CONTENTS

# Robust Saliency Detection Algorithm via Multi-Level Graph Structure and Accurate Background Queries Selection

Ziyu Shu, Yongan Shu

*a) Abstract:* In the field of saliency detection, many graph-based algorithms use boundary pixels as background seeds to estimate the background and foreground saliency,which leads to significant errors in some of pictures. In addition, local context with high contrast will mislead the algorithms. In this paper, we propose a novel multilevel bottom-up saliency detection approach that accurately utilizes the boundary information and takes advantage of both region-based features and local image details. To provide more accurate saliency estimations, we build a three-level graph model to capture both region-based features and local image details. By using superpixels of all four boundaries, we first roughly figure out the foreground superpixels. After calculating the RGB distances between the average of foreground superpixels and every boundary superpixel, we discard the boundary superpixels with the longest distance to get a set of accurate background boundary queries. Finally, we propose the regularized random walks ranking to formulate pixel-wise saliency maps. Experiment results on two public datasets indicate the significantly promoted accuracy and robustness of our proposed algorithm in comparison with 7 state-of-the-art saliency detection approaches.

*b) Index terms::* saliency detection, manifold ranking, regularized random walks ranking, accurate background queries selection, multi-level graph structure.

.

Anhui Provincial Natural Science Foundation founded project (1408085MF125).

## I. Introduction

One of the most important capabilities of our human visual system is to figure out salient objects from a complicated visual scene. Such capability is also very important for computational visual systems. By using this capability, visual systems can identify and process the most salient objects at first, which can extraordinary save the time and get rid of the information overload problem. Saliency detection is used to imitate the human visual system to identify the most salient parts of an image and neglect the remaining parts. It has been widely applied to numerous vision problems including image segmentation [1], object recognition [2], image compression [3], content based image retrieval [4].

In computer vision, both bottom-up models [1, 5–10] and top-down models [4, 11–13] can detect a salient object. Top-down models analyze task-driven visual attention, which often entail supervised learning with class labels from a large set of training sets [13, 14]. Bottom-up models are fast, data-driven, and pre-attentive [10] that always model saliency by visual distinctness or rarity using low-level image information such as contrast, color, texture, and boundary. Bottom-up models are usually faster to execute and easier to adapt to various images than top-down models. In this paper, we propose a bottom-up model to detect salient objects in images.

Existing methods always use visual cues of foreground objects for saliency detection, e.g., color [5, 15], distinct patterns [16], focuses [17]. Recently, methods using the background cues, especially using the boundaries of images to detect salient objects have been developed [10, 18, 19]. These methods use three or four boundaries of the images as background queries to detect salient object. As for saliency detection, an image is represented by a set of nodes to be labeled, and the labeling task is transformed to an energy minimization problem [20, 21] or a random walks problem [22].

We observe that although background regions usually contain image boundaries, using boundaries as background queries may also cause some mistakes. On the one hand, salient objects may touch one or two boundaries (such as a portrait) and cause mistakes when using all four boundaries as background queries (Figure 1), on the other hand, existing methods sometimes select wrong boundaries (Figure 2), which makes things even worse. In this work, we first use all four boundaries as background queries to roughly estimate the salient objects, then we calculate the RGB distances between every boundary superpixel and salient objects

Ziyu Shu is with NYU Langone Medical Center, New York University, USA, email:zs919@nyu.edu

Yongan Shu is the corresponding author and he is with the Computer Science and Technology Institute, Anhui University, Anhui 230039, China, email: shuya@mail.ustc.edu.cn.

to get the accurate background queries. With a set of accurate background queries, we can figure out salient objects better than previous boundary selection methods.

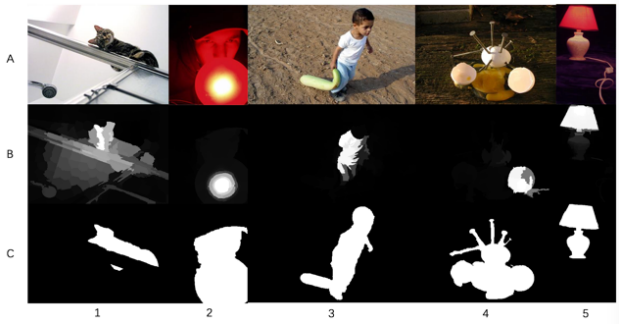


Fig. 1. The shortcoming of using all boundaries as background queries. Row A: Original image. Row B: Solutions using all four boundaries as background. Row C: Ground Truth. In the picture, we can see that using all four boundaries may wrongly regard a part of background as a salient object (B1) or lose parts of a salient object (B2, B3, B4, B5).

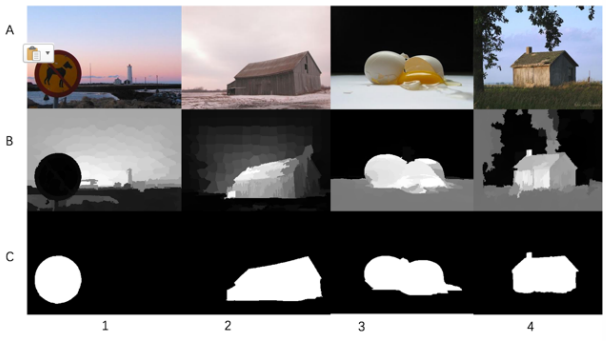


Fig. 2. The shortcoming of using three boundaries as background queries. Row A: Original image. Row B: Solutions using three boundaries as background. Row C: Ground Truth. In the picture, we can see that this method may wrongly regard large area of background as salient objects (B2, B3) or even come up with a totally wrong solution (B1, B4).

