Neurocomputing 329 (2019) 433–446

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Robust pixelwise saliency detection via progressive graph rankings

Lihua Wang[a,b], Bo Jiang[b], Zhengzheng Tu[b], Amir Hussain[c], Jin Tang[b,*]

[a] School of Computer Science and Technology, Huangshan University, Huangshan 245041, China
[b] School of Computer Science and Technology, Anhui University, Hefei 230601, China
[c] School of Computing, Merchiston Campus, Edinburgh Napier University, Edinburgh EH10 5DT, Scotland, U.K.

ARTICLE INFO

Article history:
Received 14 July 2018
Revised 27 October 2018
Accepted 28 October 2018
Available online 5 November 2018

Communicated by Shen Jianbing Shen

Keywords:
Saliency detection
Manifold ranking
Absorbing Markov chain
Regularized random walk
Progressive graph ranking

ABSTRACT

In this paper, we propose a novel saliency detection method based on superpixel-to-pixel level optimization. First, we segment the input image into superpixels under four scales. For each scale, we construct a k-regular basic graph with these superpixels as nodes. Furthermore, we enlarge the basic graph with virtual absorbing nodes and utilize absorbing Markov chain ranking to calculate background-based saliency. Second, for each scale, we obtain robust foreground queries from the previous result, and use manifold ranking to obtain foreground-based saliency. Third, a regularized random walk ranking based on the pixelwise graph for each scale is used to diffuse the saliency values among pixels. Finally, we obtain four saliency maps for the input image and integrate them together for the final saliency map. Extensive experiments on several challenging datasets reveal that the proposed method performs better in terms of precision, recall and *F*-measure values. Despite complex backgrounds, our method performs better in detecting small and/or multiple salient objects than other state-of-the-art methods as a whole.

1. Introduction

Saliency detection aims to identify outstanding attractive objects or regions in an image. It plays an important role in computer vision. In recent years, this research has been widely used in image quality assessment [1], object detection and recognition [2–4], image/video adaptive compression [5] and video tracking [6–8], among other applications.

From the input data perspective, saliency detection models can be classified into video saliency detection [9–12] and image saliency detection [13–38]. We focus on the latter. By the detection goal, saliency models can be grouped into two categories: eye fixation prediction and salient object detection. The former models aim on eye fixation, being used to predict the most attractive pixels in an image [9,10,13–17,39]. With the increasing requirement for high detection performance, the detection goal is preferred to detect salient objects or regions quickly [18–38,40–43]. From the perspective of machine learning mechanisms, saliency detection models can be divided into unsupervised models and supervised models. The unsupervised approaches take only the underlying image features into consideration, such as color, orientation, texture and so on [9,10,13–38]. In contrast, the supervised approaches typically learn salient object detectors based on the ground truth of training samples. Supervised methods [39–43] usually perform better than unsupervised methods because they utilize a large number of training data to learn a more discriminative mode for saliency model detection problem. When the training data are not available in some special cases, the unsupervised methods can be used.

Some unsupervised methods are based on contrast [13,18,22,24,44]. One of the earliest saliency models using local contrast is proposed by Itti et al. [13]. They utilize center-surround differences and combine multiscale image features (colors, intensity and orientation) into a single topographical saliency map. Jiang et al. [18] introduce a center-weighted local contrast to calculate the initial saliency map and use edge detection to obtain the shape. Then, the saliency maps and shapes are incorporated into the energy minimization model and update iteratively to generate the final saliency map. The local contrast-based methods prefer to produce high saliency values near object boundaries; however, these methods fail to highlight the whole salient objects. The global contrast-based methods alleviate this problem. Cheng et al. [22] introduce a global contrast-based method that consider the spatial relations simultaneously. Tong et al. [24] construct a saliency model by using global contrast information via CIELAB color, RGB color, texture, local binary pattern (LBP) and histogram of oriented gradient (HOG) features.

To reduce the computation complexity of saliency detection, most existing methods first segment the input image to obtain superpixels and use them as basic calculation units [24–31,33–38,45].

* Corresponding author.
E-mail address: jtang99029@foxmail.com (J. Tang).

https://doi.org/10.1016/j.neucom.2018.10.061
0925-2312/

Furthermore, graph based methods have achieved great success because they inherently consider the intrinsic structure in the image [25,26,31–38]. Yang et al. [25] propose a graph-based manifold ranking (MR) for saliency detection. First, they segment the input image into superpixels and rank the similarity of the superpixels based on the background and foreground queries. The method can well detect salient objects surrounded by a homogenous background. Unfortunately, the detection performance may be inaccurate for the images with complex backgrounds and small or multiple salient objects.

To obtain more robust performance, we propose a novel method based on MR via progressive graph rankings. We first use absorbing Markov chain ranking to obtain robust foreground queries. Then we use manifold ranking to obtain foreground-based saliency maps. Finally, we use the regularized random walk ranking to obtain pixelwise saliency maps. Experimental results show that our method is effective for small and multiple salient objects, even in complex backgrounds.

The main contributions of the paper are summarized as follows:

- We utilize the absorbing Markov chain to obtain robust foreground queries which provides a well labelled query/seed data for the following manifold ranking;
- We propose a saliency estimation model by adapting random walks on progressive graphs, which can well represent an image for modeling the behavior of salient objects.
- We use multiscale saliency integration to detect small and multiple salient objects more robustly in presence of complex backgrounds.

The remainder of the paper is organized as follows. Section 2 reviews fundamental properties of unsupervised saliency detection via background and foreground priors and graph-based methods. Section 3 details our saliency detection model. In Section 4, experimental results are presented, and we describe quantitative and qualitative comparisons between our method and other state-of-the-art methods. Finally, our conclusions are presented in Section 5.

2. Related work

In this section, we briefly review previous works from three aspects: saliency detection methods based on foreground cues, background cues and graphs.

Some previous works use the center prior to prefer the image center regions as foreground regions. However, these approaches involve strict assumptions about the size and location of the foreground. Furthermore, the method fails when salient objects are far from the image center. To overcome this problem, many works try to predict the salient regions by using a convex hull [26,27]. Yang et al. [26] apply the convex hull of points of interest to estimate salient objects and then smooth the initial saliency map by minimizing a saliency energy function with graph regularization. Xie et al. [27] use the convex hull to provide a coarse region estimation, from which they extract useful information to compute the prior distribution and the observation likelihood. Finally, they use the Bayesian framework to compute the final saliency values. Generally, these methods based on foreground cues usually result in extra background noises. Wang et al. [46] design a stereopsis-aware saliency bias that objects within the comfort zone or popping out from the screen are more visually important than others for stereoscopic thumbnail creation.

In addition, many works introduce boundary priors and treat image boundary regions as background [28–31,44]. Wei et al. [28] exploit the boundary and connectivity priors and define the saliency of an image patch as the geodesic distance to the virtual background node. Li et al. [29] mine the image boundaries to reconstruct dense and sparse errors, which are propagated and combined into a Bayesian framework. Wang et al. [30] take both the background and foreground into consideration. They first calculate a background-based saliency map and then use it to compute a foreground-based saliency map. Finally, they unify the two maps. Tu et al. [31] introduce a minimum spanning tree to formulate real-time salient object detection. Furthermore, they use boundary dissimilarity measure to alleviate the shortage of the distance transform. Wang et al. [44] use gradient flow optimization via both intra-frame boundary information and inter-frame motion information to indicate salient regions in video saliency detection.

Graph-based saliency detection models focus on the intrinsic structures in images and have achieved high performance in recent years [32–38]. To take the computation complexity into account, the image is segmented into blocks/superpixels, which are defined as graph nodes. To generate reliable superpixels more efficiently, several algorithms have been proposed recently, such as Turbo pixels [47], SLIC [48], DBSCAN [49] and so on. The feature distances between nodes are used as edge weights. Gopalakrishnan et al. [32] segment the image into 8×8 pixel blocks, then construct both fully connected graph and k-regular graph to model the global and local characteristics of the salient regions. Yan et al. [33] propose a hierarchical model to optimally select saliency values in a tree, which is especially useful when dealing with small-scale and high-contrast situations. Jiang et al. [34] introduce absorbing Markov chain to calculate saliency values. They construct a graph with four boundary nodes as absorbing nodes and the others as transient nodes. The absorption time from the transient node to the absorbing nodes is regarded as its saliency value. Sun et al. [35] use only the top and left boundaries for simplicity. They use absorbing Markov chain to obtain the preliminary saliency map and rank the relevance of nodes with foreground cues to obtain a better result. Wang et al. [36] add the direct connections between each node and the boundary nodes and use the geodesic distances as weights. Li et al. [37] use the image boundary selection method to remove one boundary and carry out a two-stage MR [25]. Furthermore, to improve the foreground saliency during the second stage, they propose the regularized random walk ranking to obtain the pixelwise saliency map. Their further work in [38] present a reversion correction process to increase robustness and take it as a common optimization algorithm for saliency detection methods based on boundary prior.

3. Our approach

Overall, our method contains four stages. The first two stages are based on superpixel level, and the other two are based on pixel level. First, we segment the image into superpixels with four scales. We use a method to generate robust queries and compute the background-based saliency. Second, we obtain the foreground-based saliency in the same way as the second stage in manifold ranking [25]. Third, we diffuse the saliency values from superpixels to pixels with a saliency refinement model. The final saliency map is obtained by integrating the four saliency maps obtained in the third stage. The diagram of the proposed method is shown in Fig. 1.

