Generic Promotion of Diffusion-Based Salient Object Detection

Peng Jiang [1] Nuno Vasconcelos [2] Jingliang Peng [1] *

[1]Shandong University, Jinan, Shandong, China

[2]University of California, San Diego, CA, USA

jump@mail.sdu.edu.cn, nuno@ucsd.edu, jpeng@sdu.edu.cn

Abstract

In this work, we propose a generic scheme to promote any diffusion-based salient object detection algorithm by original ways to re-synthesize the diffusion matrix and construct the seed vector. We first make a novel analysis of the working mechanism of the diffusion matrix, which reveals the close relationship between saliency diffusion and spectral clustering. Following this analysis, we propose to re-synthesize the diffusion matrix from the most discriminative eigenvectors after adaptive re-weighting. Further, we propose to generate the seed vector based on the readily available diffusion maps, avoiding extra computation for color-based seed search.

As a particular instance, we use inverse normalized Laplacian matrix as the original diffusion matrix and promote the corresponding salient object detection algorithm, which leads to superior performance as experimentally demonstrated.

1. Introduction

The aim of saliency detection is to identify the most salient pixels or regions in a digital image which attract humans' first visual attention. Results of saliency detection can be applied to other computer vision tasks such as image resizing, thumbnailing, image segmentation and object detection. Due to its importance, saliency detection has received intensive research attention resulting in many recently proposed algorithms.

In the field of saliency detection, two branches have developed, which are visual saliency detection [4,9,10,12–15, 19,29,34,39,41] and salient object detection [1,5–7,11,16–18,20–25,27,32,33,35,37,38,40,42]. While the former tries to predict where the human eye focuses on, the latter aims to detect the whole salient object in an image. Saliency in both branches can be computed in a bottom-up fashion using low level features [1,6,7,9,10,12–16,21,22,25,29,32–34,37, 38,40–42], in a top-down fashion by training with certain samples driven by specific tasks [4,17,19,20,23,24,27,39], or in a way of combining both low level and high level features [5,11,18,35]. In this paper, we focus on bottom-up salient object detection.

Salient object detection algorithms usually generate bounding boxes, binary foreground and background segmentation,or saliency maps which indicate the saliency likelihood of each pixel. Over the past several years, contrast based methods [1,6,11,32] significantly promote the benchmark of salient object detection. However, these methods usually miss small local salient regions or bring some outliers such that the resultant saliency maps tend to be nonuniform. To tackle these problems, diffusion-based methods [16,24,33,38] use diffusion matrices to propagate saliency information of seeds to the whole salient object. While most of them focus on how to generate good seed vectors, they have made little investigation on how to generate good diffusion matrices.

In this work, we aim at a generic scheme that promotes any diffusion-based salient object detection algorithm by constructing a good diffusion matrix and a good seed vector at the same time. First of all, we investigate the working mechanism of the diffusion matrix through eigen-analysis to find that the final saliency of a node (called focus node) is equal to a weighted sum of all the non-zero seed saliency values, with the weights determined by the similarity in diffusion map (see Sec. 2.2) between the corresponding seed node and the focus node. Further, since the diffusion map is formed by the eigenvectors and eigenvalues of the diffusion matrix, the process of saliency diffusion has a close relationship with spectral clustering. Inspired by the theories of spectral clustering [26,31], we propose to re-synthesize the diffusion matrix using only the most discriminative eigenvectors after adaptive re-weighting. Further, with the highly discriminative diffusion maps at hand, we propose to construct the seed vector based on the correlations between the non-border nodes, as measured by the similarities between their diffusion maps, which is time-efficient by avoiding an extra pass on color-based seed search.

*Corresponding author.

To the best of our knowledge, we for the first time explicitly reveal the close relationship between saliency diffusion and spectral clustering and, correspondingly, propose a generic and systematic scheme to promote any diffusion-based saliency detection algorithm. As a particular instance, we in this work use inverse normalized Laplacian matrix as the original diffusion matrix and promote the corresponding salient object detection method. As demonstrated by comprehensive experiments and analysis, the promotion leads to superior performance in salient object detection. Finally, the source code and experimental results of the proposed scheme are shared for research uses.

2. Diffusion-Based Methods

A diffusion-based salient object detection method usually segments an image into N superpixels first by an algorithm such as SLIC [2]. Then, it constructs a graph $G = (V, E)$ with superpixels as nodes v_i, $1 \leq i \leq N$, and undirected links between node pairs (v_i, v_j) as edges e_{ij}, $1 \leq i, j \leq N$. We make a close-loop graph by connecting the nodes at the four borders of the image to each other and then connecting each node to the nodes neighboring it and the nodes sharing common boundaries with its neighboring nodes. Thus, the distance between two nodes close to two different borders will be shortened by a path through borders.

The weight w_{ij} of the edge e_{ij} is defined as

$$w_{ij} = e^{-\frac{\|v_i - v_j\|_2}{\sigma^2}} \tag{1}$$

where v_i and v_j represent the mean colors of two nodes, respectively, in the CIE LAB color space, and σ is a constant that controls the strength of the weight. Given G and its affinity matrix $W = [w_{ij}]_{N\times N}$, the degree matrix D is defined as $D = diag\{d_{11}, ..., d_{NN}\}$, where $d_{ii} = \sum_j w_{ij}$.

2.1. Diffusion Matrix and Seed Vector

Diffusion-based methods such as [16,24,38] all share the similar main formula:

$$y = A^{-1}s, \tag{2}$$

where A^{-1} is the diffusion matrix (also called ranking matrix or propagation matrix), s is the seed vector (or query vector), and y is the final saliency vector to be computed. Here s usually contains preliminary saliency information of a portion of nodes, that is to say, usually s is not complete and we need to propagate the partial saliency information in s to the whole salient region along the graph to obtain the final saliency map [24]. The diffusion matrix A^{-1} is designed to fulfill this task.

