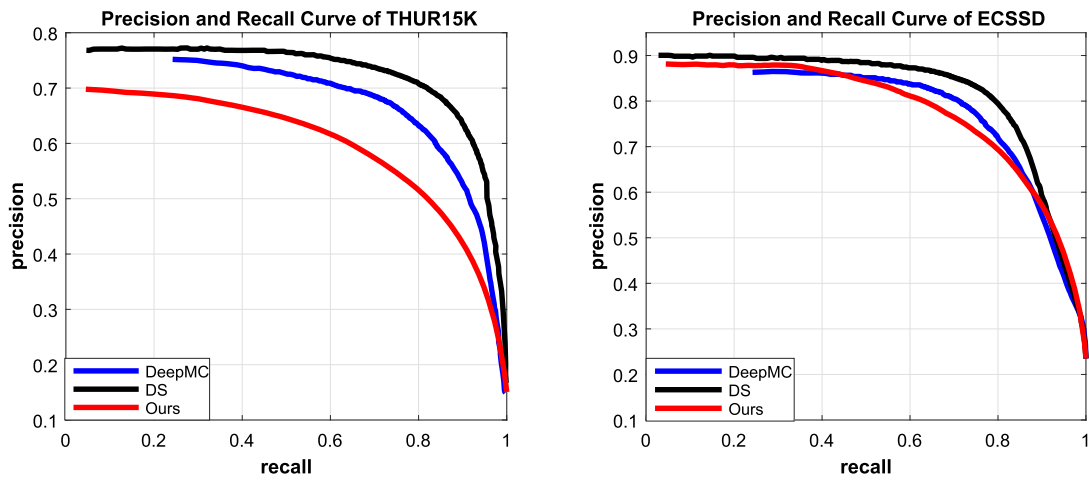


Fig. 11 Quantitative comparisons with deep learning methods in terms of PR curve. **a** THUR15K dataset. **b** ECSSD dataset