We also observe that the local contrast and textures play important roles in saliency detection. When the color of salient object is similar to the background, local contrast and texture like edges can help us to figure out the salient object. But sometimes, they also mislead algorithms into neglecting the salient object, e.g., eyes on people's face, local textures on objects' surface (Figure 3). To deal with this problem, we construct a multi-level graph structure to simultaneously capture both local and global structure information of an image, where we use Gaussian filter to smooth the input image on different levels. With a multi-level graph structure, our algorithm can not only get information from edges and textures, but also get rid of the bad influence of the high local contrast and textures.

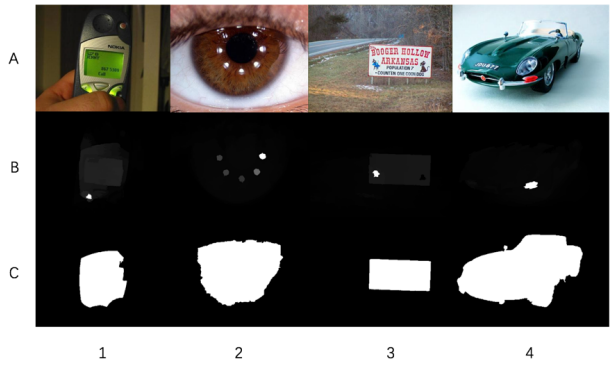


Fig. 3. The error caused by local textures. Row A: Origin image. Row B: Error solutions caused by local textures. Row C: Ground Truth.

Specifically, we use both manifold ranking and regularized random walks ranking [10, 18] to improve the overall quality of the saliency

map. Figure 4 shows the main steps of the proposed algorithm. In the first step, we use each of the boundary superpixels as background queries to roughly label the salient object. Then, we calculate the differences between salient object and each of the boundary superpixels and select the superpixels which are most likely to be background queries. With these more accurate background queries, we calculate the saliency map again to improve the output. In the third step, we use the salient object got from the first step as foreground queries to improve the quality of the saliency map. In the fourth step, we use pixel-wise regularized random walks ranking to get a new saliency map. At the last step, we use Otsu method [23] to binarize the saliency map to get the final solution.

The main contributions of this work are summarized as follows:

1) We construct a multi-level graph model to capture the characteristics of both local textures and long-range spatial connections between pairwise pixels.
2) We raise a new method to select background queries more accurately.
3) Experimental results on four benchmark data sets show that the proposed algorithm performs more efficiently and favorably than the state-of-the-art saliency detection methods.

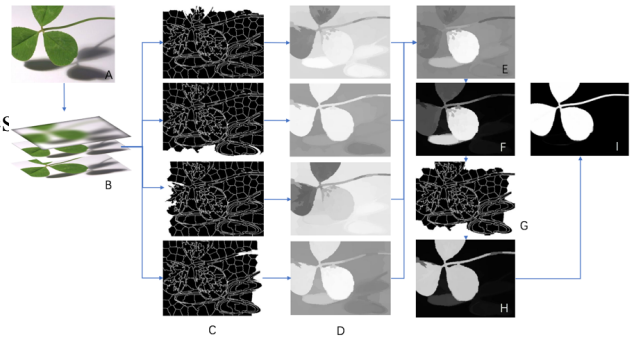


Fig. 4. Main steps of the propose method. A: Input image, B: Three level graph model, C: Using every boundary as background query to estimate saliency map, D: Output of step C, E: Average of step D, F: Estimating foreground saliency map, G: Selecting accurate background queries (white area), H: Calculating background saliency, I: Calculating foreground saliency map, apply Otsu method.

## II. Background

In this section, we will provide a brief review of the manifold ranking model, regularized random walks ranking model as preliminary knowledge.

### A. Manifold Ranking

Manifold ranking is first used in pattern classification [24, 25]. Given a dataset $A = \{a_1, \ldots, a_m, a_{m+1}, \ldots, a_n\}$, where $n$ is the total number of elements, the first $m$ elements are the queries and the rest of them are the unknown elements which we want to rank according to their relevances to the queries. Let $G = (V, E)$ denote the graph structure of $A$, where the node set $V$ denotes every point of the data set $A$, and the edge set $E$ denotes all the connections of any

two nodes in $V$, where $e_{ij} = 1$ means points $a_i$ and $a_j$ are labeled as "connected" and $e_{ij} = 0$ otherwise . From $G = (V, E)$ we can get a matrix $W$ which assigns each pair of points $a_i$ and $a_j$ a weight (distance), such as Euclidean or RGB color distance, the weight of adjacent nodes is defined as:

$$w_{ij} = exp(-\frac{\|a_i - a_j\|^2}{\sigma^2}), \tag{1}$$

where $a_i$ and $a_j$ are the mean CIELAB colors of the two nodes $i$ and $j$, $\sigma$ is a controlling constant. The remaining undefined elements of $W$ are all assigned as zero.