Different algorithms derive diffusion matrices and seed vectors in different ways. Work [38] uses inverse Laplacian matrix L^{-1} as the diffusion matrix and uses binary background and foreground indication vectors as the seed vectors in two stages, respectively. Correspondingly, the formula of saliency diffusion is

$$y = L^{-1}s \tag{3}$$

where $L = D - W$. Work [24] computes s by combining hundreds of saliency features F with learned weight w ($s = Fw$), and uses inverse normalized Laplacian matrix L_{rw}^{-1} as the diffusion matrix. Correspondingly, the formula of saliency diffusion is

$$y = L_{rw}^{-1}s \tag{4}$$

where $L_{rw} = D^{-1}(D - W)$. Work [16] duplicates the superpixels around the image borders as the virtual background absorbing nodes, and sets the inner nodes as transient nodes. Then, the entry of seed vector $s_i = 1$ if node v_i is transient node and $s_i = 0$ otherwise. Correspondingly, the formula of saliency diffusion is

$$y = (I - P)^{-1}s = L_{rw}^{-1}s \tag{5}$$

where $P = D^{-1}W$ and P is called transition matrix. Note that Eq. 5 is derived from but not identical to the original formula in reference [16] and the derivation process is described in the supplementary.

According to [26], L_{rw} is preferable to L for the spectral clustering since the former often leads to better intra-cluster coherency and clustering consistency. Therefore, we use L_{rw} in this work to explain and demonstrate our proposed scheme.

2.2. Diffusion Map

Diffusion-based salient objection detection algorithms (*e.g.*, [16, 24, 38]) usually use a positive semi-definite matrix, A, to define the diffusion matrix. Thus, A can be decomposed as $A = U_A \Lambda_A {U_A}^T$ where Λ_A is a diagonal matrix formed from the eigenvalues λ_{A_l}, $l = 1, 2, \ldots, N$, and the columns of U_A are the corresponding eigenvectors u_{A_l}, $l = 1, 2, \ldots, N$, of A. According to spectral decomposition theories, each element, $\tilde{a}(i, j)$, of A^{-1} can then be expressed as

$$\tilde{a}(i,j) = \sum_{l=1}^{N} \lambda_{A_l}^{-1} u_{A_l}(i) u_{A_l}(j). \tag{6}$$

and each entry, y_i, of y as

$$\begin{aligned} y_i &= \sum_{j=1}^{N} s_j \sum_{l=1}^{N} \lambda_{A_l}^{-1} u_{A_l}(i) u_{A_l}(j) \\ &= \sum_{j=1}^{N} s_j \left\langle \Psi_{A_i}, \Psi_{A_j} \right\rangle, \end{aligned} \tag{7}$$

$$\Psi_{A_i} = [\lambda_{A_1}^{-\frac{1}{2}} u_{A_1}(i), ..., \lambda_{A_N}^{-\frac{1}{2}} u_{A_N}(i)] \quad (8)$$

where $\langle \cdot, \cdot \rangle$ is the inner product operation. According to [8], Ψ_{A_i} is called diffusion map (diffusion map at time $t = -\frac{1}{2}$ to be more exactly) at the i-th data point (node).

Based on Eq.s 7 and 8, we make a novel interpretation of the working mechanism of diffusion-based salient object detection: the saliency of a node (called focus node) is determined by all the seed values in the form of weighted sum, with each weight determined by diffusion map similarity (measured by inner product) between the corresponding seed node and the focus node. In other words, seed nodes having more diffusion map similarity to the focus node will influence more on the focus node's saliency. This matches our intuition that similar (distinct) nodes should in general have similar (distinct) saliency values. This interpretation directly leads to our novel ways to re-synthesize the diffusion matrix and construct the seed vector, as detailed in the following two sections, respectively.

3. Re-Synthesis of Diffusion Matrix

As analyzed in Sec. 2.2, nodes with similar (distinct) diffusion maps tend to obtain similar (distinct) saliency values according to Eq.s 7 and 8. Therefore, the process of saliency diffusion is closely related to the clustering of the nodes based on their diffusion maps. Further, diffusion maps are derived from the eigenvalues and eigenvectors of the diffusion matrix, *i.e.*, we from a matrix by putting the weighted eigenvectors in columns and each row of the matrix gives one node's diffusion map (see Eq. 8). As such, the diffusion-map-based clustering is almost identical in form to the standard spectral clustering of the nodes [26, 31].

According to spectral clustering theories [26, 31], only a subset of the eigenvectors are the most discriminative, while the rest are not as discriminative or even cause confusions to the clustering. Therefore, in order to increase the discriminative power of the diffusion maps, we are motivated to keep only the most discriminative while discarding the rest of the eigenvectors. This can be fundamentally achieved by re-synthesizing the diffusion matrix from the most discriminative eigenvectors, as detailed below.

3.1. Constant Eigenvector

The eigenvalues, λ_l, and eigenvectors, u_l, $1 \leq l \leq N$, of L_{rw} are ordered such that $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_N$ with $u_1 = \mathbf{1}$ [31]. Some works (*e.g.*, [38]) avoid zero eigenvalues by approximately setting $\widetilde{L} = D - 0.99W$ such that $\widetilde{L}$ is always invertible. Assuming $\widetilde{\lambda}_l$ and $\widetilde{u}_l$, $1 \leq l \leq N$, are the corresponding eigenvalues and eigenvectors of $\widetilde{L}_{rw} = D^{-1}(D - 0.99W)$, it can be proven that $\widetilde{u}_l = u_l$ and $\widetilde{\lambda}_l = 0.99\lambda_l + 0.01$. Thus, $0.01 = \widetilde{\lambda}_1 \leq \widetilde{\lambda}_2 \leq \ldots \leq \widetilde{\lambda}_N$ with $\widetilde{u}_1 = \mathbf{1}$.

