Neurocomputing 175 (2016) 336–347

Contents lists available at ScienceDirect

Neurocomputing

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

Robust manifold-preserving diffusion-based saliency detection by adaptive weight construction

Keren Fu [a,b], Irene Y.H. Gu [b], Chen Gong [a], Jie Yang [a,*]

[a] Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, China
[b] Signal Processing Group, Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden

ARTICLE INFO

Article history:
Received 4 June 2015
Received in revised form
13 October 2015
Accepted 19 October 2015
Communicated by Liang Lin
Available online 30 October 2015

Keywords:
Salient object detection
Saliency map
Manifold
Graph-based diffusion

ABSTRACT

Graph-based diffusion techniques have drawn much interest lately for salient object detection. The diffusion performance is heavily dependent on the edge weights in graph representing the similarity between nodes, and are usually set through manually tuning. To improve the diffusion performance, this paper proposes a robust diffusion scheme, referred to as manifold-preserving diffusion (MPD), that is built jointly on two assumptions for preserving the manifold used in saliency detection. The smoothness assumption reflects the conditional random field (CRF) property and the related penalty term enforces similar saliency on similar graph neighbors. The penalty term related to the local reconstruction assumption enforces a local linear mapping from the feature space to saliency values. Graph edge weights in the above two penalties in the proposed MPD method are determined adaptively by minimizing local reconstruction errors in feature space. This enables a better adaption of diffusion on different images. The final diffusion process is then formulated as a regularized optimization problem, taking into account of initial seeds, manifold smoothness and local reconstruction. Consequently, when applied to saliency diffusion, MPD provides a higher performance upper bound than some existing diffusion methods such as manifold ranking. By utilizing MPD, we further introduce a two-stage saliency detection scheme, referred to as manifold-preserving diffusion-based saliency (MPDS), where boundary prior, Harris convex hull, and foci convex hull are employed for deriving initial seeds and a coarse map for MPD. Experiments were conducted on five benchmark datasets and compared with eight existing methods. Our results show that the proposed method is robust in terms of consistently achieving the highest weighted F-measure and lowest mean absolute error, meanwhile maintaining comparable precision–recall curves. Salient objects in different background can be uniformly highlighted in the output final saliency maps.

1. Introduction

Salient object, or region detection is an important research topic in computer vision [1,2]. Given an image, the main aim is to detect and uniformly emphasize objects attracting visual attention in the image, meanwhile suppress irrelevant background. Its applications to vision and graphics are numerous, especially in topics requiring object-level priors such as "proto object" detection [3], segmentation [4,5], content-based image editing [6–9], and image retrieval [10]. In the past decade, a variety of models is proposed, including heuristic color contrast-based models [11,12,9,13–16], learning-based models [17–19], segmentation-assisted approaches [20–23], and graph-based saliency modeling [24–27].

Among graph-based saliency modeling, graph-based diffusion [26–28] has recently been studied for saliency detection with good performance. To conduct saliency diffusion, an input image is first represented by a graph, followed by computing the unified formulation as follows:

$$\mathbf{s} = \mathbf{A}^*\mathbf{y} \tag{1}$$

where $\mathbf{A}^*$ is a global pair-wise propagation matrix, $\mathbf{y}$ is a seed vector that gives a preliminary assessment of saliency level of graph nodes, and $\mathbf{s}$ is the diffused result.

Aiming at improving the diffusion quality and detection performance, this paper proposes a novel and robust diffusion method, referred to as *manifold-preserving diffusion* (MPD). MPD builds jointly upon two assumptions on data manifold, namely the *smoothness* [29–31] and *local reconstruction* [32,33], for better preserving the manifold for saliency detection. The proposed MPD

* Corresponding author.
E-mail addresses: fkrsuper@sjtu.edu.cn (K. Fu), irenegu@chalmer.se (I.Y.H. Gu), goodgongchen@sjtu.edu.cn (C. Gong), jieyang@sjtu.edu.cn (J. Yang).

http://dx.doi.org/10.1016/j.neucom.2015.10.066
0925-2312/

hence is a new way to compute $\mathbf{A}^*$ for the diffusion process in saliency detection. Based on the two assumptions, we introduce two penalties in the diffusion model. As described later, edge weights of graph in the two penalties are determined *adaptively* by solving two optimization problems. This enables better adaption of diffusion on different images. By utilizing MPD, we further introduce a two-stage saliency detection scheme, referred to as manifold-preserving diffusion-based saliency (MPDS), where boundary prior, Harris convex hull, and foci convex hull are employed for deriving initial seeds and a coarse map for MPD. Consequently, better salient object detection can be obtained in various background.