3.1. Basic graph construction

First, for an input image, we use the simple linear iterative clustering (SLIC) algorithm [48] to segment the image into superpixels, and treat these superpixels as nodes in a basic graph. The basic graph model [25] is denoted by $G^b = (V^b, E^b)$, where V^b is the node set and E^b is the edge set between these nodes. Considering the global color features and the compactness of

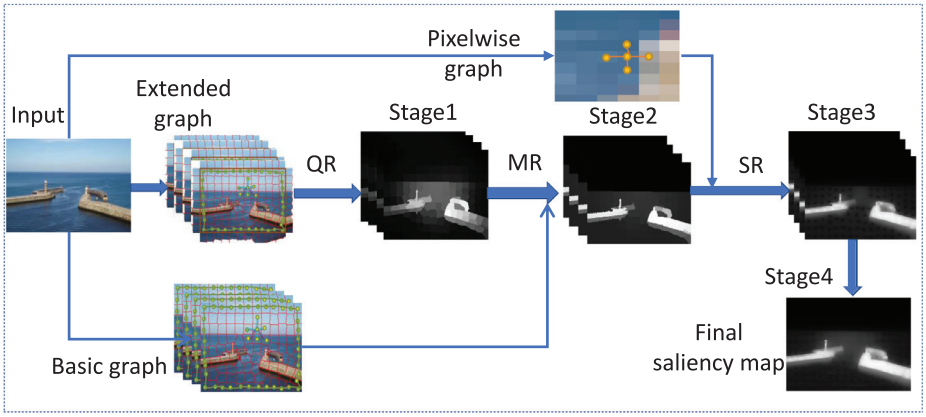
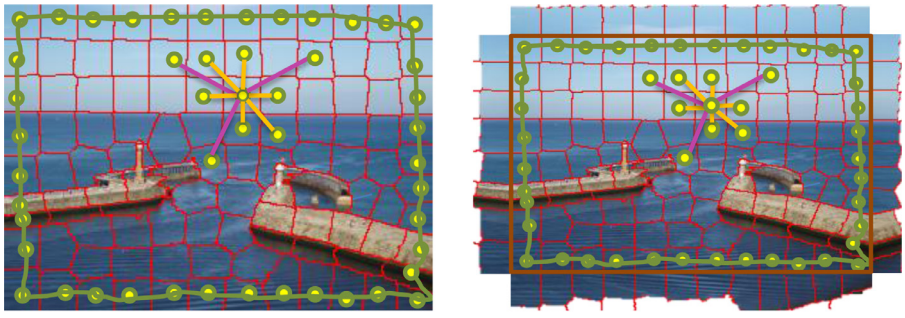


Fig. 1. Diagram of the proposed method. From left to right: QR: query generation, MR : manifold ranking [25], SR: saliency refinement.



(a) (b)

Fig. 2. Graph constructions. (a) The basic graph model; (b) The extended graph model, which is based on the basic graph. Then we copy the four boundary superpixels and add them into the extended graph as nodes.

salient objects, the k-regular graph ($k = 2$) is used for the edge connections between nodes in which each superpixel connects only with its direct neighbors and second-level neighbors. Meanwhile, for complex background situations, in order to reduce the color distance between background nodes, we fully connect the superpixels on four boundaries. All of the above connections form the edge set E^b. The basic graph is shown in Fig. 2(a). The weight of the edge is defined as the distance between two nodes in their CIELAB color space [25], which is computed as

$$w_{ij} = e^{-\frac{\|\mathbf{c}_i - \mathbf{c}_j\|}{\sigma^2}} \tag{1}$$

where $\mathbf{c}_i$ and $\mathbf{c}_j$ are the average CIELAB values of node i and j, respectively, and σ is a control constant.

3.2. Background-based saliency via absorbing Markov chain ranking

We use the algorithm based on the absorbing Markov chain [34] to obtain the background-based saliency map. To obtain a full-size saliency map, we extend the four boundary nodes of the basic graph model in Fig. 2(a), i.e., copy the four boundary superpixels as r virtual absorbing nodes. The nodes in the basic graph model are taken as t transient nodes. We define the extended graph as $G^e = (V^e, E^e)$, where V^e and E^e are obtained by adding the virtual absorbing nodes into V^b and E^b, as shown in Fig. 2(b) [34]. The time that a transient node randomly walks before being absorbed is regarded as the saliency value of the transient node.

To calculate the absorbing time of transient nodes, we first obtain the transition probability matrix $\mathbf{P}$ [34] based on the extended graph. Furthermore, the form of transition matrix $\mathbf{P}$:

$$\mathbf{P} = \begin{pmatrix} \mathbf{Q}_{t\times t} & \mathbf{R}_{t\times r} \\ \mathbf{0}_{r\times t} & \mathbf{I}_{r\times r} \end{pmatrix} \tag{2}$$

where $\mathbf{Q}$ represents the transition probabilities between transient nodes, $\mathbf{R}$ is the transition probability matrix of transient nodes and absorbing nodes, and $\mathbf{I}$ is an identity matrix.

After $\mathbf{Q}$ is computed, the fundamental matrix of $\mathbf{P}$ is then computed as $\mathbf{T} = (\mathbf{I} - \mathbf{Q})^{-1}$ [34]. The element $\mathbf{T}_{ij}$ represents the expected time that the transient node i takes to reach another transient node j. The sum of each row in matrix $\mathbf{T}$ is taken as the total expected time of a transient node before being absorbed [34]

$$\mathbf{y}_i = \sum_j \mathbf{T}_{ij} \tag{3}$$

By normalizing, the background-based saliency can be calculated as follows:

$$\mathbf{s_i}^b = \frac{\mathbf{y}_i - \min(\mathbf{y})}{\max(\mathbf{y}) - \min(\mathbf{y})} \tag{4}$$

Assigning the superpixel saliency values calculated in (4) to the corresponding pixels, we obtain the initial saliency map. In this step, the absorbing Markov chain tends to give higher saliency values for superpixels near the image center. However, this is inaccurate when salient objects are not near the image center. To overcome this problem, we first suppress the background by using an updating operation in [34]. Fig. 3 shows the comparison of the background-based saliency map between MR and the proposed method. Obviously, the proposed method obtains more accurate foreground objects than MR; however, from the corresponding ground truth perspective, the salient objects produced by the proposed method are still not complete and the background is not suppressed effectively.

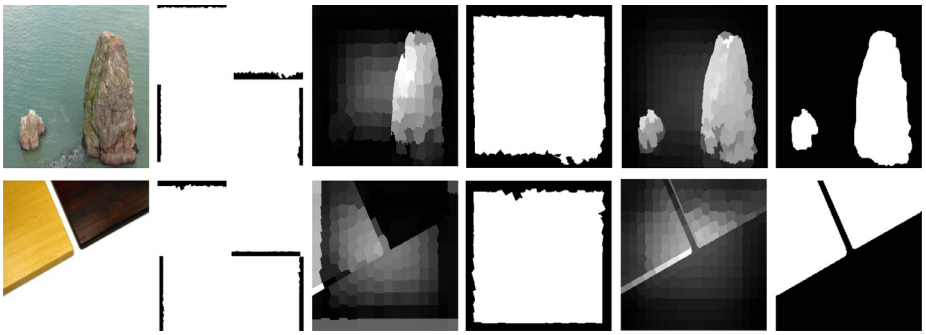


Fig. 3. Background-based saliency. From left to right: input image, background queries via MR [25], background-based saliency map via MR [25], background queries via the proposed method, background-based saliency map via the proposed method and ground truth.

3.3. Foreground-based saliency via MR

After obtaining the initial saliency values, we binarize them with an adaptive threshold [25] to obtain the query vector $\mathbf{q} = [q_1, q_2, \ldots, q_n]^T$, in which $q_i = 1$ if node i is a query and $q_i = 0$ otherwise. The graph is constructed as the same as the basic graph, while we balance the weight by the control constant in (1) to obtain the new edge weight. $\mathbf{W}^f$ is the new weight matrix, and the element $w'_{ij} = e^{-\frac{\|\mathbf{c}_i - \mathbf{c}_j\|}{\sigma_l^2}}$. We use the method in [25] to determine the foreground based saliency, which is calculated by

$$\mathbf{s}^f = (\mathbf{D}^f - \alpha \mathbf{W}^f)^{-1}\mathbf{q} \tag{5}$$

where $\mathbf{D}^f$ is the degree matrix of $\mathbf{W}^f$, and α is the control constant.

Assigning the saliency values in each superpixel calculated via (5) to the corresponding pixels, we obtain the saliency map based on the foreground, as shown in Fig. 4. The saliency map will lose certain salient regions if the robust foreground queries cannot be obtained in MR(2nd and 3rd columns in Fig. 4). The proposed method can well obtain the salient objects and effectively suppress the background after the foreground ranking (4th and 5th columns in Fig. 4). The salient objects of the image on the second row in Fig. 4 are near the boundary of the image. In the first stage of the proposed method, the boundary nodes are all treated as background in Fig. 3. However, after the foreground ranking in the second stage, the salient objects have been significantly corrected. The saliency maps show that the superpixel segmentation algorithm results in unsmooth salient maps in Fig. 4. Therefore, we use the subsequent diffusion method with a regularized random walk to obtain more smooth saliency results.

3.4. Saliency refinement with regularized random walk ranking

We use the method in [37], which utilized the prior saliency distribution in a regularized random walk, to obtain pixelwise saliency. The graph is based on the basic graph, which extends the nodes to the pixel level. The edge weight between pixel i and j is defined as k_{ij}, and the optimal ranking of queries is computed by solving the following optimization problem [37]:

$$\min_p \sum_{ij} (p_i - p_j)^2 k_{ij} + \frac{\mu}{2}\|\mathbf{p} - \tilde{\mathbf{s}}^f\|^2 \tag{6}$$

where $\mathbf{p}$ is the pixelwise saliency vector, and $\tilde{\mathbf{s}}^f$ is pixelwise saliency values of $\mathbf{s}^f$.

The saliency map after regularization on the pixel level is shown in Fig. 5d. Relative to the second stage, the white salient points in the salient objects are the optimal results after applying the regularized random walk. The saliency map has been greatly improved, but the white spots are too discrete. Therefore, in the next step, we will use multiple scales to obtain smoother results.

3.5. Multiscale saliency integration

We use four scales in this work, which are defined as L1, L2, L3 and L4, in which the number of superpixels is 200, 250, 300 and 350, respectively. We also complete three stages of saliency calculation, which are the background-based saliency, the foreground-based saliency and the saliency diffusion. During each stage, we calculate saliency in each scale simultaneously. Fig. 5 shows the results of each stage in the L2 level and the final integration, which is the average value of the third stage. As shown in Fig. 5e), the final saliency maps are smoother after the multiscale integration and more similar to the ground truth.

The main steps are summarized in Algorithm 1.

4. Experimental analysis

To evaluate the performance of our method, we compare our experimental results with 22 state-of-the-art saliency detection methods on five benchmark datasets: BFS [30], CA [19], COV [15],

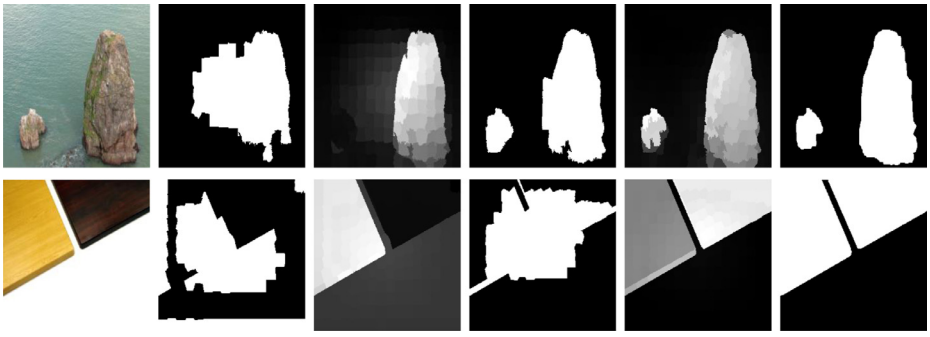


Fig. 4. Foreground-based saliency. From left to right: input image, foreground queries via MR [25], foreground-based saliency map via MR [25], foreground queries via the proposed method, foreground-based saliency map via the proposed method and ground truth.