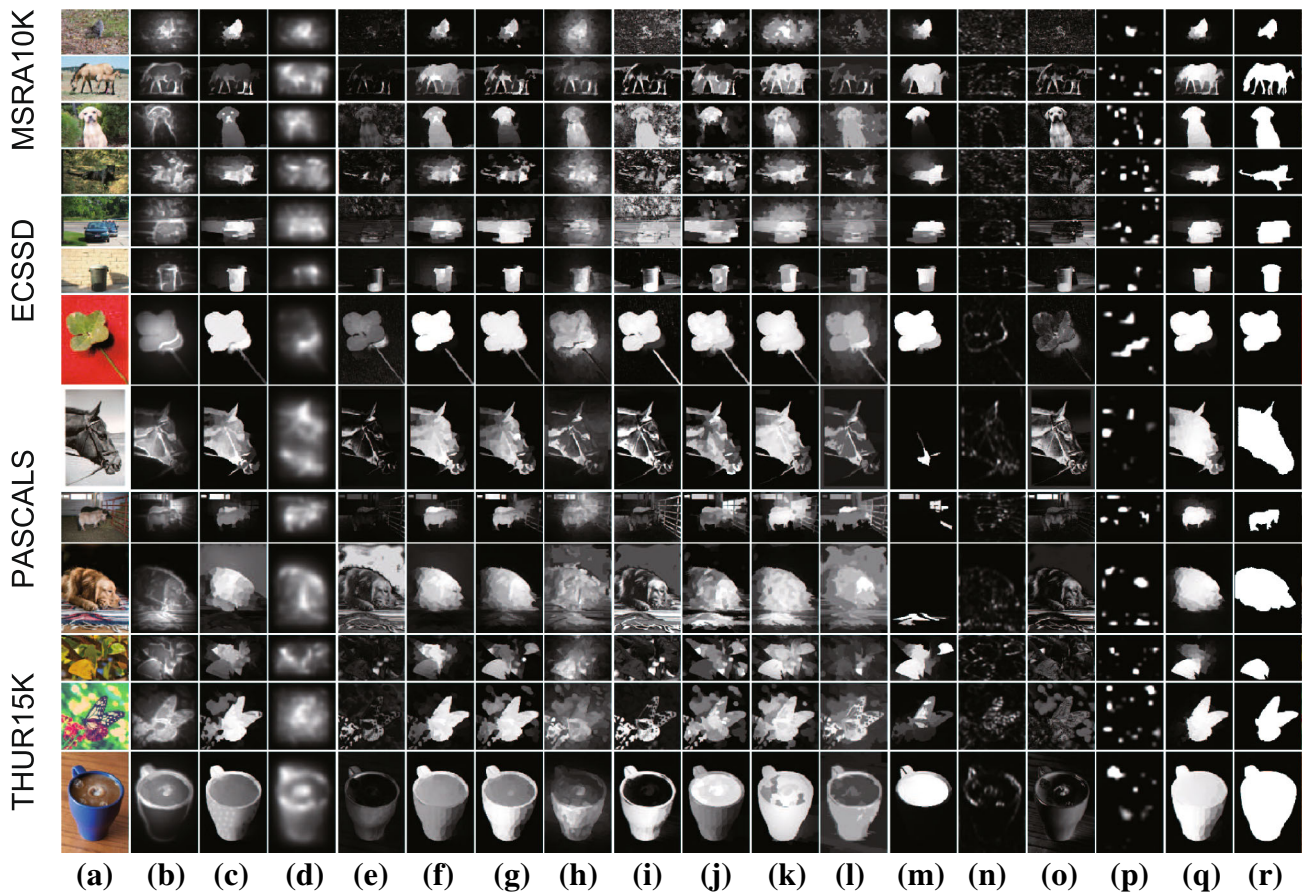


Fig. 12 Visual comparisons of the saliency maps on four datasets. **a** Input image. **b** PCA. **c** RRW. **d** GB. **e** FT. **f** MR. **g** RBD. **h** LR. **i** HC. **j** GS. **k** WLR. **l** RC. **m** TLLT. **n** SR. **o** LC. **p** IT. **q** Ours. **r** Ground Truth

Table 5 Efficiency comparisons (seconds per image)

Method	Ours	PCA	RRW	GB	FT	MR	RBD	LR	HC	GS	WLR	RC	TLLT	SR	LC	IT
Time	0.676	2.354	0.906	0.659	0.072	0.311	0.258	11.862	0.016	0.234	1.060	0.135	1.127	0.040	0.009	0.305
Code ^a	M	M	M	M	C	M	M	M	C	M	M	C	M	M	C	M

^a ‘M’ denotes the code which is written in MATLAB, while ‘C’ indicates the code which is written in C/C++

5.3.3 Visual comparisons

Figure 12 shows some visual comparisons from four datasets. In the case of simple background (see the third, sixth, seventh, eighth and thirteenth exemplars), our model can detect salient object successfully. The saliency values can be assigned uniformly for all regions in salient object. In the case of complex scenes (see the second, fifth, ninth, tenth and twelfth exemplars), the other methods usually miss some parts of the salient objects or mistake for the background regions, while the proposed method can pop out the saliency objects successfully. In the case of foreground and background having similar appearance (see the first, fourth and eleventh exemplars), our model separates almost the entire salient object from the background, while other methods often fail. All of the visual comparison results demonstrate the effective and robust of the proposed multi-graph framework and the refinement process.

5.3.4 Efficiency comparisons

The average runtime on ECSSD database of the proposed method and other methods is presented in Table 5. All of the experiments run in MATLAB R2016a and a 64-bit PC with Intel(R) Xeon(R) E3-1231 v3 CPU @ 3.40 GHz and 8GB RAM. As shown in Table 5, the methods using C/C++ implementation run fast. IT using MATLAB implementation also run fast for its simple formulation. Our method takes about 0.676 seconds per image. It is faster than most of the methods with MATLAB implementation. The graph construction takes 0.421 seconds per image while solving MMR, and the refinement process takes a short time. The reason is that the closed-form solution of MMR just needs a small-scale matrix inversion operation.