In one of the related studies on diffusion-based methods, Yang et al. [26] propose a manifold ranking-based saliency detector that employs the graph-based manifold ranking [30] to diffuse energy from four image borders. In their work, $\mathbf{A}^*$ has the form $(\mathbf{D}-\alpha\mathbf{W})^{-1}$, where $\mathbf{W}$ denotes the graph affinity matrix with entry w_{ij}, $\mathbf{D}$ is the diagonal degree matrix whose ith diagonal entry is $d_i=\sum_j w_{ij}$, and α is a constant. One can see that $\mathbf{A}^*$ in their case is a deterministic function on $\mathbf{W}$. In [26], manually tuned edge weights of graph are used for $\mathbf{W}$, where the parameter is fixed on all images. Furthermore, only the smoothness assumption is concerned. The proposed MPD differs from [26] by utilizing two assumptions and adaptive weights. Since different images have different contents and color contrast, using manually tuned edge weights is less desirable and can degrade diffusion quality. The proposed method also differs from another work [27], where their diffusion is based on geodesic distance.

The main contributions of this paper are threefold:

(i) We propose an effective graph-based diffusion method: manifold-preserving diffusion (MPD) that jointly exploits the assumptions of smoothness and local reconstruction on the manifold.
(ii) We derive two types of graph edge weights by adaptively minimizing local reconstruction errors in feature space. Hence the method is more suitable to be applied on different images. This is different from previous work where the edge weights of graph are controlled by manually tuned parameter such as bandwidth.
(iii) We introduce a two-stage saliency detection scheme: manifold-preserving diffusion-based saliency (MPDS), that leverages MPD together with boundary prior, Harris convex hull, and foci convex hull. The proposed MPDS achieves better performance than 8 recently published methods on 5 benchmark datasets.

The remainder of the paper is organized as follows. Section 2 reviews related work on salient object detection. Section 3 describes the proposed method in details, including the big picture on the proposed method, graph construction, manifold-preserving diffusion (MPD), and the two-stage manifold-preserving diffusion-based saliency detection (MPDS). Results from experiments and comparisons are given in Section 4. Finally, the conclusion is drawn in Section 5.

2. Related work

We classify existing methods into four categories: heuristic color contrast-based methods, learning-based methods, segmentation-assisted approaches, and graph-based saliency modeling. Methods beyond these four categories fall into the fifth category. For more details, readers are also referred to the comprehensive surveys [1,2].

Heuristic color contrast-based methods: Methods of this category model saliency using local or global color statistics. The underlying assumption is that salient objects are unique in color and present high color contrast to the rest parts of an image. Many methods for computing such contrast-based saliency had been proposed since 2006. Zhai et al. [11] introduce image histograms which only model luminance channel to calculate pixel-level saliency. Achanta et al. [12] provide a saliency approximation by subtracting the average color from low-pass filtered result of an image. This operation of [12] is equivalent to combining center-surround differences of all bandwidth to detect objects of different sizes. Goferman et al. [9] combine local and global features to estimate patch saliency in multi-scales. To consider both local and global factors, they compute saliency of a certain patch as its contrast to the nearest patches in feature space. Under this framework, inner parts of an object are often attenuated due to the edge preference. Cheng et al. [13] extend the method in [11] and incorporate color histograms. A regional contrast saliency measure is proposed as the contrast to other regions. Jiang et al. [14] also use regional contrast to define saliency. Instead, they use only context information from neighborhood of a region. Perazzi et al. [15] propose "saliency filter", which formulates complete contrast and saliency estimation using high dimensional Gaussian filters. Wang et al. [16] compute pixel-wise image saliency by aggregating complementary appearance contrast measures with spatial priors. A more recent method [34] computes contrast-based saliency as dissimilarity/similarity to carefully selected background/foreground seeds. Most of the above contrast-based saliency are straightforward to compute, though the performance is often less satisfactory on images with complex background.