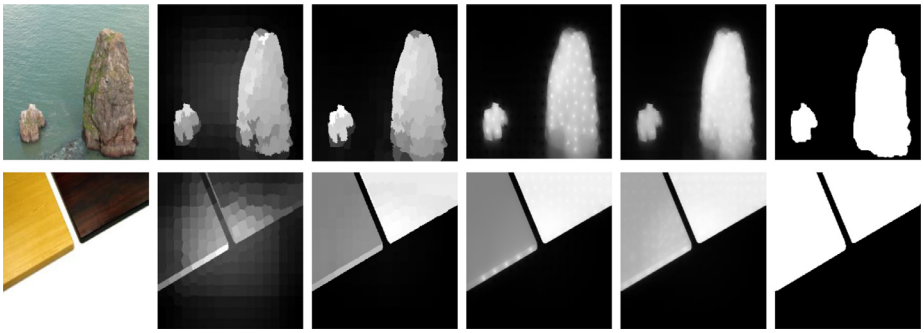


Fig. 5. The result comparisons of each stage via the proposed method. From left to right: (a) input images; (b)–(d) saliency maps of background-based, foreground-based and regularized random walk in L2 scale, respectively; (e) the saliency maps of multiscale integration; (f) ground truth.

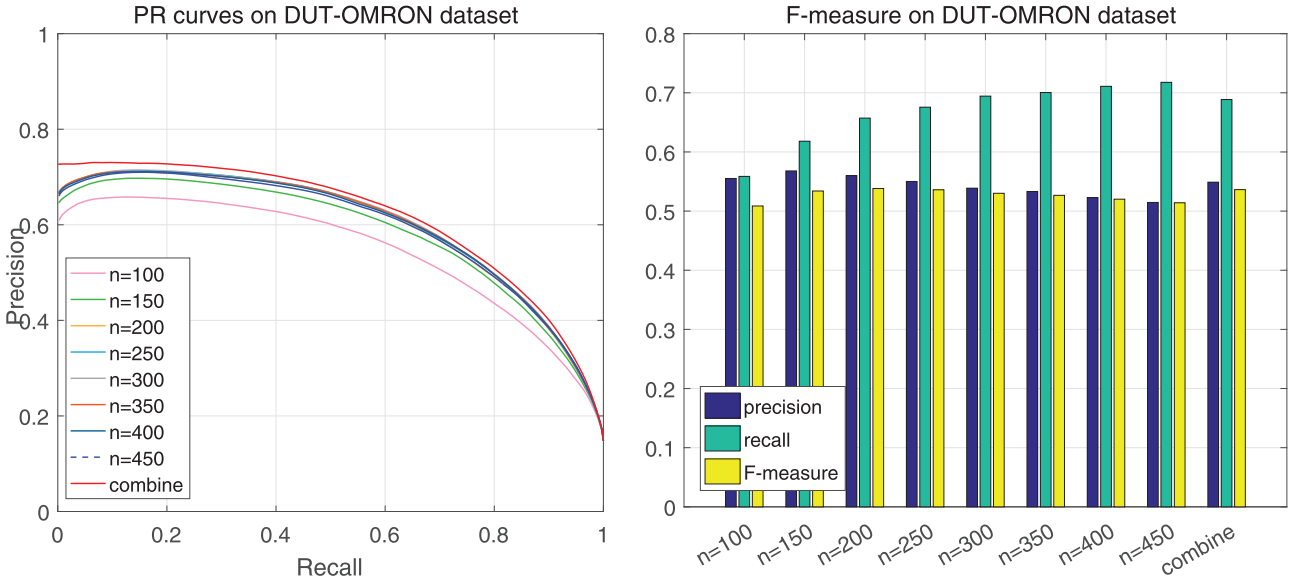


Fig. 6. Evaluation results of different scales on DUT-OMRON. *n* indicates the number of superpixels.

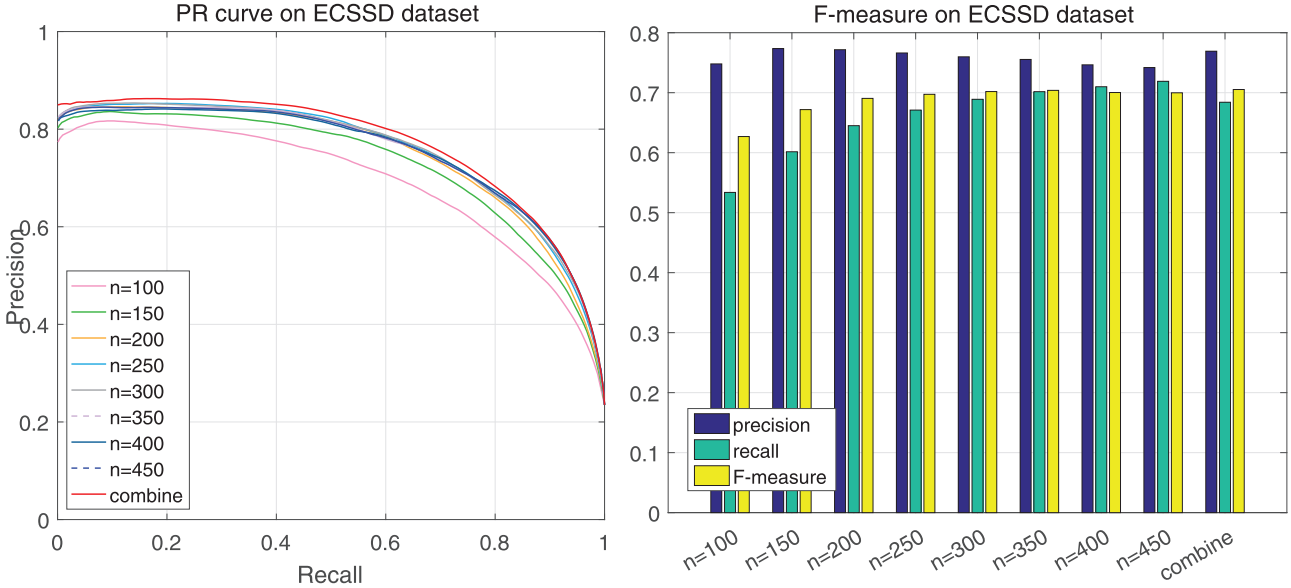


Fig. 7. Evaluation results of different scales on ECSSD. *n* indicates the number of superpixels.

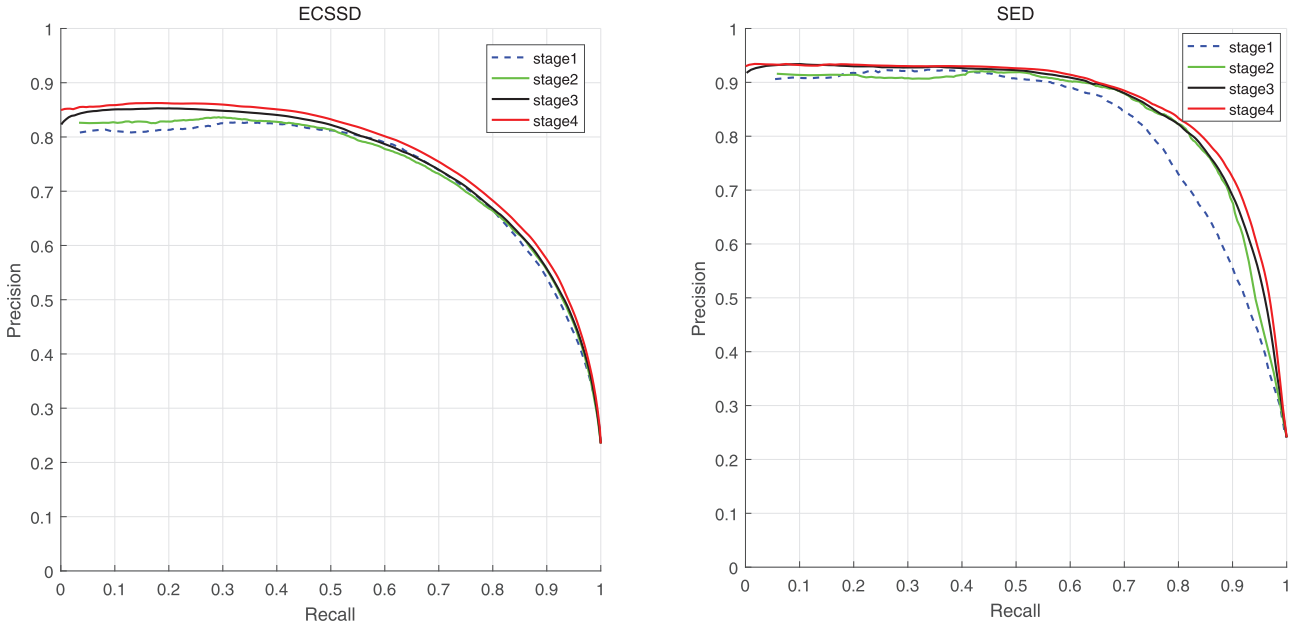


Fig. 8. PR curves of the proposed method on each stage.