Let output probability values be $f = [f_1, \ldots, f_n]^T$. To get the output, we need to solve the following problem:

$$f^* = arg\min_f \frac{1}{2}\Big\{\sum_{i,j=1}^{n} w_{ij}\Big\|\frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}}\Big\|^2 + \mu\sum_{i=1}^{n}\Big\|f_i - x_i\Big\|^2\Big\}, \tag{2}$$

where we use $X = \{x_1, \ldots, x_n\}$ to indicate the labeled queries. If $a_i$ belongs to the queries, $x_i = 1$ and if not, $x_i = 0$. $\mu$ is a controlling parameter. The solution of this problem is given in [10, 25] as:

$$f^* = \Big(D - \frac{1}{1+\mu}W\Big)^{-1}X, \tag{3}$$

where the degree matrix $D$ is the diagonal matrix with (i, i)-element equal to the sum of the i-th row of $W$:

$$D = diag(d_1 \ldots d_n), \tag{4}$$

where $d_i = \sum_j w_{ij}$

### B. Regularized Random Walks Ranking

Random walk is a mathematical formalization of a random sequence path, which leads an element to a seed location with the highest likelihood [26]. We begin by defining a precise notion for a graph $A = \{a_1, \ldots, a_n\}$ as above, $G = (V, E)$, where $A$ means all the elements of the graph. Then we choose $k$ elements from $A$ as the seed nodes to partition the elements into background nodes and foreground nodes. So what we get is $A = \{a_m^T, a_u^T\}$, in which $a_m^T$ are seed nodes and $a_u^T$ are unknown nodes. We can also get the weight matrix $W$ and the degree matrix $D$ which are similar to those in the section II.A. After that, we define the $n \times n$ Laplacian matrix $L$ as:

$$L_{ij} = \begin{cases} d_i, & if \quad i = j \\ -w_{ij}, & if \quad a_i\, and\, a_j\, are\, neighbor\, nodes \\ 0, & otherwise \end{cases} \tag{5}$$

Especially, we use $i$ and $j$ as element subscripts to indicate pixel-wise graphs other than superpixel-wise graphs.

Similar to Section II.A, we let $G^k = [(g_m^k)^T, (g_u^k)^T]$ denote the probability vector of $A$ for label $k$, which are partitioned into two parts. $(g_m^k)^T$

denotes seed nodes and $(g_u^k)^T$ denotes unknown nodes. We assign $g_m^k = 1$ if $a_m$ belongs to group $k$. The optimized $g^k$ is achieved by minimizing the Dirichlet integral [26],

$$Dir[g^k] = \frac{1}{2}(g^k)^T L(g^k) = \frac{1}{2}[(g_m^k)^T (g_u^k)^T] \begin{bmatrix} L_m & H \\ H^T & L_u \end{bmatrix} \begin{bmatrix} g_m^k \\ g_u^k \end{bmatrix} \quad (6)$$

The regularized random walks ranking is extended from the random walk model [18], we apply an additional constraint on the original Dirichlet integral to let it close to the previous saliency distribution,

$$Dir[g^k] = \frac{1}{2}(g^k)^T L(g^k) + \frac{\mu}{2}(g^k - X)^T (g^k - X), \quad (7)$$

where $\mu$ is the same controlling parameter used in Eq.2. By differentiating $Dir[g^k]$ , we get the following solution,

$$g_u^k = (L_u + \mu I)^{-1}(-H^T g_m^k + \mu X_u^k) \quad (8)$$

## III. The Proposed Algorithm

The proposed saliency detection algorithm consists of five major steps. The first step establishes new graph structure to improve the quality of the following steps. The second step uses all the superpixels on four boundaries as background queries to generate a rough saliency map. The third step removes 25% of the boundary superpixels with the lowest probability belonging to the background and uses the new background queries to generate an accurate saliency map. The forth step generates foreground saliency estimation based on the complementary values of the background saliency map. The fifth step extracts seeds from the forth step and calculates the pixel-wise saliency map with our proposed regularized random walks ranking and applies Otsu method.

### A. Multilevel Graph Model

Salient objects are likely to appear at different scales, furthermore, some of them even contain high contrast area as shown in Figure 3. In that case, we consider to multiply quantize the input image with a multi-level graph structure by applying two Gaussian filters (hsize=[55,55], sigma=15 and hsize=[5,5], sigma=15) to the original image. The lower levels of this hierarchical graph describe more detailed image structures. This will help the algorithm to capture local features like edges and textures. The higher levels encode holistic visual information, which helps the algorithm get rid of the misleading of some extreme local features (See Figure5).

The three-level weight matrix W is established similar to the single-level weight matrix stated in section II.A. As shown in Figure5, the weight of inter level adjacent nodes (Red, blue

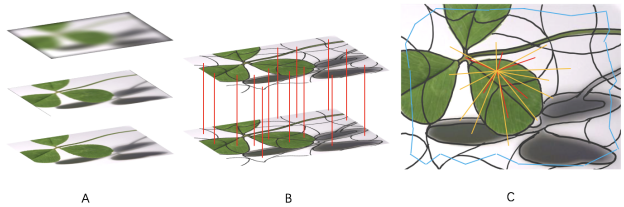


Fig. 5. The graph structure. A: Using Gaussian filter to create a three-level graph structure. B: Intra-level connections,each node is connected to itself in different levels. C: Red: Neighboring nodes with shared edges are connected to each other. Orange: Each node is also connected to the neighbor nodes of its own neighbors. Blue: Any two nodes from the four boundaries of the graph are treated as connected. Purple: Each node is connected to itself in different levels.

and green) is defined by Eq.1, the weight of intra level adjacent nodes (red) is empirically set to 5.