Though not zero, $\widetilde{\lambda}_1 = 0.01$ is still a small value. As a result, the constant eigenvector $\widetilde{u}_1$ with no discriminative information has a significant influence on the nodes' diffusion maps and, correspondingly, suppresses other eigenvectors and weakens the discriminative power of diffusion maps. Therefore, our solution is to discard the constant eigenvector and re-synthesize the diffusion matrix.

3.2. Eigengap

Except u_1 that is a constant vector, the more u_l ($l \in [2, N]$) is to the front of the ordered array, the more indicative it usually is for the clustering. For instance, we visualize in Fig. 1 a leading portion (excluding u_1) of the ordered array of eigenvectors for each of four sample images. From Fig. 1, we see that, for each sample image, the first few eigenvectors well indicate node clusters while the later ones often convey less information about or even confuse the clustering. The key is how to determine the exact cutting point before which the eigenvectors should be kept and after which discarded.

In practice, L_{rw} often exhibits an eigengap, *i.e.*, a few of its eigenvalues before the eigengap are much smaller than the rest. Specifically, we denote the eigengap of L_{rw} as r and define it as

$$\begin{aligned} r &= \underset{l}{\operatorname{argmax}} |\Delta\Upsilon_l|, \\ \Delta\Upsilon_l &= \lambda_l - \lambda_{l-1},\ l = 2, \ldots, N. \end{aligned} \quad (9)$$

Usually, Eq. 9 is called eigengap heuristic. According to [26], some leading eigenvectors (except u_1) before the eigengap are usually good cluster indicators which can capture the data cluster information with good accuracy (as observed in Fig. 1), meanwhile the location of the eigengap often indicates the right number of data clusters. Further, the larger the difference between the two successive eigenvalues at the eigengap is, the more important the leading eigenvectors are, since u_l is weighted by $\lambda_l^{-\frac{1}{2}}$ in diffusion map Ψ (see Eq. 8). Ideally, the eigenvalues before the eigengap are close to zero while the rest are much larger, which means that the leading eigenvectors (except u_1) will dominate the behavior of the diffusion map.

With the eigengap identified, we then keep only the eigenvectors prior to the eigengap excluding the constant u_1, which are usually the most discriminative ones for the task of node clustering. It may sometimes happen that $r = 2$ according to Eq. 9, meaning that all the eigenvectors will be filtered out. In this case, we assume the position of the second largest $|\Delta\Upsilon_l|$ as the eigengap.

3.3. Discriminability

In some cases, an eigenvector may only distinguish a tiny region from the background, *e.g.*, u_5 , u_6 in the second row and u_6 in the last row of Fig. 1. Usually, these tiny regions

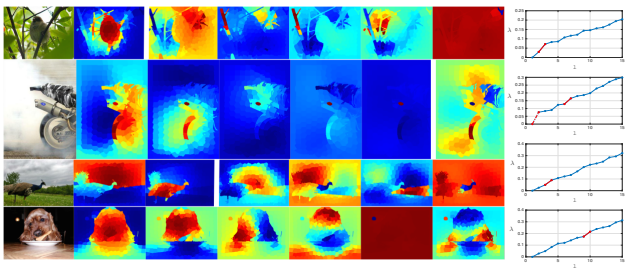

Figure 1. Visualization of normalized eigenvectors by color coding. Pixels in each node are assigned a single color and nodes with similar values in an eigenvector are colored similarly. The eight columns show the source images (SRC), the corresponding eigenvectors (u_2-u_7) and eigenvalue curves (λ), respectively. we use a white margin between successive eigenvectors to indicate an eigengap (All the eigenvectors, u2 to u7, are before the eigengap, if there is no white margin in that row.). Besides, on the eigenvalue curves, we use red solid segments to indicate the final eigengaps and a red dash segment to indicate an initial eigengap of $r = 2$ to be reset. Ground truth saliency of the source images are shown in Fig. 4.

are less likely to be the salient regions we search for. Besides, these tiny regions often have been captured by other leading eigenvectors as well. Therefore, such eigenvectors have low discriminability and may even worsen the final results by overemphasizing tiny regions.Therefore, we evaluate the discriminability of eigenvector u_l by its variance $var(u_l)$, and filter out eigenvectors with variance values below a threshold, v. Specifically, we formulate the discriminability indicator of u_l as

$$dc(u_l) = \begin{cases} 0, & var(u_l) < v \\ 1, & else \end{cases}, \qquad \widetilde{DC} = diag\{dc(u_2), \ldots, dc(u_r)\} \tag{10}$$

where $\widetilde{DC}$ is the matrix with dc values of all the eigenvectors as its diagonal elements. When re-synthesizing the diffusion matrix (see Sec. 3.4), we use $\widetilde{DC}$ as a weighting matrix, which equals to selecting each u_l by a binary factor of $dc(u_l)$.