6 Conclusion

In this paper, we have presented a novel multi-graph-based salient object detection method. An input image is represented as multiple graphs. Nodes of multiple graphs are connected in different ways based on their peculiarity. Different from the traditional propagation frameworks which only use a single graph, a novel propagation framework is proposed by aggregating multiple graph information. Moreover, the propagation result is improved by using a refinement model which is also based on multiple graph information. Comprehensive experiments on four datasets have demonstrated the effect of our method, especially the ability to detect salient object in complex scenes. Since deep learning methods can perform satisfactorily in many complex scenes, the depth cue will be considered to integrated into our model in the future.

Acknowledgements This work was supported by National Natural Science Foundation of China under Grant No. 61602244.

Compliance with ethical standards

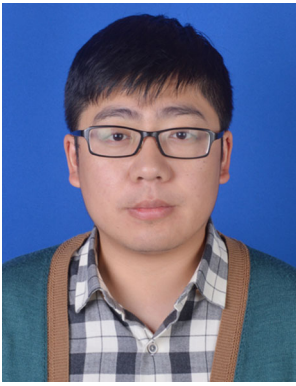
Conflict of interest The authors declared that they have no conflicts of interest to this work.

References

- Borji, A., Cheng, M.M., Jiang, H., Li, J.: Salient object detection: a benchmark. *IEEE Trans. Image Process.* **24**, 5706–5722 (2015)
- Ren, Z., Gao, S., Chia, L.T., Tsang, I.W.H.: Region-based saliency detection and its application in object recognition. *IEEE Trans. Circuits Syst. Video Technol.* **24**, 769–779 (2014)
- Guo, C., Zhang, L.: A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. *IEEE Trans. Image Process.* **19**, 185–198 (2010)
- Zhang, J., Feng, S., Li, D., Gao, Y., Chen, Z., Yuan, Y.: Image retrieval using the extended salient region. *Inf. Sci.* **399**, 154–182 (2017)
- Zhi, X.-H., Shen, H.-B.: Saliency driven region-edge-based top down level set evolution reveals the asynchronous focus in image segmentation. *Pattern Recognit.* **80**, 241–255 (2018)
- Maybank, S.J.: A probabilistic definition of salient regions for image matching. *Neurocomputing* **120**, 4–14 (2013)
- Yang, J., Yang, M.-H.: Top-down visual saliency via joint crf and dictionary learning. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, pp. 2296–2303 (2012)
- Hou, X., Zhang, L.: Saliency detection: a spectral residual approach. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
- Li, Y., Hou, X., Koch, C., Rehg, J. M., Yuille, A. L.: The secrets of salient object segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 280–287 (2014)
- Kim, J., Han, D., Tai, Y. W., Kim, J.: Salient region detection via high-dimensional color transform. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 883–890 (2014)
- Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: *Advances in Neural Information Processing Systems*, pp. 545–552 (2007)
- Wang, W., Wang, Y., Huang, Q., Gao, W.: Measuring visual saliency by site entropy rate. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, pp. 2368–2375 (2010)
- Mai, L., Niu, Y., Liu, F.: Saliency aggregation: a data-driven approach. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, pp. 1131–1138 (2013)
- Gong, C., Tao, D., Liu, W., Maybank, S. J., Fang, M., Fu, K., Yang, J.: Saliency propagation from simple to difficult. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2531–2539 (2015)
- Zhang, Q., Lin, J., Li, W., Shi, Y., Cao, G.: Salient object detection via compactness and objectness cues. *Vis. Comput.* **34**, 473–489 (2018)
- Qi, W., Han, J., Zhang, Y., Bai, L.: Saliency detection via Boolean and foreground in a dynamic Bayesian framework. *Vis. Comput.* **33**, 209–220 (2017)
- Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M.-H.: Saliency detection via graph-based manifold ranking. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, pp. 3166–3173 (2013)