Learning-based methods: Methods in this category estimate image saliency by machine learning. The basic idea is to learn weights of features for saliency computation. Jiang et al. [17] perform pre-segmentation for an input image and extract abundant discriminative features from each region. A random forest regressor trained is adopted to map features to a regional saliency score. Liu et al. [18] segment salient objects by aggregating pixel saliency cues in a conditional random field (CRF). The linear weights for those cues are learned under the maximized likelihood (ML) criteria by tree-reweighted belief propagation. Recently, Wang et al. [35] aim at segmenting objects-of-interest as well but solve the problem in a general joint deep learning framework, where two convolutional neural networks are employed collaboratively to boost the detection and segmentation performance. Mai et al. [19] propose a data-driven approach for aggregating saliency maps output by existing saliency detection models using a CRF. Weights for aggregation are learned in a data-driven way from most similar images retrieved from a pre-defined dataset. Lu et al. [28] learn optimal combination of seeds by maximizing figure-ground segregation. Learning-based methods can achieve good performance in complex scenarios attributed to the learning, however, high computational cost is usually needed due to feature extraction and learning.

Segmentation-assisted methods: Methods in this category aim at generating good segmentation, usually in hierarchy or multi-scale, to facilitate saliency computation. Lu et al. [20] exploit the concavity context in a scene and detect concave arcs from multi-scale segmentation. The detected arcs then contribute to a figure-ground segmentation phase. Yan et al. [21] propose a hierarchical saliency detection method that merges regions according to user-defined scales. Each region in a hierarchy is evaluated by using local contrast and location prior. Cheng et al. [22] measure saliency by hierarchical soft abstraction. They form a 4-layer hierarchical structure including pixel layer, histogram layer, GMM layer and clustering layer with an index table to associate cross-layer relations efficiently. Saliency estimation using color contrast and distribution is conducted on the coarse layers and then propagated to the pixel layer. Jiang et al. [23] find potential salient regions by maximizing a submodular objective function. The problem is solved efficiently by finding a closed-form

harmonic solution on the constructed graph for an input image. The saliency of a region is modeled in terms of appearance and spatial location. These methods, benefiting from some optimized segmentation phases, can make entire objects emphasized and hence boost the final performance.

Graph-based saliency modeling: These methods represent an image by using a graph, where natures of salient objects, such as high color contrast and compact color distribution, are modeled. Gopalakrishnan et al. [24] perform random walks on graphs to find salient objects. The global pop-out and compactness properties of salient objects are modeled in random walks by the equilibrium access time. Wei et al. [25] propose to treat boundary parts of an image as the background. The patch saliency is defined as the shortest geodesic distance on a graph to image boundary. Some other methods propagate saliency energy from labeled seeds to the entire image through graph. Yang et al. [26] propagate saliency via graph-based manifold ranking from four image borders separately. Four saliency maps generated are then multiplied to achieve the final one. Fu et al. [27] perform diffusion from a coarse energy map based on geodesic distance. Since many graph-based diffusion models are related to CRFs, methods of this category can emphasize holistic objects and achieve relatively good performance. The proposed method belongs to this category. We propose a novel graph-based diffusion technique MPD with adaptive weights. Our method is aimed at providing more robust diffusion for saliency detection.

Other methods: Other works include: Shen et al. [36] solve saliency detection as a low rank matrix recovery problem, where salient objects are represented by a sparse matrix (noise) and background is indicated by a low rank matrix. This sparse and low rank assumption may hardly be satisfied in complex scenes, leading to unsatisfactory results. A Bayesian framework is adopted in [37]. First, saliency points are applied to get a coarse location of the saliency region. Based on the rough region, a prior map is computed for the Bayesian model. Margolin et al. [38] define patch saliency as L1-norm in PCA coordinates and combine it with color contrast saliency. Li et al. [39] model patch saliency by dense and sparse reconstruction errors, where the dictionaries for reconstruction are obtained from image boundary.

3. The proposed method

3.1. The big picture on saliency detection

Fig. 1 shows the big picture on saliency detection in this paper. The proposed diffusion method, referred to as the manifold-preserving diffusion (MPD), first involves the adaptive construction of the reconstruction matrix **A** and the affinity matrix **W** from a superpixel-based graph $G=(V,E)$. The estimations of the two matrices are formulated as two optimization problems under some constraints towards minimizing local reconstruction errors in feature space. With **A** and **W** computed, next MPD defines the diffusion process as a regularized optimization problem, taking into account of initial seeds, manifold smoothness and local reconstruction. A closed-form solution exists for MPD. By utilizing MPD, the proposed manifold-preserving diffusion-based saliency (MPDS) incorporates boundary prior, Harris convex hull, and foci convex hull so as to derive initial seeds and a coarse map for diffusion. MPDS is a two-stage detection scheme, whose final output is a final saliency map that highlights salient objects in the image. The following subsections describe the proposed method in detail, including graph construction, manifold-preserving diffusion (MPD), and manifold-preserving diffusion-based saliency (MPDS).