3.4. Integration

Finally, discarding u_1 and λ_1 and using the eigenvectors and eigenvalues prior to the eigengap and the weighting matrix, $\widetilde{DC}$, we re-synthesize the original diffusion matrix, L_{rw}^{-1}, to $\widetilde{A}^{-1}$ by

$$\begin{aligned} &\widetilde{U} = [u_2, \ldots, u_r]; \\ &\widetilde{\Lambda}^{-1} = diag\{\lambda_2^{-1}, \ldots, \lambda_r^{-1}\}; \\ &\widetilde{A}^{-1} = \widetilde{U}\widetilde{\Lambda}^{-1}\widetilde{DC}\widetilde{U}^T. \end{aligned} \tag{11}$$

4. Seed Vector Construction

Other diffusion-based saliency object detection methods usually generate the seed vector based on low-level features. Since we already have the highly discriminative diffusion maps of the re-synthesized diffusion matrix, we propose to construct our seed vector directly based on them. Besides yielding good accuracy in seed value estimation, this approach is time-efficient since we avoid an extra pass of color-based preliminary saliency search.

Specifically, we compute each entry, $\tilde{s}_i$, $1 \leq i \leq N$, of the seed vector, $\tilde{s}$, by summing the inner products of its diffusion map with those of all the m non-border nodes. Assuming that the non-border nodes are assigned the smallest indices, we compute $\tilde{s}$ by

$$\tilde{s}_i = \sum_{j=1}^{m} \left\langle \Psi_{\tilde{A}_i}, \Psi_{\tilde{A}_j} \right\rangle \tag{12}$$

or, equivalently,

$$\tilde{s} = \widetilde{A}^{-1} x \tag{13}$$

where $\tilde{s} = (\tilde{s}_1, \ldots, \tilde{s}_N)$ and $x = (x_1, \ldots, x_N)$ with $x_i = 1$ if v_i is a non-border node and $x_i = 0$ otherwise. A more in-depth analysis of the working mechanism of the proposed seed vector construction method is made in the supplementary.

5. Combination

We use $\tilde{s}$ as the seed vector and $\widetilde{A}^{-1}$ as the diffusion matrix, and compute the saliency vector y as

$$\begin{aligned} y &= \widetilde{A}^{-1}\tilde{s}, \\ &= (\widetilde{A}^{-1})^2 x, \end{aligned} \tag{14}$$

Thereafter, we obtain the saliency map S by assigning the value of y_i to the corresponding node v_i, $1 \leq i \leq N$. The main steps of the proposed salient object detection algorithm are summarized in Algorithm 1.

6. Experiments and Analysis

6.1. Datasets and Evaluation Methods

Our experiments are conducted on two datasets: the MSRA10K dataset [6, 7] with $10K$ images and the ECSSD dataset [37] with $1K$ images. Each image in these datasets is associated with a human-labeled ground truth. In order to study the performance of saliency detection algorithms, we adopt prevalently used evaluation protocols including precision-recall (PR) curves [1], F-measure score which is a weighted harmonic mean of precision and recall [1], mean overlap rate (MOR) score [18] and area under ROC curve (AUC) score [24]. Further, we propose to measure the quality of a diffusion matrix by constrained optimal seed efficiency (COSE), as described in Sec. 6.4.

Algorithm 1 Promoted Diffusion-Based Salient Object Detection

Input: An image on which to detect the salient object.

1: Segment the input image into superpixels, use the superpixels as nodes, connect border nodes to each other and connect close nodes to construct a graph G, and compute its degree matrix D and weight matrix W.
2: Compute $L_{rw} = D^{-1}(D-W)$ and its eigenvalues and eigenvectors.
3: Estimate the eigengap of L_{rw} by Eq. 9, discard the first constant eigenvector and the eigenvectors after the eigengap.
4: Re-weight the remaining eigenvectors by discriminability as computed by Eq. 10.
5: Form the re-synthesized diffusion matrix $\widetilde{A}^{-1}$ by Eq. 11 and compute the seed vector $\widetilde{s}$ by Eq. 13.
6: Compute the final saliency vector y by Eq. 14.

Output: The saliency vector y representing the saliency value of each superpixel.

In the experiments, we evaluate different diffusion matrices by visual saliency promotion and constrained optimal seed efficiency, as detailed in Sec. 6.3 and Sec. 6.4, respectively, and compare different salient object detection algorithms, as detailed in Sec. 6.5. At last, we show in Sec. 6.6 the effects of different steps in the proposed diffusion matrix re-synthesis algorithm.

6.2. Experimental Settings

We empirically choose $\sigma = 10$ in Eq. 1 and set $v = 300$ in Eq. 10. In order to avoid zero eigenvalues, we approximately set $\widetilde{L}_{rw} = D^{-1}(D - 0.99W)$ and $\widetilde{L} = D - 0.99W$ when comparing diffusion matrices, as done in reference [38]. However, our diffusion matrix is directly re-synthesized from $L_{rw} = D^{-1}(D - W)$. When comparing with other salient object detection methods in Sec. 6.5, we further use standard image processing techniques to increase the contrast of the final saliency maps.

6.3. Promotion of Visual Saliency

Visual saliency detection predicts human fixation locations in an image, which are often indicative of salient objects around. Therefore, we use the detected visual saliency as the seed information, and conduct diffusion on it to detect the salient object region in an image. In other words, we promote a visual saliency detection algorithm by diffusion for the task of salient object detection.

In this experiment, we use the results of nine visual saliency detection methods (*i.e.*, IT [15], AIM [4], GB [12], SR [14], SUN [41], SeR [34], SIM [29], SS [13] and COV [9]) on the MSRA10K dataset as the seed vectors, respectively, and compare the saliency detection results before and after diffusion. For the diffusion, we test three matrices including $\widetilde{A}^{-1}$, $\widetilde{L}^{-1}$ and $\widetilde{L}_{rw}^{-1}$. The PR curves of the nine visual saliency detection methods before and after diffusion by $\widetilde{A}^{-1}$, $\widetilde{L}^{-1}$ and $\widetilde{L}_{rw}^{-1}$ are plotted in Fig. 2(a), (b) and (c), respectively.