18. Mao, M., Lu, J., Zhang, G., Zhang, J.: Multirelational social recommendations via multigraph ranking. *IEEE Trans. Cybern.* **47**(12), 4049–4061 (2017)
19. Wu, J., Pan, S., Zhu, X., Zhang, C., Wu, X.: Positive and unlabeled multi-graph learning. *IEEE Trans. Cybern.* **47**(4), 818–829 (2017)
20. Yang, X., Wang, M., Tao, D.: Robust visual tracking via multigraph ranking. *Neurocomputing* **159**, 35–43 (2015)
21. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 1254–1259 (1998)
22. Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.Y.: Learning to detect a salient object. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**, 353–367 (2011)
23. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1597–1604 (2009)
24. Shen, X., Wu, Y.: A unified approach to salient object detection via low rank matrix recovery. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 853–860 (2012)
25. Cheng, M.M., Mitra, N.J., Huang, X., Torr, P.H.S., Hu, S.M.: Global contrast based salient region detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 569–582 (2015)
26. Li, X., Zhao, L., Wei, L., Yang, M.-H., Wu, F., Zhuang, Y., Ling, H., Wang, J.: Deepsaliency: multi-task deep neural network model for salient object detection. *IEEE Trans. Image Process.* **25**(8), 3919–3930 (2016)
27. Kuen, J., Wang, Z., Wang, G.: Recurrent attentional networks for saliency detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3668–3677 (2016)
28. Lee, G., Tai, Y.-W., Kim, J.: Deep saliency with encoded low level distance map and high level features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 660–668 (2016)
29. Hou, Q., Cheng, M.-M., Hu, X., Borji, A., Tu, Z., Torr, P.: Deeply supervised salient object detection with short connections. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, pp. 5300–5309 (2017)
30. Hu, P., Shuai, B., Liu, J., Wang, G.: Deep level sets for salient object detection. In: CVPR, vol. 1, p. 2 (2017)
31. Wang, Q., Zheng, W., Piramuthu, R.: Grab: visual saliency via novel graph model and background priors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 535–543 (2016)
32. Sun, J., Lu, H., Liu, X.: Saliency region detection based on markov absorption probabilities. *IEEE Trans. Image Process.* **24**(5), 1639–1649 (2015)
33. Li, C., Yuan, Y., Cai, W., Xia, Y., Feng, D. D. et al.: Robust saliency detection via regularized random walks ranking. In: CVPR, pp. 2710–2717 (2015)
34. Wang, J., Lu, H., Li, X., Tong, N., Liu, W.: Saliency detection via background and foreground seed selection. *Neurocomputing* **152**, 359–368 (2015)
35. Tao, D., Cheng, J., Song, M., Lin, X.: Manifold ranking-based matrix factorization for saliency detection. *IEEE Trans. Neural Netw. Learn. Syst.* **27**(6), 1122–1134 (2016)
36. Xia, C., Li, J., Chen, X., Zheng, A., Zhang, Y.: What is and what is not a salient object? learning salient object detector by ensembling linear exemplar regressors. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 4321–4329 (2017)
37. Zhou, D., Bousquet, O., Lal, T. N., Weston, J., Schölkopf, B.: Learning with local and global consistency. In: Advances in Neural Information Processing Systems, pp. 321–328 (2004)
38. Zhou, D., Weston, J., Gretton, A., Bousquet, O., Schölkopf, B.: Ranking on data manifolds. In: Advances in Neural Information Processing Systems, pp. 169–176 (2004)
39. Perazzi, F., Krähenbühl, P., Pritch, Y., Hornung, A.: Saliency filters: contrast based filtering for salient region detection. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, pp. 733–740 (2012)
40. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(11), 2274–2282 (2012)
41. Jiang, B., Zhang, L., Lu, H., Yang, C., Yang, M.-H.: Saliency detection via absorbing markov chain. In: 2013 IEEE International Conference on Computer Vision (ICCV), IEEE, pp. 1665–1672 (2013)