With the connections between levels, the multi-level saliency inference can capture both local and long-range labeling cues of the foreground and background, which highly improve the capability of the proposed algorithm (See Figure6).

The connection matrix $E$ of our multilevel graph model is defined as follow:

1) Each node is connected to its neighboring nodes.
2) Each Node is connected to the nodes sharing common boundaries with its neighboring nodes.
3) Each node is connected to the nodes on the four boundaries.
4) Each node is also connected to itself on different levels.
5) Any pair of the four boundary nodes of the graph are treated as connected.

Then, the weight matrix W [24, 25] is established based on the connection matrix $E$ using Eq.1

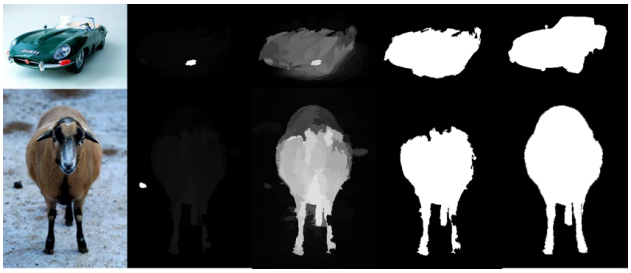

Fig. 6. The improvement of our graph structure. From left to right, original input image, output without using the three-level graph structure, output using the three-level graph structure, output after applying Otsu method, ground truth.

### *B. Background Saliency Estimation*

Previous methods [10, 18, 19, 27] suppose that foreground objects may touch one of the four boundaries, so they discard the boundary with the biggest RGB Euclidean distance. But on the one hand, foreground object may not only touch one of the four boundaries (See Figure7), on the other hand, the boundary with the largest Euclidean distance may not be the right boundary to be discarded (See Figure8). To solve the problem, we propose a new superpixel based method to select background saliency queries accurately. First, instead of calculating the Euclidean distance between any two of the

boundaries and removing one of the four boundaries, we use all the four boundaries as background saliency queries to estimate the foreground object. That is, the superpixels on every boundary will be labeled as one in the indication vector X in Eq.2, while the other superpixels are labeled as zero. Four ranking results $f_l^*$ will be achieved using Eq.3. We use Eq.9 [18] to get the complement foreground based saliency map and Eq.10 to combine all four saliency maps.

$$S_l(i) = 1 - f_l^*(i),\ i = 1, \ldots, n \tag{9}$$

$$S_a(i) = \frac{1}{4} \prod_l S_l(i) \tag{10}$$

We then use Otsu method to binarize the saliency map to roughly label foreground superpixels.

Next, we use the RGB Euclidean distance between every boundary superpixels and the estimated foreground object superpixels to select the query superpixels:

$$diff_l = \sqrt{(\overline{r_l} - \overline{R})^2 + (\overline{g_l} - \overline{G})^2 + (\overline{b_l} - \overline{B})^2}, \tag{11}$$

where , $\overline{r_l}$, $\overline{g_l}$ and $\overline{b_l}$ indicate average R, G and B value of every boundary superpixel respectively, $\overline{R}$, $\overline{G}$ and $\overline{B}$ represent the average R, G and B values of all estimated foreground object superpixels respectively.

At last, we discard 25% of boundary pixels with smallest saliency probability distances and choose the rest of them as accurate background boundary queries. Using these accurate queries, we can get a new saliency map $S_b$ through Eq.3. As shown in Figure7 and Figure8, our new method preforms much better than the existing boundary selection methods.

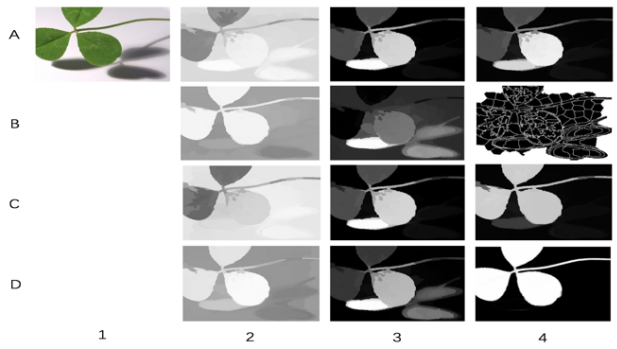


Fig. 7. Accurate boundary superpuxels selection method. 1A: Original input image, the salient object touch both up and left boundaries. 2A, 2B, 2C, 2D: Using only top, bottom, left or right boundary superpixels to generate the saliency map. 3A, 3B, 3C, 3D: The saliency map output using three of the top, bottom, left and right boundary superpixels, none of them are correct. 4A: Estimated saliency map using all four boundary superpixels. 4B: Select the accurate boundary superpixels (white area). 4C: Output saliency map. 4D: Solution after applying Otsu method, the output is just the same as the ground truth in this example.