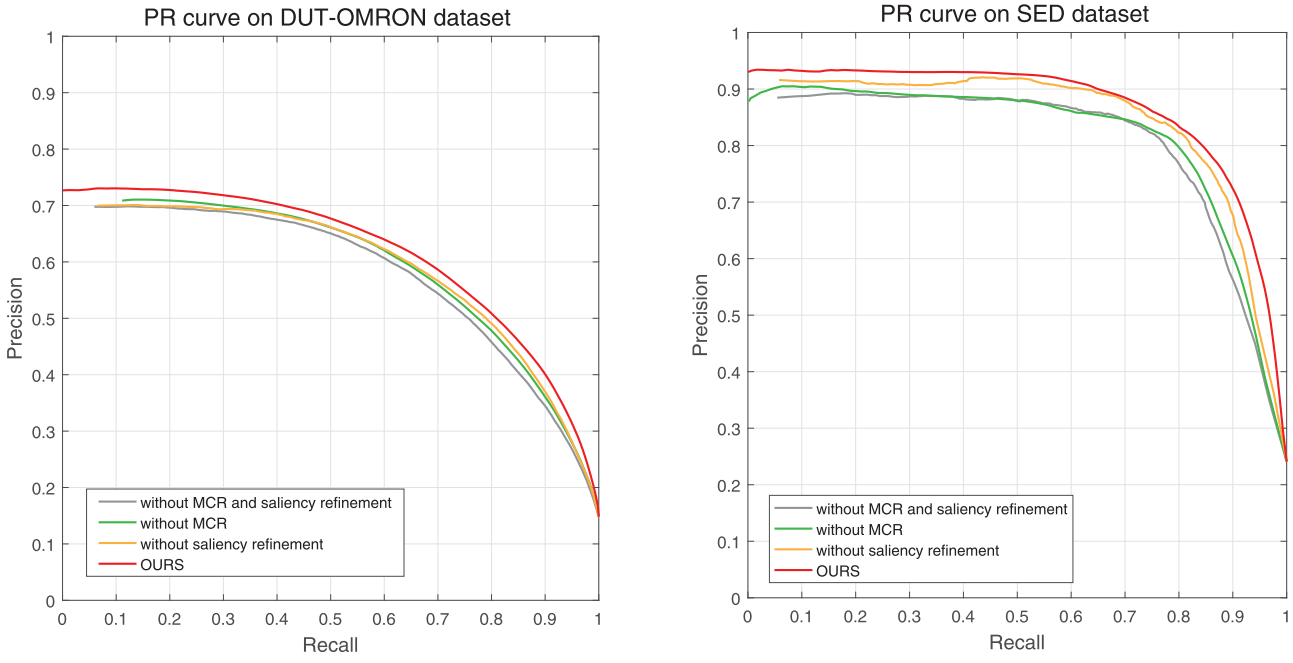


Fig. 9. PR curves of our method with different components on DUT-OMRON and SED. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

FES [20], GS [28], HS [33], LMLC [27], MC [34], MAP [35], MR [25], MS [45], PCA [21], RCRR [38], RRWR [37], SEG [10], SeR [9], SF [50], SUN [14], SWD [16], GP [51], LPS [52], PRMGR [53], DS [54] and UCF [55]. Their codes are mostly available online, and we run them to acquire the results except COV [15], FES [20], SEG [10], SeR [9], SUN [14] and SWD [16], for which we directly use the results reported by Cheng et al. [56]. For PRMGR [53], DS [54] and UCF [55] methods, we also use the results provided by authors.

4.1. Datasets and settings

The datasets used in our work are as follows: DUT-OMRON [25], SED [57], PASCAL-S [58], ECSSD [59] and SOD [60].

DUT-OMRON [25] is composed of 5168 images with pixelwise ground truth. Images in this dataset are manually selected from 140,000 pictures of natural scenes with complex backgrounds, which leads to a considerable challenge. SED [57] contains 200 natural pictures and their corresponding ground truth. The dataset is divided into SED1 with one salient object and SED2 with two salient objects, each of which consists of 100 images. Salient objects in SED2 are often located separately and far from the image center. PASCAL-S [58] contains 850 images of natural scenes, which are selected from PASCAL VOC 2010. A multitude of images in this dataset have complex backgrounds and five targets per image. ECSSD [59] contains 1000 images with semantically rich and complex backgrounds, which are selected from the BSD dataset, Pascal VOC

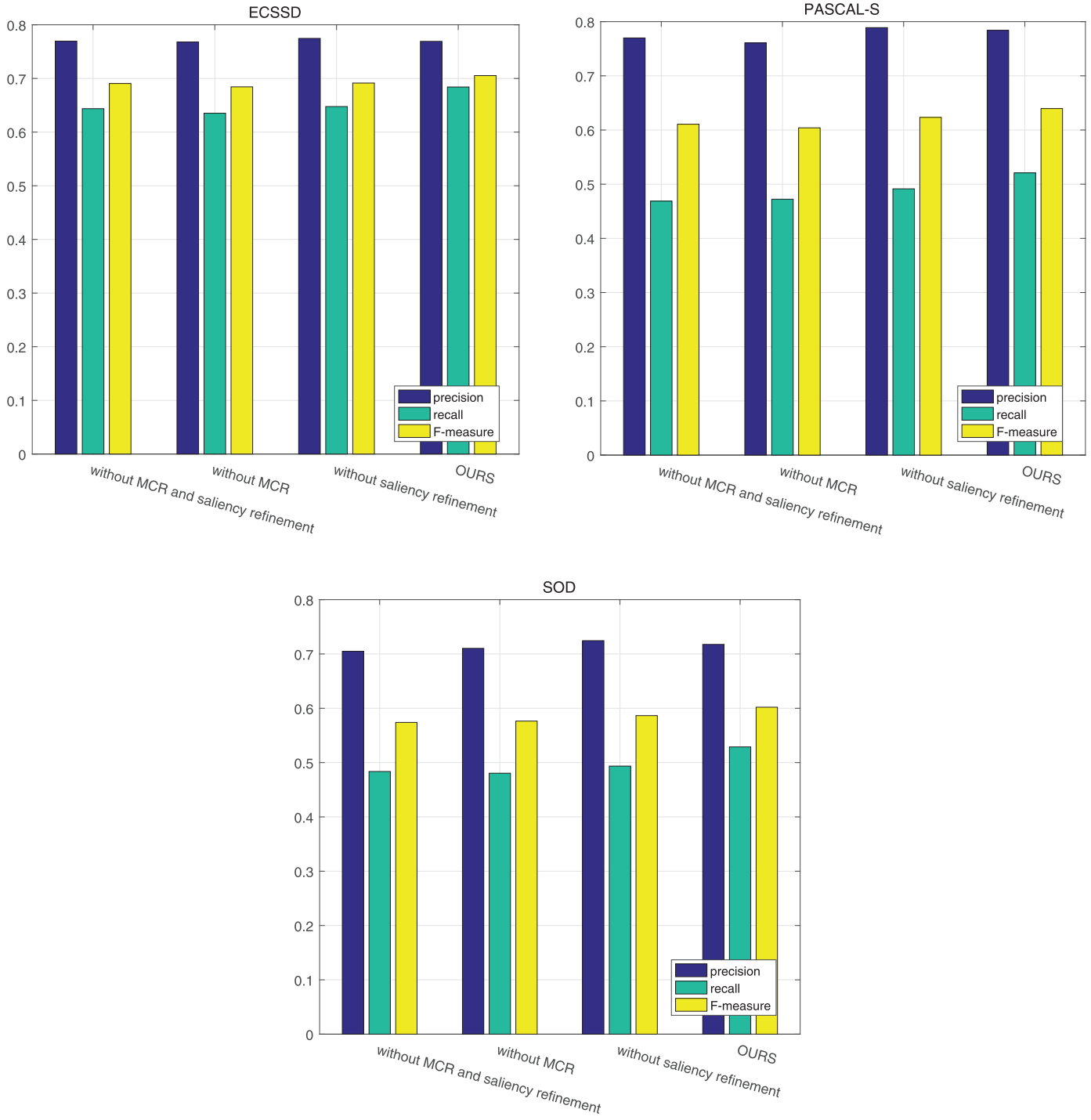


Fig. 10. *F*-measure values of our method with different components on ECSSD, PASCAL-S and SOD.

and the Internet. SOD [60] contains 300 images with salient objects usually occupying most of the area of the image and touching the image boundary.

All the experiments are run in the MATLAB platform on a PC with Intel i5-7500 CPU (3.4 GHZ) and 16 GB RAM. The parameters used in the proposed method, σ^2, σ_i^2, α and μ are set to 1/10, 1/13, 0.99 and 0.01, respectively, in all the experiments.

4.2. Evaluation metrics

We use precision-recall curves (PR curves) and *F*-measure evaluation metrics. To obtain the PR curve, we binarize the saliency maps with thresholds ranging from 0 to 255. Saliency values under the corresponding threshold are masked out to retain 256 binary masks. Then, we compare them with the ground truth to compute precision and recall at each threshold. To extensively evaluate these methods, we also use the *F*-measure metric with adaptive thresholds [61]. The *F*-measure value is the weighted average of precision and recall, which is defined as follows:

$$F_\gamma = \frac{(1+\gamma)\,precision * recall}{\gamma\, precision + recall} \tag{7}$$

We set $\gamma = 0.3$ [61] to pay more attention to precision.