3.2. Image preprocessing and graph construction

We first represent an input image by a graph $G=(V,E)$, where V is a set of vertices (or nodes) and E is a set of graph edges. Given an image, we first over-segment it into n SLIC superpixels [40]. Each superpixel, denoted as $v_i, i\in\{1:n\}$, is treated as a node in V. Superpixels v_i and v_j that satisfy either $\{v_j\in N_i\}$ or $\{\exists k, v_j\in N_k, v_k\in N_i\}$ are connected to form an edge in E, where N_i, N_k denote a set of spatial adjacent superpixels of v_i, v_k, respectively. Such connections lead to a 2-ring graph topology (Fig. 2). Besides, arbitrary boundary superpixels are connected with each other since they are very likely to belong to same background regions. An illustration for this graph structure is shown in Fig. 2.

Hereafter, we use notation "$i\sim j$" to indicate that v_i and v_j are graph neighbors, and "$i\nsim j$" otherwise. We extract a d dimensional feature vector $\mathbf{f}_i$ from each superpixel v_i. In practice, it is the mean CIELab color of each superpixel. We have extracted other color features like RGB color but the performance seems degenerated. This is not surprising because CIELab space characterizes human vision property and is more appropriate for saliency detection [15,13,26].

3.3. The proposed manifold-preserving diffusion (MPD)

In the following subsections, two $n\times n$ matrices **A**, **W** are first defined based on the constructed graph. **W** is a symmetrical affinity matrix with entry w_{ij} encoding similarity between vertices v_i and v_j, and will be used for manifold smoothness assumption. **A** is a reconstruction matrix with entry a_{ij} encoding reconstruction contribution of v_i to v_j. It will be used for local reconstruction on manifold. Note that since only connected nodes lead to a relationship (i. e. Markov random field), w_{ij}/a_{ij} between non-connected superpixels are set to 0, leading to sparse **W** and **A**.

3.3.1. Adaptive estimation of reconstruction matrix *A*

Matrix **A** is related to the manifold reconstruction penalty which is based on the assumption that a node on a graph can be linearly reconstructed by its graph neighbors in feature space. Such linear relationship on high dimensional manifold should be preserved when projecting to lower dimensional space. Inspired by locally linear embedding (LLE) [32], computation of **A** in our work is formulated by minimizing the overall reconstruction error

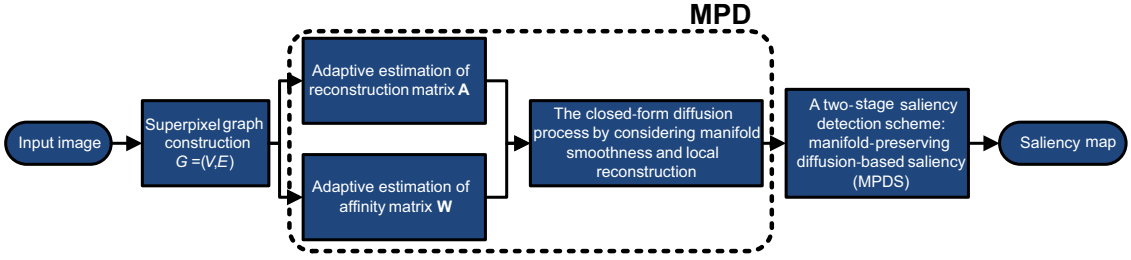


Fig. 1. The big picture on saliency detection in this paper.

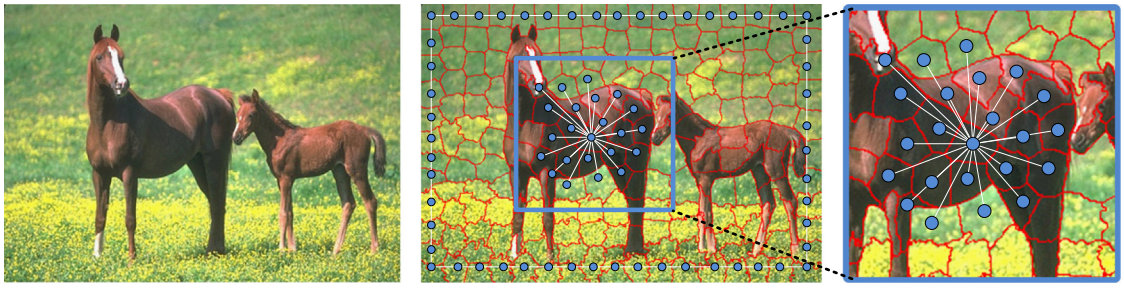


Fig. 2. Superpixel segmentation (superpixel boundaries are in red) and graph topology. Blue circle dots refer to graph vertices and white lines refer to graph edges. For illustrative purpose, only connections as a rectangle around the image boundary is visualized but note that there are connections between arbitrary boundary superpixels. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