### C. Foreground Saliency Estimation

We get the foreground saliency from section III.A by complementary subtraction of the background saliency estimation ($S_a$), which may not be accurate enough. Background queries could not always produce enough information for the algorithm to fully figure out the foreground object, especially in cases where the

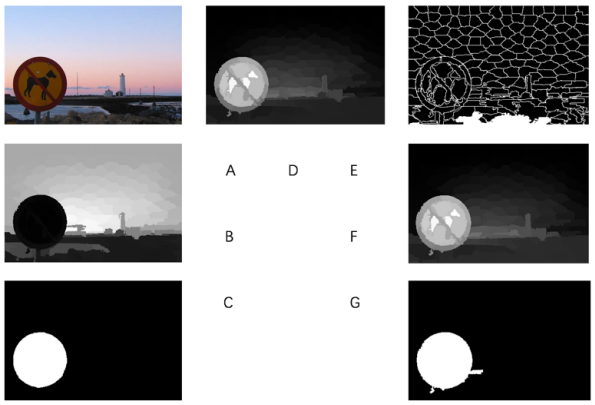


Fig. 8. The improvement of our accurate boundary superpixels selection method. A: Original input image. B: Discard the wrong boundary (top) by using the method stated in section II.C, and the solution is totally wrong. C: Ground truth. D: Estimation saliency map using all four boundary superpixels. E: Discard the accurate boundary superpixels (white area). F: Output saliency map. G: Solution after applying Otsu method, the same as ground truth.

foreground object has complicated structures or textures, or similar patterns to the background. In that case, we propose foreground-query-based saliency estimation subsequently after background-query-based saliency estimation to improve the algorithm [18].

Again, we use Otsu method to extract the foreground queries. Then we can get foreground saliency estimation $S_c$ by using Eq.2 and Eq.3, which will be used in the following regularized random walks ranking method.

### *D. Saliency Map Formulation by Regularized Random Walks Ranking*

The saliency detection solutions mentioned above are based on the output of SLIC superpixel segmentation, which may not perform well in details because the lack of precision of the SLIC segmentation. To overcome these disadvantages, we use a regularized random walks ranking model which is based on every pixel but not superpixel to improve the saliency maps [20].

$$Dir[g^k] = \frac{1}{2}(g^k)^T L(g^k) + \frac{\mu}{2}(g^k - X)^T(g^k - X), \tag{12}$$

where $X$ is an $N \times 1$ pixel-wise indication vector inheriting the values of $S_c$, $L$ is an $N \times N$ matrix defined in Eq.5, $N$ is the total pixel number of the image. To generate pixel-wise matrix $L$ we define the connection matrix $E$ as:

1) Neighboring nodes with shared edges are connected to each other.
2) Each node is also connected to the neighbor nodes of its own neighbors.
3) Any two nodes from the four boundaries of the graph are treated as connected.
4) Each node is connected to boundary nodes on the four sides of the image.

We use two thresholds $t_f$ and $t_b$ as follows to extract foreground and background seeds respectively:

$$t_f = \frac{mean(S_c) + max(S_c)}{2} \tag{13}$$

$$t_b = mean(S_c) \tag{14}$$

That is, the pixels with $X_u > t_f$ are selected as foreground seeds and $X_u < t_b$ as background seeds. The seeds are assigned to $g_M^k$, $k = 1, 2$, where $k = 1$ indicates the background label, and $k = 2$ indicates the foreground label. The optimized solution of Eq.12 is Eq.8

$$g_u^k = (L_u + \mu I)^{-1}(-H^T g_m^k + \mu X_u^k). \quad (15)$$

We set k=2 to select the foreground possibility $g^2$.

Then, we multiply $g^2$ with $S_c$ to combine the superpixel based saliency map and pixel based saliency map as following:

$$S_d = g^k \times S_c \quad (16)$$

### *E. Otsu Binarization*

We use Otsu method [23] to optimized threshold of $S_d$. Figure9 and Figure10 show the effect of Otsu method.

The main process of our proposed algorithm is summarized in Algorithm 1.

Algorithm 1: Robust Saliency Detection Algorithm via Multi-Level Graph Structure and Accurate Background Queries Selection

---

**Input**: An image and related controlling parameters

1) Establish the three-level graph structure based on superpixels, calculate weight matrix $W$ and $D$ with Eq.1 and Eq.4 respectively.

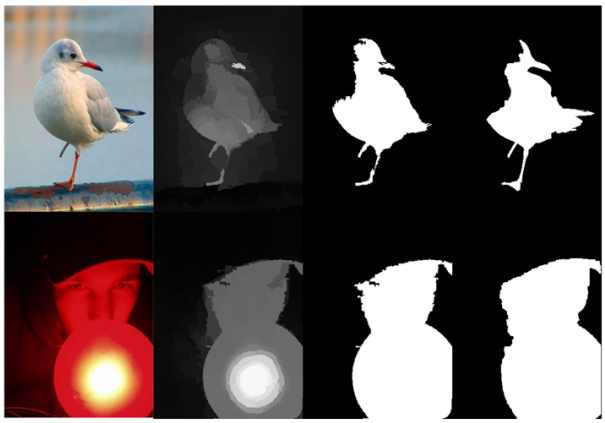

Fig. 9. The improvement of Otsu method. From left to right, original images, solutions without binarization, solutions with Otsu binarization, ground truth.

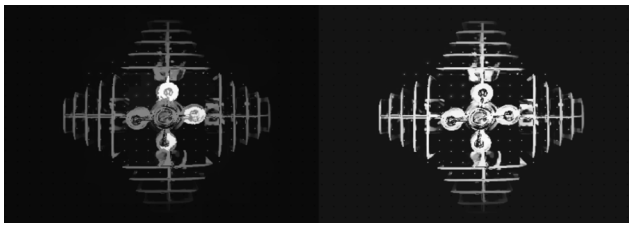

Fig. 10. The improvement of Otsu method. Left: Solution without Otsu binarization. Right: Solution with Otsu binarization.