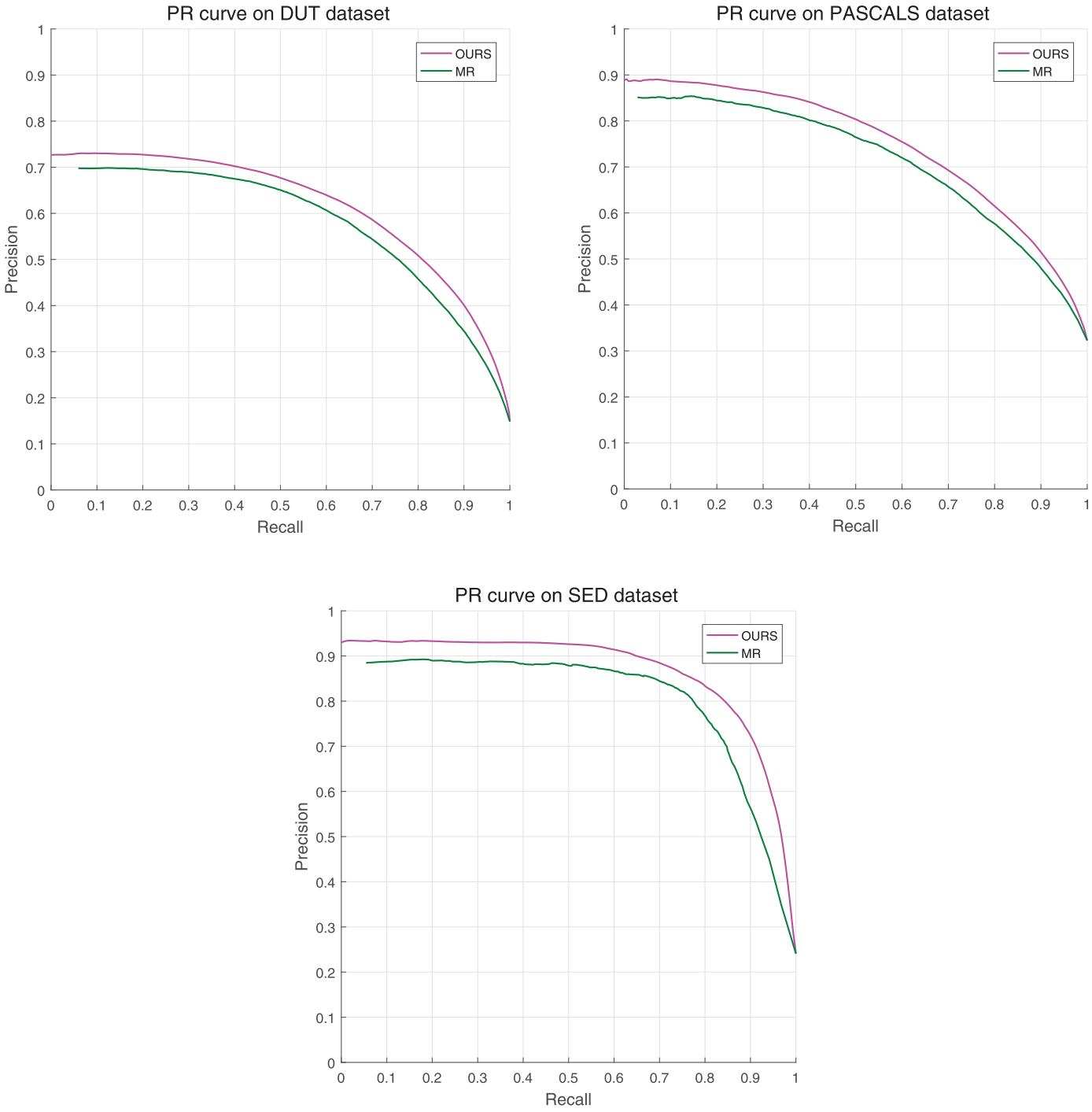


Fig. 11. PR curves of MR and our method on DUT-OMRON, SED and SOD.

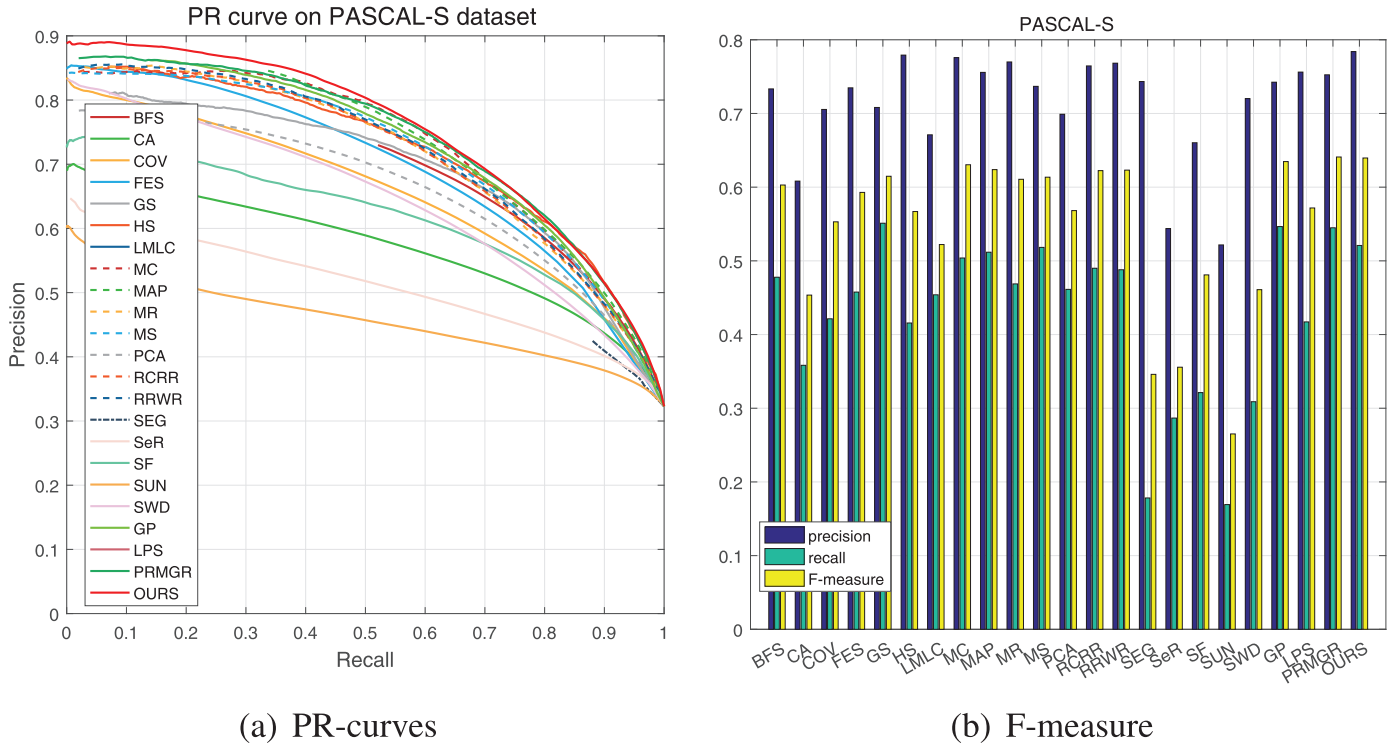
4.3. Components analysis

4.3.1. Single scale vs. multiscale

We run the proposed method on different scales with different numbers of superpixels. We select from 100 to 450 superpixels with an increment of 50. Figs. 6 and 7 show the results on DUT-OMRON and ECSSD, respectively. The results from 100 and 150 are unfavorable, while those from 200 to 450 are slightly different. The combined curve indicates the average results according to the four scales, i.e., 200, 250, 300 and 350 superpixels. In fact, we tried to combine the results of two and three scales in our experiments. Finally, we chose these four scales because their combination achieves the best performance.

4.3.2. Stage contribution

The proposed method has four stages. We validate their efficiency on three datasets, i.e., ECSSD, SED and PASCAL-S. Fig. 8 shows the PR curves together. Stage 1 and Stage 2 refer to the background and foreground saliency detection in L2, respectively. Stage 3 is the refined results via the regularized random walk in L2, and Stage 4 is the combined result of multiple scales. The results show that with MR and regularized random walk, we



(a) PR-curves (b) F-measure

Fig. 12. Evaluation results on the PASCAL-S dataset.

Algorithm 1 Saliency detection via progressive graph rankings.

Input: An image and the related parameters: σ^2, σ_l^2, α and μ;
Output: The corresponding saliency map;

Segment the input image using SLIC;¡¡¡¡
Construct the initial graph G^b;
Construct the extended grpah G^e and use Eq(4) to obtain the first step saliency $\mathbf{s}_b$;
Initialize the foreground query $\mathbf{q}$, i.e., all elements are set to 0;
for each superpixel i **do**
  Set threshold $t = \frac{1}{N}\sum_{j=1}^{N}\mathbf{s}_b(j)$, where N is the total pixel number.¡¡¡¡
  if $\mathbf{s}_b(i) \geq t$ **then**¡¡¡¡¡¡¡¡
    Set $\mathbf{q}_i = 1$ and regard superpixel i as a query;¡¡¡¡
  end if
end for
Compute the second step saliency $\mathbf{s}^f$ with Eq(5) with graph G^b;
Refine the saliency map with Eq(6);
Use multiscale to obtain the final saliency map.

obtain a more stable saliency map, which is further improved via the multiscale integration.

4.3.3. Component analysis of our method

To evaluate the components in our method, we do experiments on five datasets under different circumstance. Fig. 9 gives the PR curves on DUT-OMRON and SED, which shows that our method is better than other combinations. The gray curve is the PR curve without using MCR and saliency refinement; The green curve is the PR curve without using MCR; The orange curve is the PR curve without using saliency refinement; The red curve is the result of saliency detection via progressive graph rankings. Fig. 10 shows the *F*-measure results of our model comparing with above combinations on ECSSD, PASCAL-S and SOD, which also shows that our method is better than other combinations.

Table 1
Comparison of *F*-measure on PASCAL-S dataset.

Methods	LMLC	MAP	MR	PCA	RCRR	SF
F-measure	0.522	0.624	0.611	0.568	0.622	0.481
Methods	GP	LPS	PRMGR	DS	UCF	OURS
F-measure	0.635	0.572	0.641	0.702	0.798	0.640

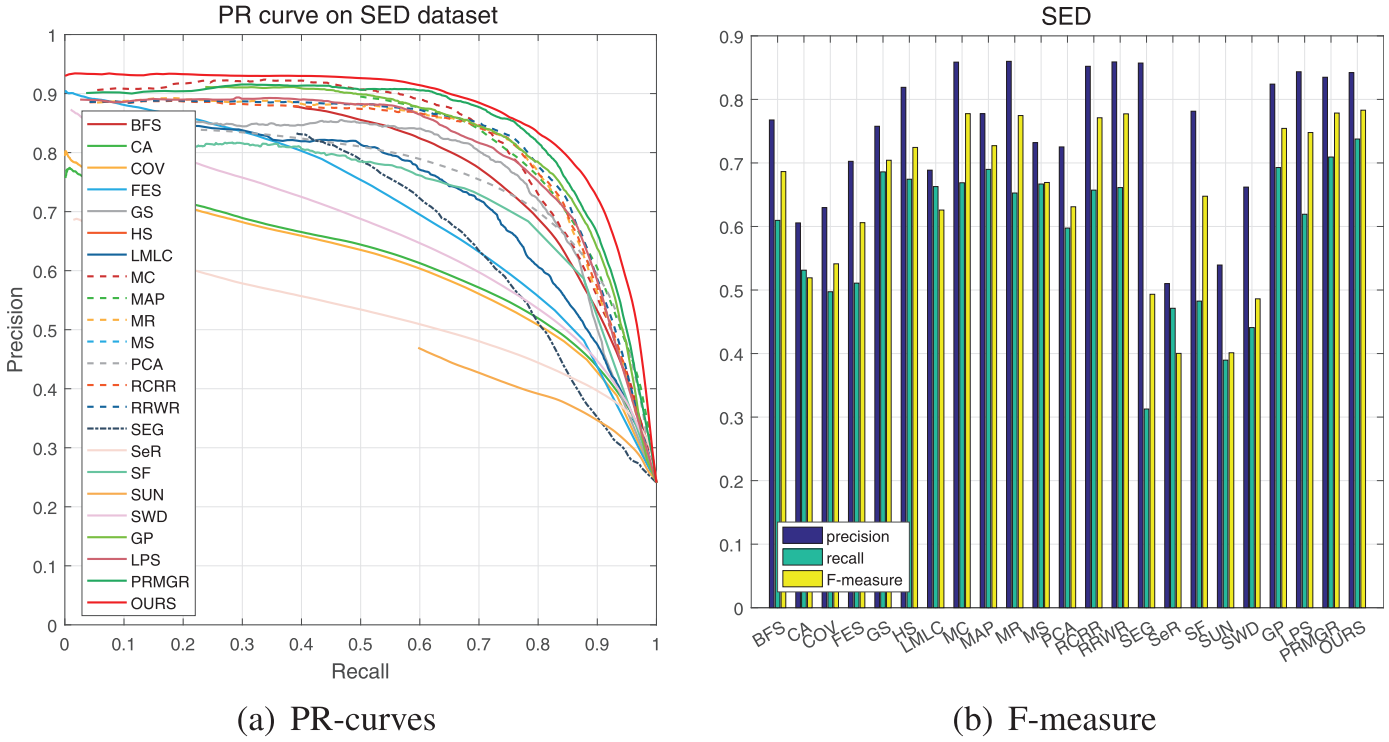
Table 2
Comparison of *F*-measure on SOD dataset.