as follows:

$$\arg\min_{a_{ij}} \sum_{i=1}^{n} \left\{ \|\mathbf{f}_i - \sum_{j,j\sim i} a_{ij}\mathbf{f}_j\|^2 + \epsilon \sum_{j,j\sim i} a_{ij}^2 \right\} \quad \text{s.t. } \forall i,\ \sum_{j,j\sim i} a_{ij} = 1,\ a_{ii} = 0\ \forall i \not\sim j,\ a_{ij} = 0 \tag{2}$$

where ϵ is a small number for regularization which guarantees unique solution of (2). It is necessary when the number of graph neighbors is larger than the feature dimension [32] (exactly our case). We set $\epsilon = 10^{-4}$ in our implementation. Condition $\sum_{j,j\sim i} a_{ij} = 1$ ensures the reconstruction is linear and shift-invariant [32], i. e. irrelevant to the coordinate origin. If we denote $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n]^T$, the matrix form of the optimization problem above becomes:

$$\arg\min_{\mathbf{A}} \|\mathbf{F} - \mathbf{A}\mathbf{F}\|_{\mathrm{F}}^2 + \epsilon\|\mathbf{A}\|_{\mathrm{F}}^2 \quad \text{s.t. } \mathbf{A}\mathbf{1} = \mathbf{1}\ \forall\ i,\ a_{ii} = 0,\ \forall i \not\sim j,\ a_{ij} = 0 \tag{3}$$

where $\mathbf{1}$ is an all–one vector. We solve each row of $\mathbf{A}$ independently according to [32], and there is an analytical solution. Noting that the difference between our case and [32] is that the numbers of non-zero entries are different in different rows of $\mathbf{A}$ in our case. The main reason is that since the superpixel-based graph is constructed for saliency detection, the numbers of graph neighbors for different nodes may not be identical, as compared with the commonly used k-nearest neighbor graph in machine learning.

3.3.2. Adaptive estimation of affinity matrix $\boldsymbol{W}$

Symmetrical matrix $\mathbf{W}$ encodes similarity between graph nodes. In machine learning, the most common way to define such similarity is by Gaussian kernel, i. e. $w_{ij} = \exp(-\|\mathbf{f}_i - \mathbf{f}_j\|^2/2\sigma^2)$, where the standard deviation σ is related to the bandwidth of similarity function (how "close" between samples is enough "close"). A similar function was used in [26], however, σ was a fixed parameter that is empirically determined. Their disadvantage is that input images usually have different contrast and color statistics. Hence, it is unlikely that a single value σ fits well to all images. In the proposed method, we also utilize Gaussian weight for w_{ij}, however, we select an adaptive and parameter-free solution to enhance the adaptability. Furthermore, instead of assigning the same bandwidth σ to all feature components, we set $\{\sigma = \sigma_1, \ldots, \sigma_d\}$. This enables one to put different weights on different feature components. The affinity function then takes the following form:

$$w_{ij} = \exp\left\{ -\sum_{l=1}^{d} \frac{(f_{il} - f_{jl})^2}{2\sigma_l^2} \right\} \tag{4}$$

where f_{il} is the lth element of feature vector $\mathbf{f}_i$, and d is the total dimension of $\mathbf{f}_i$. Determining σ adaptively is an ill-posed problem due to the lack of prior knowledge. Although in machine learning there are heuristics to determine σ automatically, such as the median heuristics [41] and local scaling [42], they do not work well on our superpixel graph as they are highly sensitive to the graph topology. Inspired by the aforementioned LLE [32] and also the recent advance in adaptive edge weighting [43], following the reconstruction assumption similar to (2), the estimation of $\{\sigma_1, \sigma_2, \ldots, \sigma_d\}$ is formulated by minimizing the reconstruction error as follows:

$$\arg\min_{\{\sigma_1,\sigma_2,\ldots,\sigma_d\}} \|\mathbf{F} - \mathbf{D}^{-1}\mathbf{W}\mathbf{F}\|_{\mathrm{F}}^2 \quad \text{s.t. :}\ \forall\ i,\ w_{ii} = 0;\ \forall i \not\sim j,\ w_{ij} = 0 \tag{5}$$