42. Zhang, L., Yang, C., Lu, H., Ruan, X., Yang, M.H.: Ranking saliency. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1892–1904 (2017)
43. Wei, Y., Wen, F., Zhu, W., Sun, J.: Geodesic saliency using background priors. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) Computer Vision–ECCV 2012, pp. 29–42. Springer, Berlin (2012)
44. Wu, X., Jin, Z., Zhou, J., Ma, X.: Saliency propagation with perceptual cues and background-excluded seeds. *J. Vis. Commun. Image Represent.* **54**, 51–62 (2018)
45. Lu, S., Mahadevan, V., Vasconcelos, N.: Learning optimal seeds for diffusion-based salient object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2790–2797 (2014)
46. Cheng, M.-M., Warrell, J., Lin, W.-Y., Zheng, S., Vineet, V., Crook, N.: Efficient salient region detection with soft image abstraction. In: IEEE ICCV, pp. 1529–1536 (2013)
47. Cheng, M.-M., Mitra, N.J., Huang, X., Hu, S.-M.: Salienshape: group saliency in image collections. *Vis. Comput.* **30**, 443–453 (2014)
48. Shi, J., Yan, Q., Xu, L., Jia, J.: Hierarchical image saliency detection on extended cssd. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**, 717–729 (2016)
49. Peng, H., Li, B., Ling, H., Hu, W., Xiong, W., Maybank, S.J.: Salient object detection via structured matrix decomposition. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 818–832 (2017)
50. Margolin, R., Zelnik-Manor, L., Tal, A.: How to evaluate foreground maps? In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014)
51. Zhai, Y., Shah, M.: Visual attention detection in video sequences using spatiotemporal cues. In: Proceedings of the 14th ACM International Conference on Multimedia, MM '06, pp. 815–824. ACM, New York (2006)
52. Zhu, W., Liang, S., Wei, Y., Sun, J.: Saliency optimization from robust background detection. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR '14, IEEE Computer Society, pp. 2814–2821. Washington (2014)
53. Tang, C., Wang, P., Zhang, C., Li, W.: Salient object detection via weighted low rank matrix recovery. *IEEE Signal Process. Lett.* **24**, 490–494 (2017)
54. Zhao, R., Ouyang, W., Li, H., Wang, X.: Saliency detection by multi-context deep learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1265–1274 (2015)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Ye Lu received his Ph.D degree from Nanjing University of Science and Technology in 2016 and obtained his B.Sc. cum laude from the same university in 2010. Dr. Lu is currently a Lecturer at the school of Mechanical Engineering at Nanjing University of Science and Technology, China. Contact him at luye_njust@163.com.



Xiyin Wu is currently a Ph.D student in pattern recognition and intelligence systems at the school of Computer Science and Engineering at Nanjing University of Science and Technology, China. She obtained her Bachelor's degree in information security (2012) and the Master's degree in computer technology (2015) from Jiangnan University. Contact her at xiyinwu1990@gmail.com.



Kedong Zhou received his Ph.D degree from Nanjing University of Science and Technology in 1991. He went to Germany as a visiting scholar at the Institute of thermodynamics of the Department of mechanical engineering at Universitaet der Bundeswehr, Hamburg, for one year. Dr. Zhou is currently a Professor at the school of Mechanical Engineering at Nanjing University of Science and Technology, China. Contact him at zkd81151@njust.edu.cn.



Penghan Gong received his Ph.D degree from Nanjing University of Science and Technology in 2017. He obtained his Bachelor's degree in 2005 and the Master's degree in 2008 from Shijiazhuang Mechanical Engineering College, China. Dr. Gong is currently a Lecturer at the Shijiazhuang Campus, AEU, China. Contact him at gongpenghan@163.com.