Remarkably, as shown in Fig. 2, previous visual saliency detection methods which usually can not highlight the whole salient object all get significantly boosted after diffusion with any of $\widetilde{A}^{-1}$, $\widetilde{L}^{-1}$ and $\widetilde{L}_{rw}^{-1}$. The promotion is so significant that some promoted methods even outperform some state-of-the-art salient objection detection methods, as observed by comparing Fig. 2 and Fig. 3, meaning that, with a good diffusion matrix, we can fill the performance gap between two branches of saliency detection methods.

Comparing Fig.s 2(a), 2(b) and 2(c), we observe that $\widetilde{A}^{-1}$ leads to more significant performance promotion and more consistent promoted performance than $\widetilde{L}^{-1}$ and $\widetilde{L}_{rw}^{-1}$, demonstrating higher effectiveness and robustness of the re-synthesized diffusion matrix, $\widetilde{A}^{-1}$, in visual saliency promotion.

6.4. Constrained Optimal Seed Efficiency

We prefer a diffusion matrix to use as little query information or, equally, as few non-zero seed values to derive as close saliency to the ground truth as possible. Correspondingly, for a diffusion matrix, we measure the constrained optimal saliency detection accuracy it may achieve at each non-zero seed value budget, leading to an constrained optimal seed efficiency curve, as detailed below.

Given the ground truth GT, the diffusion matrix A^{-1}, we hope to find the optimal seed vector, s, that minimizes the residual, res, computed by

$$res = GT - A^{-1}s. \tag{15}$$

Aiming to reduce the number of non-zero values in s, we turn the residual minimization to a sparse recovery problem, to solve which we adapt the algorithm of orthogonal matching pursuit (OMP) [36], as described in Alg. 2.

As shown in Alg. 2, we adapt the residual computation to $\widetilde{res} = GT - bin(A^{-1}s)$ in Step 4, where bin is the binarization operation since GT is binary; we multiply a factor $GT(j)$ in Step 1 to ensure that the non-zero seed values are selected from only the salient region; we solve the nonnegative least-squares problem in step 3 of Alg. 2 to ensure nonnegative elements of s. The adapted OMP will stop when $\|\widetilde{res}\|_2$ is below a threshold, c, or the nonnegative seed values at the salient region are all selected, as shown in Step 5 of Alg. 2. We see that the optimization process in Alg. 2 is constrained, *e.g.*, the seeds are selected from only the salient region, the optimization is conducted in a greedy fashion and so forth. Although the saliency detection performance of these resultant seed vectors provides a good reference for

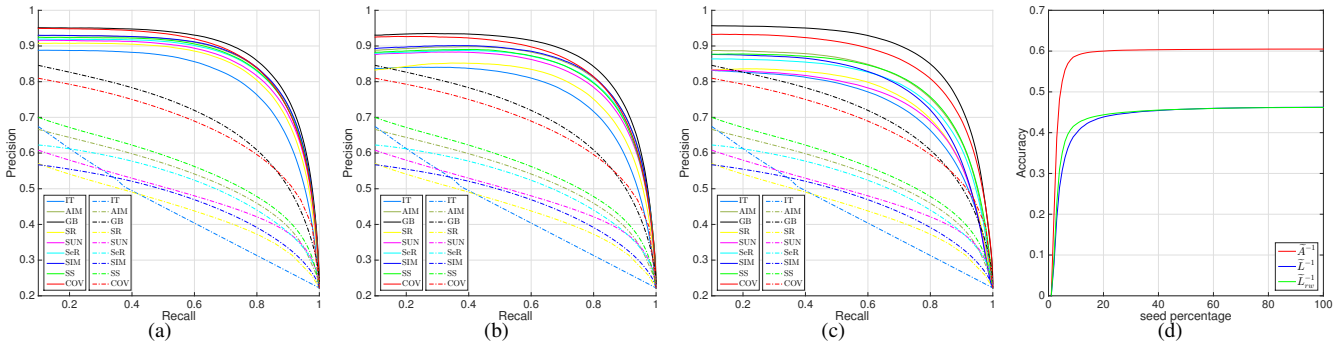

Figure 2. PR curves of nine visual saliency detection methods before (dash line) and after (solid line) diffusion by (a) $\widetilde{A}^{-1}$, (b) $\widetilde{L}^{-1}$, and (c) $\widetilde{L}_{rw}^{-1}$. The constrained optimal seed efficiency curves for $\widetilde{A}^{-1}$, $\widetilde{L}^{-1}$ and $\widetilde{L}_{rw}^{-1}$ on the MSRA10K dataset are shown in (d).