2) Use all four boundaries as background queries to estimate the foreground object with Eq.2, Eq.3, Eq.9 and Eq.10.
3) Select accurate background superpixel queries with Eq.11.
4) Use accurate queries to acquire the background saliency map $S_b$ with Eq.2 and Eq.3
5) Acquire the foreground saliency map Sc with Eq.2, Eq.3, Eq.9 and Eq.10.
6) Establish the pixel-wise graph structure

and obtain $L$ with Eq.5.

7) Get the final saliency output $S_d$ by using Eq.15 and Eq.16.
8) Apply Otsu method and reshape the output to get $S_{final}$.

**Output**: a binary map with the same size as the input image, where 1 indicates salient objects and 0 indicates background.

## IV. Experiment Results

Datasets: We use four public datasets to test our method, the MSRA10K dataset [5] which contains 10000 randomly-chosen images from the MSRA dataset [28], the DUT-OMRON dataset [10] with 5168 manually selected high quality images, the BSDS500[29] dataset and the iCoseg[30] dataset.

Experimental Parameters Setup: For the experimental comparison, we use the same parameter settings in [10, 19], where the number of superpixel is set to $n = 200$, and the two controlling parameters are set to $\mu = 0.01$ and $\sigma^2 = 0.1$ respectively.

### A. Evaluation Metrics

We use the same evaluating method in [10, 18] to evaluate our method in terms of the precision, recall and F-measure and compare with 7 state-of-the-art approaches which are evaluated in [18]. The terms are defined in [31] as:

$$precision = \frac{\sum_{i=1}^{N} G(i)S_{final}(i)}{\sum_{i=1}^{N} S_{final}(i)}, \quad (17)$$

$$recall = \frac{\sum_{i=1}^{N} G(i)S_{final}(i)}{\sum_{i=1}^{N} G(i)}, \quad (18)$$

$$F_\beta = \frac{(1+\beta^2)precision \times recall}{\beta^2 precision + recall}, \quad (19)$$

where $G(i)$ is the ground truth. In other words, precision indicates the ratio of retrieved salient pixels to all retrieved pixels, recall means the ratio of retrieved salient pixels to all salient pixels in the image, and F-measure is adopted as a weighted average between precision and recall. We set $\beta^2 = 0.3$ to grant more weight on precision, as suggested in [2].

### B. Examination of design options

We first examine the four major innovations of our proposed method, as shown in Table1 and Figure11.

The first row and the second row show the effect of using accurate boundary superpixels selection. We observe that our boundary superpixels selection method is better than the previous boundary selection method.

The second row and the third row show the contribution of using the three-level graph structure. We observe that there is a tradeoff between precision and recall (See Fig.3 B1, B2,

B3, B4. In these solutions, the value of precision is 1, and the value of recall is close to 0). By using the three-level graph structure, the performance of precision decreases slightly while the performances of recall and F-measure score increase greatly. Also, we can compensate the loss of precision by using our regularized random walks ranking method.

The forth and the fifth rows show the effect of using our new graph connection structure and combine the solution of regularized random walks ranking with manifold ranking.

Based on the Table1, all these innovations of our proposed method have contributions to the overall performance.

### C. Comparison with State-of-the-art

We evaluate our proposed algorithm against 7 state-of-the-art saliency detection methods, they are:

FT[2], GS[32], LR[33], MR[10], PCA[16], SEG[34], and RRWR [18]).

We use MSRA10K, DUT-OMRON, BSDS500 and iCoseg dataset mentioned above to evaluate our algorithm. The results are shown in Figure14. These precision-recall curves in Figure11 demonstrate that the proposed method obviously outperforms all of the state-of-the-art algorithms in all of these datasets. The proposed method is especially better than MR and RRWR, which are two of the top-performance algorithms.

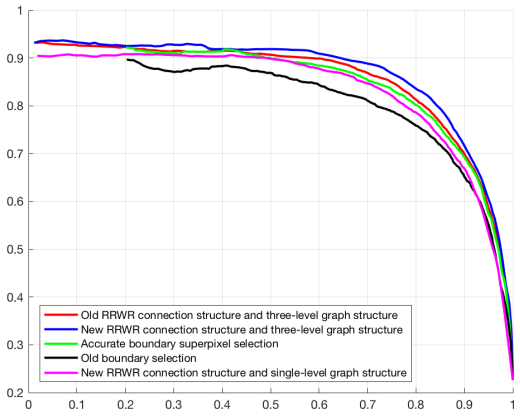


Fig. 11. Performance parameters of different innovations

After performing binarization algorithm, we also calculate F-measure score for each algorithm. The results are shown in Figure12. Our method achieves the highest F-measure score in all of the datasets. On average, our F-measure score (0.669) is $5.74\%$ over the second-best algorithm RRWR (0.632). We also calculate the variance of the F-measure score, our method's F-measure score variance (0.03363) is just $57\%$ of the second-best algorithm (RRWR, 0.058488). This means that our method is much more robust than RRWR. To provide a visual comparison of the different saliency outputs, we choose second best method and extract example saliency map outputs in Figure13. We find out that our method also generates saliency maps with clearer details and finer boundary adherences.