Methods	LMLC	MAP	MR	PCA	RCRR	SF
F-measure	0.488	0.585	0.574	0.538	0.579	0.425
Methods	GP	LPS	PRMGR	DS	UCF	OURS
F-measure	0.581	0.525	0.586	0.703	0.757	0.602

4.4. Quantitative evaluation

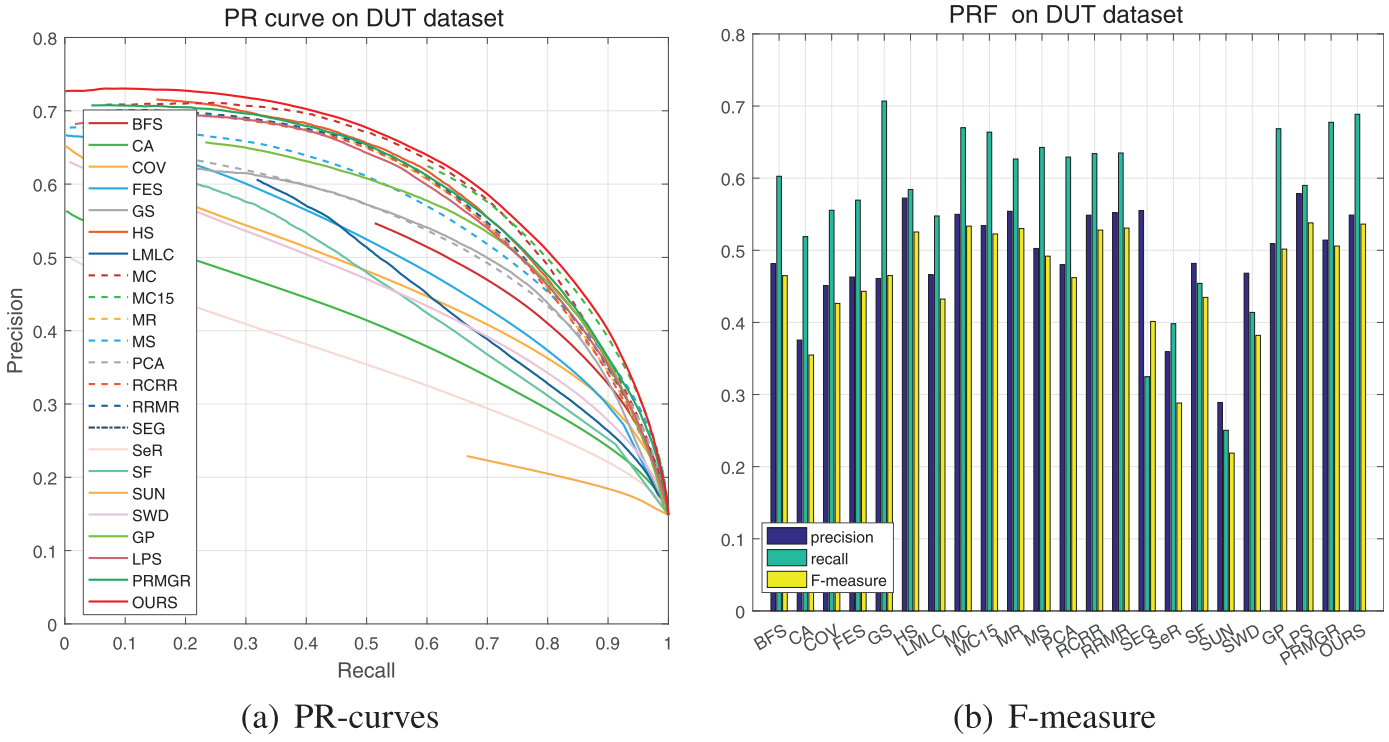
We first compare our proposed method with MR [25] to observe whether it performs better. Fig. 11 shows the PR curves on three datasets, i.e., DUT-OMRON, PASCAL-S and SED. Neither method performs well enough on DUT-OMRON, which also confirms that it is a challenging dataset. We also examine the computation complexity of both methods. Our method contains background-based ranking, foreground-based ranking and saliency refinement. Also we explore multiscale detection. Therefore, the final computation complexity is a little higher than MR [25] which contains two stages of rankings.

Then, we illustrate the quantitative evaluation of the five datasets in Figs. 12–16. In each figure, the left part shows the comparison of PR curves, while the right part shows precision, recall and *F*-measure values under adaptive thresholds. We can observe that the proposed method achieves the best performance, i.e., the largest area of PR curves and the highest *F*-measure values. The *F*-measure values of 11 state-of-the-art methods on PASCAL-S and SOD are showed in Tables 1 and 2. DS [54] and UCF [55] are



(a) PR-curves (b) F-measure

Fig. 13. Evaluation results on the SED dataset.



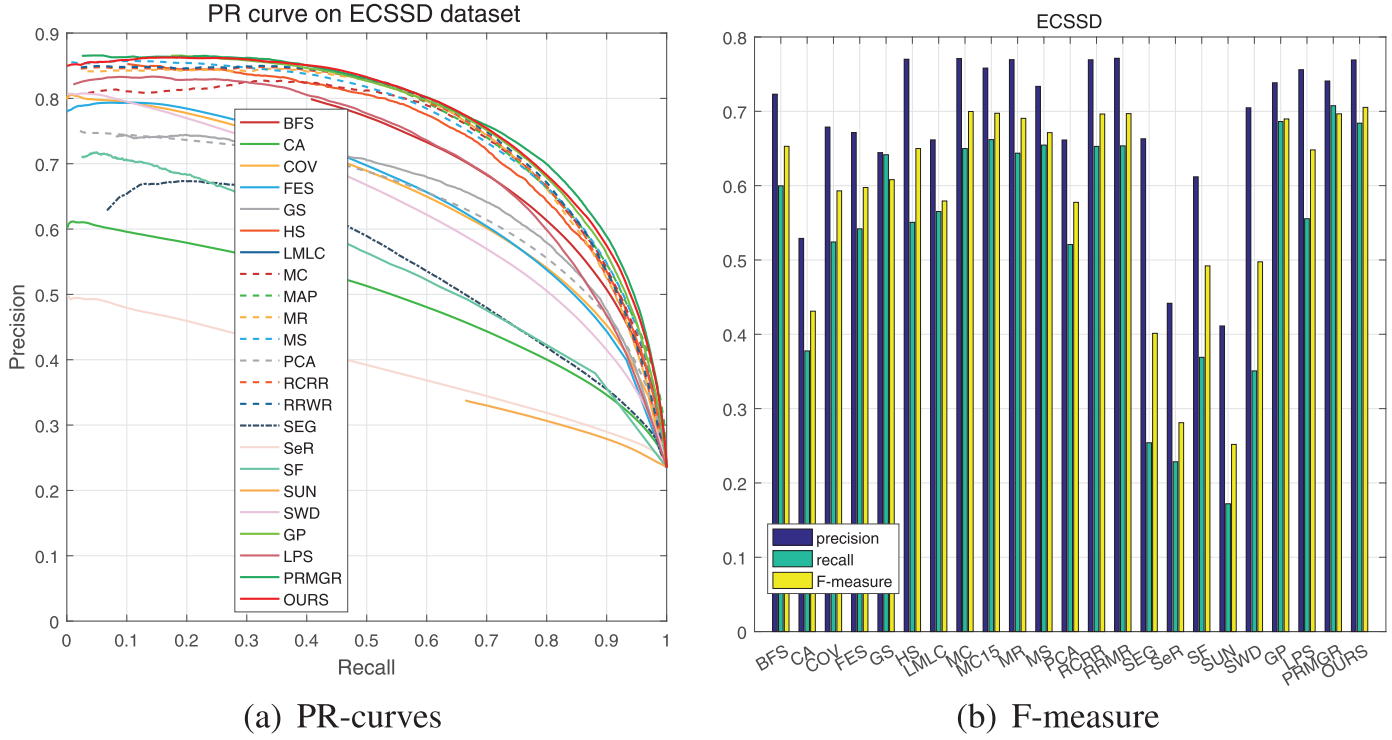
(a) PR-curves (b) F-measure

Fig. 14. Evaluation results on the DUT-OMRON dataset.

supervised methods via deep learning, for which the number of training images are 10,000 and 80,000, respectively. Except for their training time, they obtain higher performance than other unsupervised methods. Among these unsupervised methods, our method performs better than others (PRMGR [53] is as well as ours on PASCAL-S).