Dataset	Protocol	PCA	GMR	MC	DSR	BMS	HS	GC	RBD	Ours
MSRA10K	Precision	0.80289	0.89021	0.89063	0.8532	0.83237	0.88492	0.82117	0.87157	0.87807
	Recall	0.67817	0.752	0.75455	0.73813	0.72263	0.71551	0.67469	0.79522	0.78882
	F-measure	0.7702	0.85399	0.85504	0.82357	0.80419	0.83908	0.78199	0.85267	0.85573
	AUC	0.94111	0.94379	0.95074	0.95888	0.92901	0.93264	0.91169	0.95474	0.96358
	Overlap	0.57652	0.69254	0.69386	0.65398	0.63533	0.65576	0.59866	0.71582	0.71011
ECSSD	Precision	0.66047	0.76865	0.77004	0.74891	0.70377	0.76924	0.65498	0.72626	0.7376
	Recall	0.52427	0.64498	0.65227	0.64544	0.59603	0.53912	0.48608	0.66356	0.68775
	F-measure	0.62311	0.73608	0.73924	0.72219	0.67559	0.70027	0.60636	0.71076	0.72547
	AUC	0.87643	0.89127	0.91113	0.9154	0.86814	0.88534	0.80438	0.8959	0.91663
	Overlap	0.39517	0.52335	0.53065	0.51352	0.46533	0.45799	0.39145	0.52522	0.53146

Table 1. Performance statistics of different algorithms on the five protocols and the two datasets. For each dataset and protocol, the top three results are highlighted in red, blue and green, respectively.

our diffusion matrix evaluation, it should be noted that their optimal performance is constrained but not absolute.

In order to obtain the constrained optimal seed efficiency curve over the full range of nonnegative seed value budget, we set $c = 0$ in Alg. 2 and, at the i-th ($0 \leq i \leq 100$) iteration, we compute and record the pair of nonnegative seed percentage, r_i, and saliency detection accuracy, a_i, according to the following formulae:

$$r_i = \frac{100 \times \|s\|_0}{\|GT\|_0}\%, \qquad a_i = \frac{\|GT\|_2 - \|\widetilde{res}\|_2}{\|GT\|_2}. \tag{16}$$

Based on these (r_i, a_i) pairs, we can plot the OSE curve of A^{-1} on an image.

We substitute $\widetilde{A}^{-1}$, $\widetilde{L}^{-1}$ and $\widetilde{L}_{rw}^{-1}$ into Eq. 15 for A^{-1}, respectively. For each diffusion matrix, we plot the average OSE curve over all the images in the MSRA10K dataset, as shown in Fig. 2(d). From Fig. 2(d), we observe that the constrained optimal seed efficiency rises sharply at the beginning and levels off at around the nonnegative seed percentage of 30%, that $\widetilde{A}^{-1}$ exhibits significantly higher average constrained optimal seed efficiency than $\widetilde{L}^{-1}$ and $\widetilde{L}_{rw}^{-1}$, and that there is an inherent performance ceiling for each diffusion matrix while $\widetilde{A}^{-1}$ has the highest one. According to the last observation, it appears that the performance of diffusion-based saliency detection is fundamentally determined by the diffusion matrix, again emphasizing the importance in constructing a good diffusion matrix.

Algorithm 2 Adapted Orthogonal Matching Pursuit

Input: Dictionary($A_{N\times N}^{-1}$), Signal($GT_{N\times 1}$) and Stop criterion(c)

Output: Coefficient vector($s_{N\times 1}$) and Residual(res)

Initialize: $res = GT$, $Inds = \emptyset$, $FgInds = \arg_i\{GT(i) = 1\}$

Iteration:

1: $ind = \arg\max_j\{|\langle res, A^{-1}(:,j)\rangle| \cdot GT(j)\}$, $j \in FgInds$;
2: $Inds = Inds \cup ind$, $FgInds = FgInds \setminus ind$;
3: $s(Inds) = \arg\min_{\tilde{s}\geq 0}\|GT - A^{-1}(:, Inds)\tilde{s}\|_2$;
4: $\widetilde{res} = GT - bin(A^{-1}s)$,
5: **if** $\|\widetilde{res}\|_2 \geq c \wedge FgInds \neq \emptyset$ **then**
6: **Go to** 1;
7: **end if**

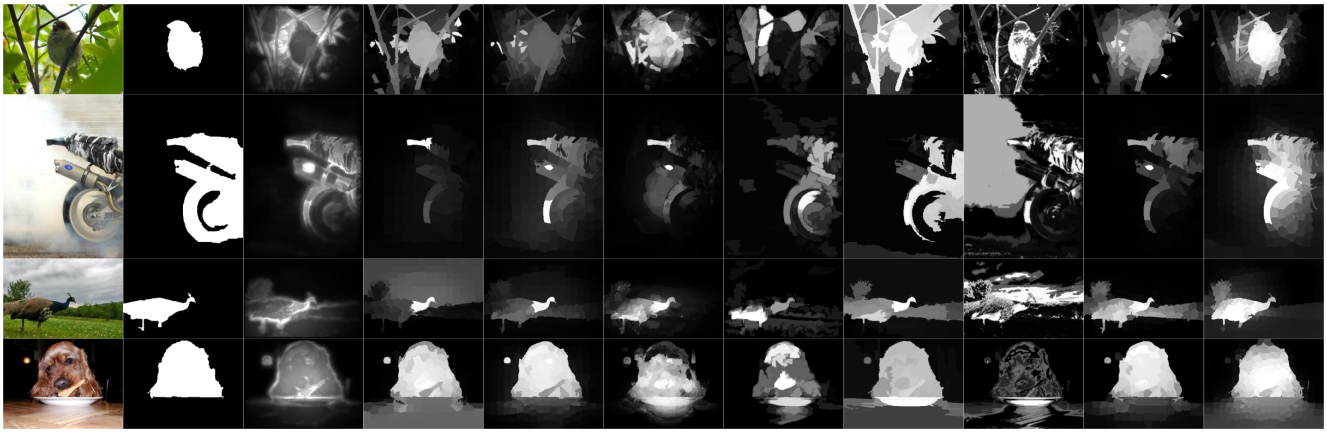

Figure 4. Visual comparison of previous approaches to our method and ground truth (GT).