TABLE I

PERFORMANCE PARAMETERS OF DIFFERENT INNOVATIONS

| | Avg.(Precision) | Var.(Precision) | Avg.(Recall) | Var.(Recall) | F |
|---|---|---|---|---|---|
| Boundary Selection | 0.7008 | 0.0547 | 0.5187 | 0.0584 | 0.6483 |
| Accurate Boundary Selection | 0.7524 | 0.0504 | 0.5281 | 0.0561 | 0.6852 |
| Three-level graph sturcture | 0.7307 | 0.0533 | 0.7150 | 0.0207 | 0.7270 |
| Previous connection structure | 0.7477 | 0.0541 | 0.5213 | 0.0593 | 0.6796 |
| Our connection structure | 0.8041 | 0.0500 | 0.6303 | 0.0352 | 0.7429 |

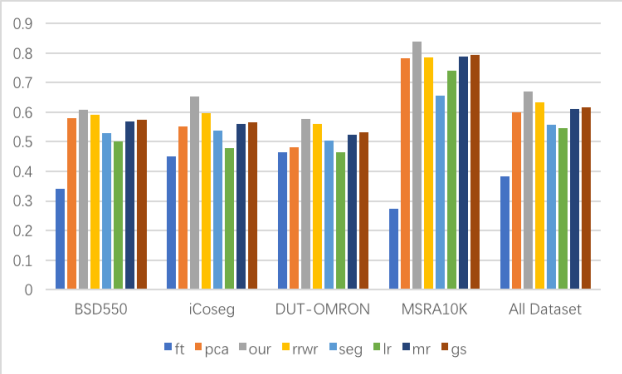


Fig. 12. The F-measure score in different datasets and in all of the four datasets.

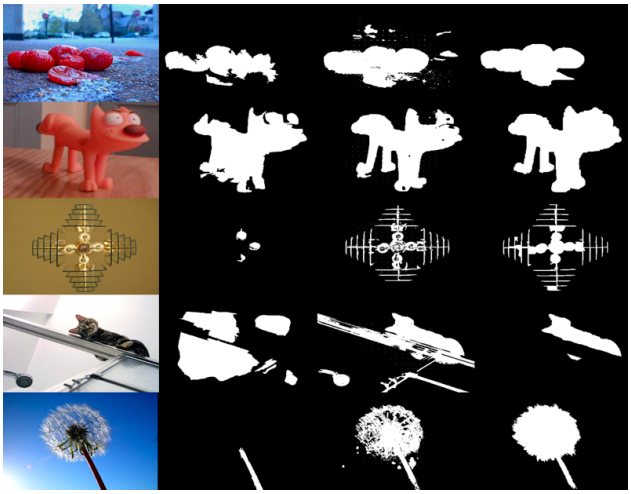

Fig. 13. Output comparison between our method and RRWR. From left to right: Original images, solutions of RRWR, solutions of our algorithm, ground truth.

## V. CONCLUSION

In this paper, we propose a novel bottom-up saliency detection method with multi-level graph structure, accurate boundary superpixel removal and regularized random walks ranking. There are four major innovations: firstly, the multi-level graph structure successfully captures both local and holistic visual information. This helps the algorithm get rid of the misleading of some high contrast local details; secondly, the accurate boundary superpixel removal effectively eliminates boundary-adjacent foreground superpixels, and thus works better than the previous boundary removal methods; thirdly, the combination of manifold ranking and regularized random walks not only helps the regularized random walks algorithm choose its seeds automatically but also improves the robustness of regularized random walks; fourthly, applying Otsu method greatly increases the performance of our method. Our approach is fully automatic and does not need any other supervision requirement. Results of experiments on two pub-

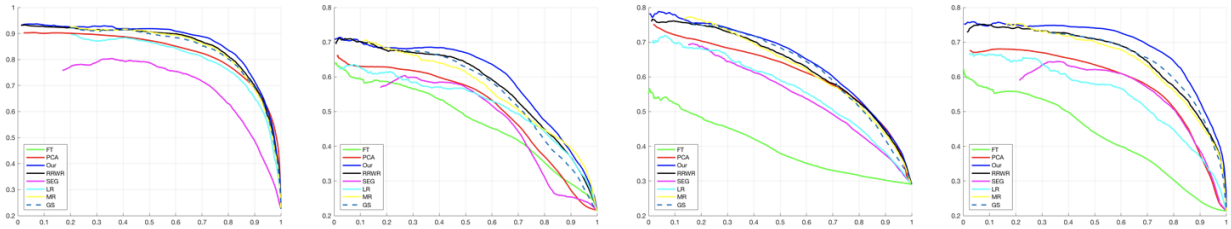


Fig. 14. The precision-recall plot in different dataset. From left to right, MSRA10K, DUT-OMRON, BSDS500 and iCoseg

lic datasets show that the proposed method significantly outperforms 7 state-of-the-art saliency detection algorithms in terms of both accuracy and robustness. In the future, we will further improve the performance of our method, and explore for more potential methods.

## REFERENCES

[1] S.Goferman, L.Zelnik-Manor, and A.Tal, "Context-aware saliency detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1915–1926, Oct 2012.

[2] U.Rutishauser, D.Walther, C.Koch, and P.Perona, "Is bottom-up attention useful for object recognition?" in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 2, June 2004, pp. II–37–II–44 Vol.2.

[3] L.Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *Trans. Img. Proc.*, vol. 13, no. 10, pp. 1304–1318, Oct. 2004. [Online]. Available: http://dx.doi.org/10.1109/TIP.2004.834657

[4] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu, "Sketch2photo: Internet image montage," *ACM Trans. Graph.*, vol. 28, no. 5, pp. 124:1–124:10, Dec. 2009. [Online]. Available: http://doi.acm.org/10.1145/1618452.1618470

[5] M.M.Cheng, N.J.Mitra, X.Huang, P.H.S.Torr, and S.M.Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, March 2015.