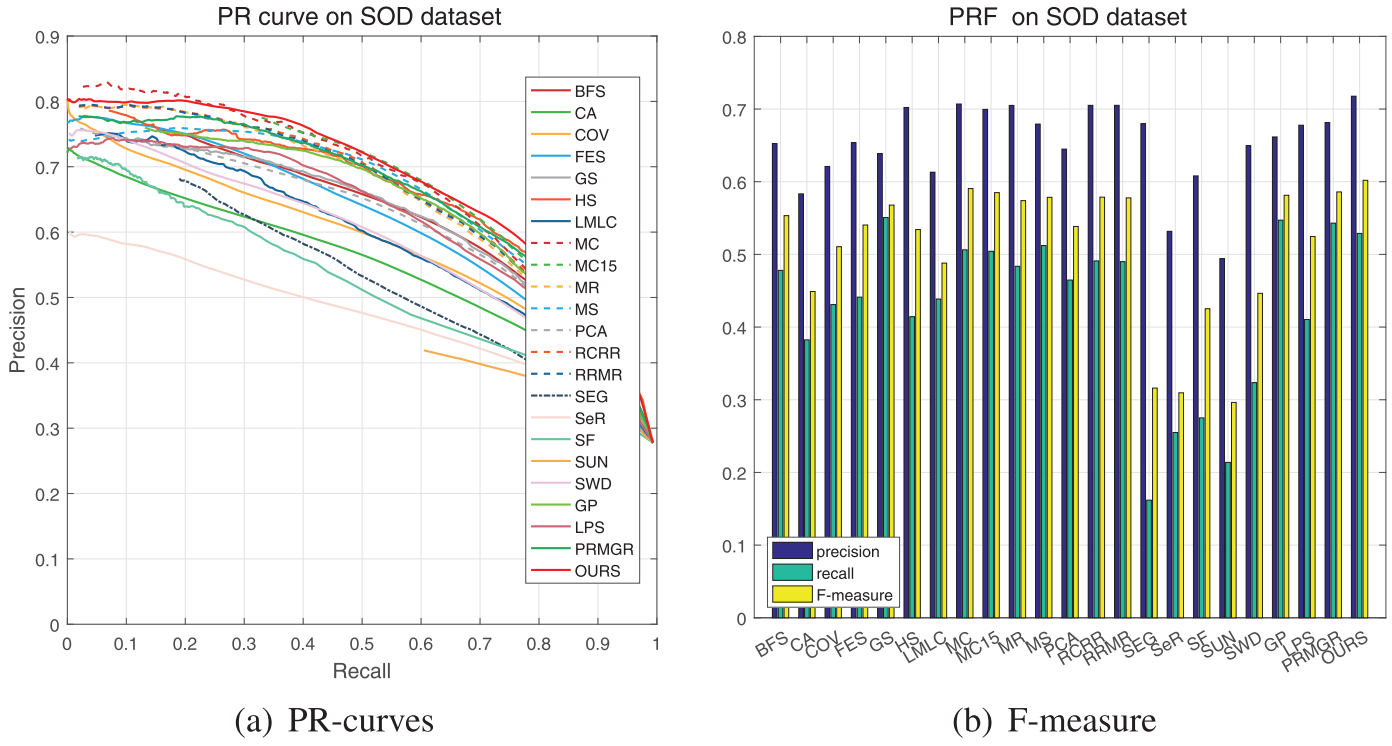
4.5. Qualitative analysis

Fig. 17 shows examples of saliency maps produced by 11 state-of-the-art methods. These methods are selected from the above methods and have demonstrated better results both in PR curves and *F*-measure values.



(a) PR-curves (b) F-measure

Fig. 15. Evaluation results on the ECSSD dataset.



(a) PR-curves (b) F-measure

Fig. 16. Evaluation results on the SOD dataset.

Next, we analyze the saliency maps generated by several related methods. BFS [30] (column d) depends too much on background priors so that it suffers from losing salient parts near image boundaries. MAP [35] (column e) uses guide filters to obtain smoother results but cannot efficiently suppress the background (7e, 11e and 14e). Meanwhile, this method also fails to fully detect salient objects when they are near the image boundary (5e, 6e, 7e and 9e). MR [25](column n), in some cases as a result of the single scale and improper foreground queries, obtains the reverse results (7n, 8n and 13n). Furthermore, this method fails to work for objects with small size (1n, 2n and 3n). RCRR [38] and RRWR [37] (columns k and l) both utilize regularized random walk to diffuse the saliency values, enabling these methods to obtain smoother results than MR [25](column n). In summary, our

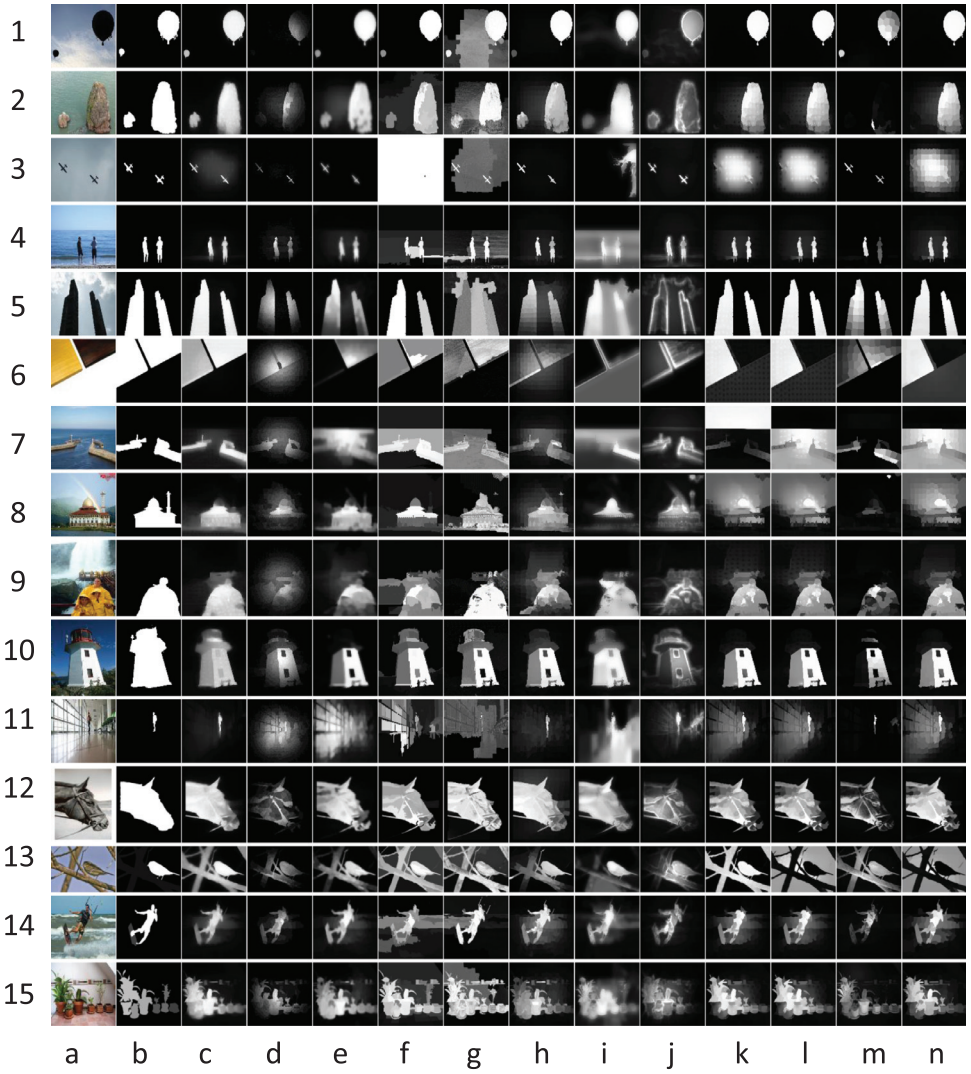


Fig. 17. Examples of saliency maps. (a) input images, (b) ground truth, (c) our results, (d) BFS [30], (e) MAP [35], (f) HS [33], (g) LMLC [27], (h) MC [34], (i) MS [45], (j) PCA [21], (k) RCRR [38], (l) RRWR [37], (m) SF [50] and (n) MR [25]. Images with two salient objects from the 1st to the 7th row are selected from SED. Some of the salient objects are far from the image center, while others are small in size. Especially in the 1st and 2nd input images, the two salient objects have largely different sizes. From the 8th to 11th rows, the input images are chosen from DUT-OMRON and the rest are from PASCAL-S.

results have the highest quality, i.e., they are generally closer to the ground truth.

5. Conclusions

We construct a novel progressive graph rankings model for pixelwise salient object detection. To provide robust foreground queries, we explore the extended graphs and utilize absorbing Markov chain ranking to obtain the background-based saliency maps. Next, we utilize the basic graphs with new edge weights to obtain the foreground-based saliency maps. The above two stages are based on superpixels, which cause the blocking phenomenon in saliency maps. Thus, in the third stage, the pixel-level graph and regularized random walk are utilized to diffuse the saliency values. Finally, we integrate the four results of stage three to smoothen the final saliency map. Quantitative and qualitative analyses show that the proposed method achieves better performance than other start-of-the-art methods. In future work, we will study how to incorporate other middle- and/or high-level features effectively.

Acknowledgment

The authors wish to thank the anonymous reviewers who helped improve the quality of the paper. This study was partially funded by the National Natural Science Foundation of China (61602001, 61602006), the Natural Science Foundation of Anhui Higher Education Institution of China (KJHS2018B06, KJ2015ZD44), the Open Research Project of Anhui Tourism Talent Cultivation Demonstration Base (YYRCYB1703) and the Research Center of Cognitive Computing, Anhui University. A. Hussain was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant no. EP/M026981/1.

References

[1] A. Li, X. She, Q. Sun, Y. Wang, X. Yi, Color image quality assessment combining saliency and FSIM, in: Proceedings of the SPIE – The International Society for Optical Engineering, 2013, p. 88780I.
[2] Z. Ren, S. Gao, L.T. Chia, W.H. Tsang, Region-based saliency detection and its application in object recognition, IEEE Trans. Circuits Syst. Video Technol. 24 (5) (2014) 769–779.