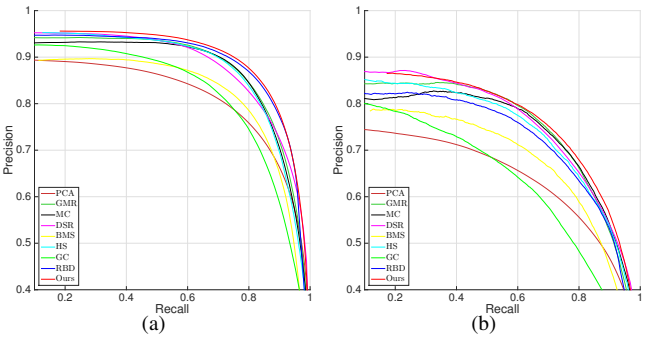

Figure 3. PR curves for all the algorithms on (a) the MSRA10K dataset [6, 7] and (b) the ECSSD dataset [37].

6.5. Salient Object Detection

We experimentally compare our method(Ours) with eight other recently proposed ones including PCA [28], GMR [38], MC [16], DSR [21], BMS [40], HS [37], GC [7] and RBD [42] on salient object detection. When evaluating these methods, we either use the results from the original authors (when available) or run our own implementations. Note that GMR, MC, DSR, and RBD have been identified as the top performers on the saliency benchmark study of work [3][1].

We plot the PR curves of all the nine methods on the MSRA10K dataset and the ECSSD dataset in Fig.s 3(a) and 3(b), respectively. Further, we provide the performance statistics on the five prevalent protocols for all the methods on the two datasets in Tab. 1. From both Fig. 3 and Tab. 1, we clearly observe that our proposed method yields top performance.

For visual comparison, we show in Fig. 4 the saliency object detection results by the benchmark methods and our method on several images in MSRA10K. From Fig. 4, we observe clearly that our method produces much closer results to the ground truth than the others. It is worth noting that, of the benchmark methods, GMR [38] and MC [16] are diffusion-based ones while our method produces much better results than them.

The average running time (without parallel programming) is 0.75s per image on the MSRA10K dataset on a machine with Intel Core i7 2.2 GHz CPU.

[1]We note that the DFRI method of work [17] achieved the best performance on this study. This, however, is a supervised learning method, which uses 3,000 out of the 10,000 MSRA10K images for training. In this work, we only consider unsupervised methods.

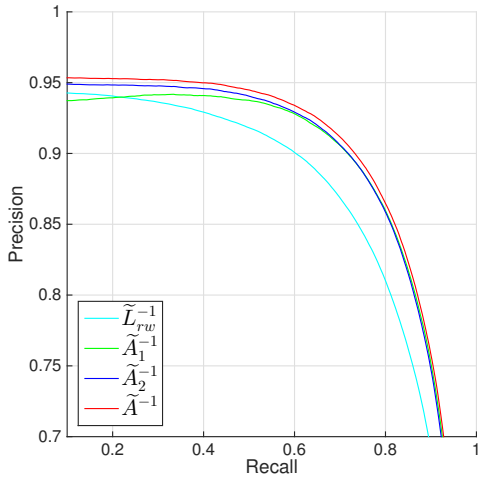

Figure 5. PR curves for diffusion matrices re-synthesized to different steps on the MSRA10K dataset [6, 7].

6.6. Effects of Steps in Diffusion Matrix Re-Synthesis

In this section, we demonstrate the effects of separate steps in the proposed diffusion matrix re-synthesis method (see Sec. 3), as detailed below.

For each test image, we start from its $\widetilde{L}_{rw}^{-1}$ and sequentially obtain $\widetilde{A}_1^{-1}$, $\widetilde{A}_2^{-1}$ and $\widetilde{A}^{-1}$ when the constant eigen-

vector is discarded, the eigenvectors after the eigengap are further filtered out and the discriminability weighting is finally conducted, respectively. Thereafter, we use $\widetilde{L}_{rw}^{-1}$, $\widetilde{A}_1^{-1}$, $\widetilde{A}_2^{-1}$ and $\widetilde{A}^{-1}$, respectively, as the re-synthesized diffusion matrix and run our saliency object detection algorithm on the test image. Experimenting on the whole MSRA10K dataset [6, 7], we obtain the PR curves for $\widetilde{L}_{rw}^{-1}$, $\widetilde{A}_1^{-1}$, $\widetilde{A}_2^{-1}$ and $\widetilde{A}^{-1}$, as plotted in Fig. 5.

From Fig. 5, we observe that discarding the constant eigenvector ($\widetilde{A}_1^{-1}$) clearly boosts the performance of $\widetilde{L}_{rw}^{-1}$, the eigengap-based eigenvector filtering ($\widetilde{A}_2^{-1}$) further improves the performance, especially in precision, and the final incorporation of discriminability ($\widetilde{A}^{-1}$) leads to the top performance.

7. Conclusions

In this work, we make a novel analysis of the working mechanism of the diffusion-based salient object detection. Through analysis, we find that the saliency of each node is formed by a weighted sum of all the seeds' saliency values, with the weights determined by the diffusion map similarities between the nodes. In order to increase the discriminative power of the diffusion maps, we keep only the most dominant eigenvectors and use them (after adaptive reweighting) to re-synthesize the diffusion matrix. Further, we construct the seed vector based on the correlations of diffusion maps between the non-border nodes, taking advantage of the diffusion maps' discriminative power while saving extra computation for color-based seed search.

The proposed scheme is a generic one which can be used to promote any diffusion-based saliency object detection algorithm. As a particular instance, we use inverse normalized Laplacian matrix, L_{rw}^{-1}, as the original diffusion matrix and promote the corresponding saliency detection algorithm. Experiments show that the promoted diffusion matrix is superior in both visual saliency promotion and constrained optimal seed efficiency, and the promoted salient object detection method advances the state of the art.