[6] K. Fu, C. Gong, J. Yang, Y. Zhou, and I. Yu-Hua Gu, "Superpixel based color contrast and color distribution driven salient object detection," *Image Commun.*, vol. 28, no. 10, pp. 1448–1463, Nov. 2013. [Online]. Available:

http://dx.doi.org/10.1016/j.image.2013.07.005

[7] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Random walks on graphs for salient object detection in images," *Trans. Img. Proc.*, vol. 19, no. 12, pp. 3232–3242, Dec. 2010. [Online]. Available: http://dx.doi.org/10.1109/TIP.2010.2053940

[8] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proceedings of the 19th International Conference on Neural Information Processing Systems*, ser. NIPS'06. Cambridge, MA, USA: MIT Press, 2006, pp. 545–552. [Online]. Available: http://dl.acm.org/citation.cfm?id=2976456.2976525

[9] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, Nov 1998.

[10] C.Yang, L.Zhang, H.Lu, X.Ruan, and M.H.Yang, "Saliency detection via graph-based manifold ranking," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, June 2013, pp. 3166–3173.

[11] S. Frintrop, G. Backer, and E. Rome, *Goal-Directed Search with a Top-Down Modulated Computational Attention System*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 117–124. [Online]. Available: https://doi.org/10.1007/11550518_15

[12] D. Gao and N. Vasconcelos, "Discriminant saliency for visual recognition from cluttered scenes," in *Advances in Neural Information Processing Systems 17*, L. K. Saul, Y. Weiss, and L. Bottou, Eds. Cambridge, MA: MIT Press, 2004, pp. 481–488. [Online]. Available: http://books.nips.cc/papers/files/nips17/NIPS2004_-0196.pdf

[13] J. Yang and M. H. Yang, "Top-down visual saliency via joint crf and dictionary learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 3, pp. 576–588, March 2017.

[14] T.Liu, Z.Yuan, J.Sun, J.Wang, N.Zheng, X.Tang, and H.Y.Shum, "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353–367, Feb 2011.

[15] A. Borji, "Exploiting local and global patch rarities for saliency detection," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, ser. CVPR '12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 478–485. [Online]. Available: http://dl.acm.org/citation.cfm?id=2354409.2354899

[16] R. Margolin, A. Tal, and L. Zelnik-Manor, "What makes a patch distinct?" in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, June 2013, pp. 1139–1146.

[17] P. Jiang, H. Ling, J. Yu, and J. Peng, "Salient region detection by ufo: Uniqueness, focusness and objectness," in *Proceedings of the 2013 IEEE International Conference on Computer Vision*, ser. ICCV '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 1976–1983. [Online]. Available: http://dx.doi.org/10.1109/ICCV.2013.248

[18] C. Li, Y. Yuan, W. Cai, Y. Xia, and D. D. Feng, "Robust saliency detection via regularized random walks ranking," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 2710–2717.

[19] L. Zhang, C. Yang, H. Lu, X. Ruan, and M. H. Yang, "Ranking saliency," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 9, pp. 1892–1904, Sept 2017.

[20] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Proceedings of the 12th European Conference on Computer Vision - Volume Part III*, ser. ECCV'12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 29–42. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-33712-3_3

[21] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR '14. Washington, DC, USA: IEEE Computer Society, 2014, pp. 2814–2821. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2014.360

[22] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang, "Saliency detection via absorbing markov chain," in *The IEEE International Conference on Computer Vision (ICCV)*, December 2013.

[23] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, Jan 1979.

[24] D. Zhou and B. Schölkopf, *Learning from Labeled and Unlabeled Data Using Random Walks*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 237–244. [Online]. Available: https://doi.org/10.1007/978-3-540-28649-3_29

[25] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf, "Ranking on data manifolds," in *Proceedings of the 16th International Conference on Neural Information*

*Processing Systems*, ser. NIPS'03. Cambridge, MA, USA: MIT Press, 2003, pp. 169–176. [Online]. Available: http://dl.acm.org/citation.cfm?id=2981345.2981367

[26] L. Grady, "Random walks for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1768–1783, Nov 2006.

[27] Q. Wang, W. Zheng, and R. Piramuthu, "Grab: Visual saliency via novel graph model and background priors," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[28] A. Borji, D. N. Sihite, and L. Itti, *Salient Object Detection: A Benchmark*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 414–429. [Online]. Available: https://doi.org/10.1007/978-3-642-33709-3_30

[29] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2010.161

[30] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen, "icoseg: Interactive co-segmentation with intelligent scribble guidance," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 3169–3176.

[31] G. Hripcsak and A. S. Rothschild, "Agreement, the f-measure, and reliability in information retrieval," *Journal of the American Medical Informatics Association*, vol. 12, no. 3, pp. 296–298, 2005. [Online]. Available: + http://dx.doi.org/10.1197/jamia.M1733

[32] Y. Wei, F. Wen, W. Zhu, and J. Sun, *Geodesic Saliency Using Background Priors*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 29–42. [Online]. Available: https://doi.org/10.1007/978-3-642-33712-3_3

[33] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 853–860.

[34] E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä, *Segmenting Salient Objects from Images and Videos*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 366–379. [Online]. Available: https://doi.org/10.1007/978-3-642-15555-0_27