[3] F. Gao, J. You, J. Wang, J. Sun, E. Yang, H. Zhou, A novel target detection method for SAR images based on shadow proposal and saliency analysis, Neurocomputing 267 (C) (2017) 220–231.
[4] C. Li, X. Wang, L. Zhang, J. Tang, H. Wu, L. Lin, Weld: weighted low-rank decomposition for robust grayscale-thermal foreground detection, IEEE Trans. Circuits Syst. Video Technol. 27 (4) (2017) 725–738.
[5] C. Guo, L. Zhang, A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression, IEEE Trans. Image Process. 19 (1) (2009) 185–198.
[6] S. Stalder, H. Grabner, L.V. Gool, Dynamic objectness for adaptive tracking, in: Proceedings of the Asian Conference on Computer Vision (ACCV), 2012, pp. 43–56.
[7] C. Li, H. Cheng, S. Hu, X. Liu, J. Tang, L. Lin, Learning collaborative sparse representation for grayscale-thermal tracking, IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc. 25 (12) (2016) 5743–5756.
[8] C. Li, L. Lin, W. Zuo, J. Tang, M.-H. Yang, Visual tracking via dynamic graph learning, IEEE Trans. Pattern Anal. Mach. Intell. (2018), doi:10.1109/TPAMI.2018.2864965.
[9] H.J. Seo, P. Milanfar, Static and space-time visual saliency detection by self-resemblance, J. Vis. 9 (12) (2009) 1–27.
[10] E. Rahtu, J. Kannala, M. Salo, J. Heikkil, Segmenting Salient Objects from Images and Videos, Springer, Berlin Heidelberg, 2010.
[11] W. Wang, J. Shen, R. Yang, F. Porikli, Saliency-aware video object segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 40 (1) (2018a) 20–33.
[12] W. Wang, J. Shen, L. Shao, Video salient object detection via fully convolutional networks, IEEE Trans. Image Process. 27 (1) (2018b) 38–49.
[13] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1254–1259, doi:10.1109/34.730558.
[14] L. Zhang, M.H. Tong, T.K. Marks, H. Shan, G.W. Cottrell, Sun: a Bayesian framework for saliency using natural statistics, J. Vis. 8 (7) (2008) 32.1–20.
[15] E. Erdem, A. Erdem, Visual saliency estimation by nonlinearly integrating features using region covariances., J. Vis. 13 (4) (2013) 11.
[16] L. Duan, C. Wu, J. Miao, L. Qing, Y. Fu, Visual saliency detection by spatially weighted dissimilarity, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 473–480.
[17] S. Li, C. Zeng, S. Liu, Y. Fu, Merging fixation for saliency detection in a multilayer graph, Neurocomputing 230 (C) (2016) 173–183.
[18] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, S. Li, Automatic salient object segmentation based on context and shape prior, in: Proceedings of the British Machine Vision Conference (BMVC), 2011, pp. 1–12.
[19] S. Goferman, L. Zelnik-Manor, A. Tal, Context-aware saliency detection, IEEE Trans. Pattern Anal. Mach. Intell. 34 (10) (2012) 1915–1926.
[20] H.R. Tavakoli, E. Rahtu, J. Heikkil, Fast and efficient saliency detection using sparse sampling and kernel density estimation, in: Proceedings of the Scandinavian Conference on Image Analysis (SCIA), 2011, pp. 666–675.
[21] M. Ran, A. Tal, L. Zelnikmanor, What makes a patch distinct? in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 1139–1146.
[22] M.-M. Cheng, N.J. Mitra, X. Huang, P.H.S. Torr, S.-M. Hu, Global contrast based salient region detection, IEEE Trans. Pattern Anal. Mach. Intell. 37 (3) (2015) 569–582, doi:10.1109/TPAMI.2014.2345401.
[23] L. Zhang, Q. Zhou, Salient object detection via proposal selection, Neurocomputing 295 (2018) 59–71, doi:10.1016/j.neucom.2018.01.050.
[24] N. Tong, H. Lu, Y. Zhang, R. Xiang, Salient object detection via global and local cues, Pattern Recognit. 48 (10) (2015) 3258–3267.
[25] C. Yang, L. Zhang, H. Lu, X. Ruan, M.H. Yang, Saliency detection via graph-based manifold ranking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013a, pp. 3166–3173.
[26] C. Yang, L. Zhang, H. Lu, Graph-regularized saliency detection with convex-hull-based center prior, IEEE Signal Process. Lett. 20 (7) (2013b) 637–640.
[27] Y. Xie, H. Lu, M.H. Yang, Bayesian saliency via low and mid level cues., IEEE Trans. Image Process. 22 (5) (2013) 1689–1698.
[28] Y. Wei, F. Wen, W. Zhu, J. Sun, Geodesic saliency using background priors, in: Proceedings of the European Conference on Computer Vision (ECCV), Springer, 2012, pp. 29–42.
[29] X. Li, H. Lu, L. Zhang, R. Xiang, Saliency detection via dense and sparse reconstruction, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2013, pp. 2976–2983.
[30] J. Wang, H. Lu, X. Li, N. Tong, W. Liu, Saliency detection via background and foreground seed selection, Neurocomputing 152 (C) (2015) 359–368.
[31] W.C. Tu, S. He, Q. Yang, S.Y. Chien, Real-time salient object detection with a minimum spanning tree, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2334–2342.
[32] V. Gopalakrishnan, Y. Hu, D. Rajan, Random walks on graphs for salient object detection in images, IEEE Trans. Image Process. 19 (12) (2010) 3232–3242.
[33] Q. Yan, L. Xu, J. Shi, J. Jia, Hierarchical saliency detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 1155–1162.
[34] B. Jiang, L. Zhang, H. Lu, C. Yang, M.H. Yang, Saliency detection via absorbing Markov chain, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2013, pp. 1665–1672.
[35] J. Sun, H. Lu, X. Liu, Saliency region detection based on Markov absorption probabilities, IEEE Trans. Image Process. 24 (5) (2015) 1639–1649.
[36] Q. Wang, W. Zheng, R. Piramuthu, Grab: visual saliency via novel graph model and background priors, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 535–543.
[37] C. Li, Y. Yuan, W. Cai, Y. Xia, Robust saliency detection via regularized random walks ranking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 2710–2717.
[38] Y. Yuan, C. Li, J. Kim, W. Cai, D.D. Feng, Reversion correction and regularized random walk ranking for saliency detection, IEEE Trans. Image Process. 27 (3) (2018) 1311–1322.
[39] W. Wang, J. Shen, Deep visual attention prediction, IEEE Trans. Image Process. 27 (5) (2018) 2368–2378.
[40] W. Wang, J. Shen, L. Shao, F. Porikli, Correspondence driven saliency transfer, IEEE Trans. Image Process. 25 (2016) 5025–5034.
[41] N.D.B. Bruce, C. Catton, S. Janjic, A deeper look at saliency: feature contrast, semantics, and beyond, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 516–524.
[42] Z. Lu, Z. Fu, T. Xiang, P. Han, L. Wang, X. Gao, Learning from weak and noisy labels for semantic segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (3) (2016) 486–500.
[43] J. Chen, J. Chen, H. Lu, Z. Chi, CNN for saliency detection with low-level feature integration, Neurocomputing 226 (C) (2017) 212–220.
[44] W. Wang, J. Shen, L. Shao, Consistent video saliency using local gradient flow optimization and global refinement., IEEE Trans. Image Process. 24 (11) (2015) 4185–4196.
[45] N. Tong, H. Lu, L. Zhang, R. Xiang, Saliency detection with multi-scale superpixels, IEEE Signal Process. Lett. 21 (9) (2014) 1035–1039.
[46] W. Wang, J. Shen, Y. Yu, K.L. Ma, Stereoscopic thumbnail creation via efficient stereo saliency detection, IEEE Trans. Visual. Comput. Gr. PP (99) (2016). 1–1
[47] A. Levinshtein, A. Stere, K.N. Kutulakos, D.J. Fleet, S.J. Dickinson, K. Siddiqi, Turbopixels: fast superpixels using geometric flows, IEEE Trans. Pattern Anal. Mach. Intell. 31 (12) (2009) 2290–2297.
[48] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Ssstrunk, Slic superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2274–2282.
[49] J. Shen, X. Hao, Z. Liang, Y. Liu, W. Wang, L. Shao, Real-time superpixel segmentation by DBSCAN clustering algorithm, IEEE Trans. Image Process. 25 (12) (2016) 5933–5942.
[50] P. Krahenbuhl, Saliency filters: contrast based filtering for salient region detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 733–740.
[51] P. Jiang, N. Vasconcelos, J. Peng, Generic promotion of diffusion-based salient object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 217–225.
[52] H. Li, H. Hu, Z. Lin, X. Shen, B. Price, Inner and inter label propagation: salient object detection in the wild, IEEE Trans. Image Process. 24 (10) (2015) 3176–3186.
[53] Y. Xiao, B. Jiang, Z. Tu, J. Ma, J. Tang, A prior regularized multi-layer graph ranking model for image saliency computation, Neurocomputing 315 (2018) 234–245, doi:10.1016/j.neucom.2018.06.072.
[54] X. Li, L. Zhao, L. Wei, M.-H. Yang, F. Wu, Y. Zhuang, H. Ling, J. Wang, Deepsaliency: multi-task deep neural network model for salient object detection, IEEE Trans. Image Process. (2016), doi:10.1109/TIP.2016.2579306.
[55] P. Zhang, D. Wang, H. Lu, H. Wang, B. Yin, Learning uncertain convolutional features for accurate saliency detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 212–221.
[56] A. Borji, M.-M. Cheng, H. Jiang, J. Li, Salient object detection: a benchmark, IEEE Trans. Image Process. 24 (12) (2015) 5706–5722, doi:10.1109/TIP.2015.2487833.
[57] S. Alpert, M. Galun, R. Basri, A. Brandt, Image segmentation by probabilistic bottom-up aggregation and cue integration, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007, pp. 1–8.
[58] Y. Li, X. Hou, C. Koch, J.M. Rehg, A.L. Yuille, The secrets of salient object segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 280–287.
[59] J. Shi, Q. Yan, L. Xu, J. Jia, Hierarchical image saliency detection on extended cssd., IEEE Transactions on Pattern Analysis and Machine Intelligence 38 (4) (2016) 717–729.
[60] V. Movahedi, J.H. Elder, Design and perceptual validation of performance measures for salient object segmentation, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, pp. 49–56.
[61] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 1597–1604.

Lihua Wang received the M.S. degree in computer science from Hangzhou Dianzi University, Hangzhou, China, in 2009. She is currently a visiting scholar in Anhui University. Her research interests include computer vision and saliency detection.

Bo Jiang received the Ph.D. degree in computer science from Anhui University, Hefei, China, in 2015. He is currently an Associate Professor at computer science at Anhui University. His current research interests include image and graph matching, image feature extraction, and graph-based representation and learning.

Zhengzheng Tu received the Ph.D. degree in computer application technology from the Anhui University, Hefei, China, in 2015. She is currently an Associate Professor in computer science at Anhui University. Her current research interests include computer vision, pattern recognition theory and application and digital image processing and video intelligent analysis.

Amir Hussain received the B.Eng. degree and the Ph.D. degree in Electronic & Electrical Engineering from University of Strathclyde, Scotland, UK, in 1992 and 1997, respectively. Following postdoctoral and academic positions at West of Scotland (1996-98), Dundee (1998-2000) and Stirling Universities (2000-18) respectively, he is currently a Professor and founding Head of the Cognitive Big Data and Cybersecurity (CogBiD) Research Lab at Edinburgh Napier University, U.K. His research interests include cognitive computation, machine learning and computer vision.

Jin Tang received the Ph.D. degree in computer science from the Anhui University, Hefei, China, in 2007. He is currently a Professor in Anhui University. His current research interests include computer vision, machine learning and pattern recognition.