There are known limitations of spectral clustering [30], such as sensitivity of the eigenvectors to the scale parameter value σ^2 in Eq. 1. Various approaches have been proposed in the literature to overcome these problems. These should, in principle, be applicable to the saliency problem. We intend to investigate this in the future.

Acknowledgment

This work was partially funded by NSFC (National Natural Science Foundation of China) grant 61472223 and NSF (National Science Foundation) grant IIS-1208522. The first and the last co-authors are also from Engineering Research Center of Digital Media Technology, Ministry of Education of PRC.

References

[1] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. *CVPR*, 2009. 1, 4

[2] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE PAMI*, 2012. 2

[3] A. Borji, M.-M. Cheng, H. Jiang, and J. Li. Salient object detection: A benchmark. *arXiv eprint*, 2015. 7

[4] N. D. B. Bruce and J. K. Tsotsos. Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision*, 2009. 1, 5

[5] K. Y. Chang, T. L. Liu, H. T. Chen, and S. H. Lai. Fusing generic objectness and visual saliency for salient object detection. *ICCV*, 2011. 1

[6] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S. M. Hu. Global contrast based salient region detection. *IEEE PAMI*, 2015. 1, 4, 7, 8

[7] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and N. Crook. Efficient salient region detection with soft image abstraction. *ICCV*, 2013. 1, 4, 7, 8

[8] R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 2006. 3

[9] E. Erdem and A. Erdem. Visual saliency estimation by nonlinearly integrating features using region covariances. *Journal of Vision*, 2013. 1, 5

[10] D. Gao and N. Vasconcelos. Decision-theoretic saliency: Computational principles, biological plausibility, and implications for neurophysiology and psychophysics. *Neural Computation*, 2009. 1

[11] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. *IEEE PAMI*, 2012. 1

[12] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. *NIPS*, 2006. 1, 5

[13] X. Hou, J. Harel, and C. Koch. Image signature: Highlighting sparse salient regions. *IEEE PAMI*, 2012. 1, 5

[14] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. *CVPR*, 2007. 1, 5

[15] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE PAMI*, 1998. 1, 5

[16] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang. Saliency detection via absorbing markov chain. *ICCV*, 2013. 1, 2, 7

[17] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li. Salient object detection: A discriminative regional feature integration approach. *CVPR*, 2013. 1, 7

[18] P. Jiang, H. Ling, J. Yu, and J. Peng. Salient region detection by ufo: Uniqueness, focusness and objectness. *ICCV*, 2013. 1, 4

[19] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. *ICCV*, 2009. 1

[20] J. Kim, D. Han, Y.-W. Tai, and J. Kim. Salient region detection via high-dimensional color transform. *CVPR*, 2014. 1

[21] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang. Saliency detection via dense and sparse reconstruction. *ICCV*, 2013. 1, 7

[22] R. Liu, J. Cao, Z. Lin, and S. Shan. Adaptive partial differential equation learning for visual saliency detection. *CVPR*, 2014. 1

[23] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Shum. Learning to detect a salient object. *IEEE PAMI*, 2011. 1

[24] S. Lu, V. Mahadevan, and N. Vasconcelos. Learning optimal seeds for diffusion-based salient object detection. *CVPR*, 2014. 1, 2, 4

[25] Y. Lu, W. Zhang, H. Lu, and X. Y. Xue. Salient object detection using concavity context. *ICCV*, 2011. 1

[26] U. Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 2007. 1, 2, 3

[27] L. Mai, Y. Niu, and F. Liu. Saliency aggregation: A data-driven approach. *CVPR*, 2013. 1

[28] R. Margolin, A. Tal, and L. Zelnik-Manor. What makes a patch distinct? *CVPR*, 2013. 7

[29] N. Murray, M. Vanrell, X. Otazu, and C. A. Parraga. Saliency estimation using a non-parametric low-level vision model. *CVPR*, 2011. 1, 5

[30] B. Nadler and M. Galun. Fundamental limitations of spectral clustering. *NIPS*, 2006. 8

[31] A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. *NIPS*, 2002. 1, 3

[32] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung. Saliency filters: Contrast based filtering for salient region detection. *CVPR*, 2012. 1

[33] Z. Ren, Y. Hu, L.-T. Chia, and D. Rajan. Improved saliency detection based on superpixel clustering and saliency propagation. *ACM Multimedia*, 2010. 1

[34] H. J. Seo and P. Milanfar. Static and space-time visual saliency detection by self-resemblance. *Journal of Vision*, 2009. 1, 5

[35] X. Shen and Y. Wu. A unified approach to salient object detection via low rank matrix recovery. *CVPR*, 2012. 1

[36] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 2007. 5

[37] Q. Yan, L. Xu, J. Shi, and J. Jia. Hierarchical saliency detection. *CVPR*, 2013. 1, 4, 7

[38] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang. Saliency detection via graph-based manifold ranking. *CVPR*, 2013. 1, 2, 3, 5, 7

[39] J. Yang and M.-H. Yang. Top-down visual saliency via joint crf and dictionary learning. *CVPR*, 2012. 1

[40] J. Zhang and S. Sclaroff. Saliency detection: A boolean map approach. *ICCV*, 2013. 1, 7

[41] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell. Sun: A bayesian framework for saliency using natural statistics. *Journal of Vision*, 2008. 1, 5

[42] W. Zhu, S. Liang, Y. Wei, and J. Sun. Saliency optimization from robust background detection. *CVPR*, 2014. 1, 7