where $\mathbf{D}$ is the diagonal degree matrix of $\mathbf{W}$. Following [43], gradient descent is used to optimize $\{\sigma_1, \sigma_2, \ldots, \sigma_d\}$. Same as (3), we force the diagonal entries of $\mathbf{W}$ to zeros to avoid self-reconstruction. We start from an initial empirically determined value σ_0 by letting $\sigma_1 = \sigma_2 = \ldots = \sigma_d = \sigma_0$ (usually set conservatively large for common images), and then gradually use steepest gradient descent to optimize them towards smaller values. The iteration is terminated when one of the following conditions is satisfied:

(i) The gradient descent converges.
(ii) The graph turns into a non-connected graph when graph edge weights are $\mathbf{W}$.

Since we will use diffusion technique for saliency detection, condition (ii) is essential so that energy of seeds can be successfully passed to unlabeled nodes. In our implementation, (ii) is checked by examining whether the second smallest eigenvalue of graph Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is below a small threshold (10^{-3}). The rationale behind is that the multiplicity of zero eigenvalues of $\mathbf{L}$ equals the number of connected components in a graph [44]. In most cases, we find condition (ii) is reached prior to condition (i). It is worthy noting that although our method involves an initial value σ_0, this parameter can be set to a conservative value. Hence, the method is not sensitive to this initial value when using the optimized σ for the diffusion. Fig. 3 shows an example of determining the affinity matrix $\mathbf{W}$.

3.3.3. The closed-form diffusion process

Given a graph $G = (V, E)$ with adaptively determined $\mathbf{A}$ and $\mathbf{W}$, we define the diffused energy $\mathbf{s}$ by minimizing:

$$\arg\min_{\mathbf{s}} \underbrace{\mu \sum_{i=1}^{n} k_i(s_i - y_i)^2}_{\text{weighted fitness term}} + \underbrace{\sum_{i=1}^{n} \sum_{j,j\sim i} \frac{1}{2} w_{ij}(s_i - s_j)^2}_{\text{manifold smoothness}} + \underbrace{\lambda \sum_{i=1}^{n} (s_i - \sum_{j,j\sim i} a_{ij}s_j)^2}_{\text{manifold reconstruction}} \tag{6}$$

where $\lambda \geq 0$, $\mu > 0$ are balancing weights. s_i, y_i are the ith elements of $\mathbf{s}$ and $\mathbf{y}$ ($\mathbf{y}$ is a pre-defined seed vector, see also (1)), respectively, and $k_i > 0$ is the weighting coefficient for the ith node.

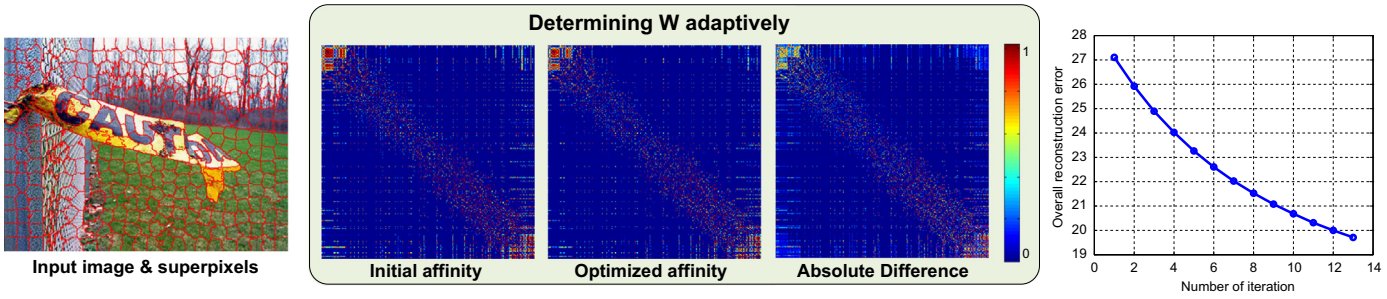


Fig. 3. Determining **W** adaptively as illustrated in Section 3.3.2. From left to right are: original input & superpixel segmentation (about 400 superpixels), initial affinity matrix ($\sigma_0 = 20$), optimized affinity matrix by gradient descent, and the absolute difference matrix between them (all scaled to [0,1] and visualized in pseudo colors), tendency of reconstruction error during iteration. In this example, 13 iterations are required and σ after optimization equals {12.92, 15.55, 16.94} for CIELab color feature.

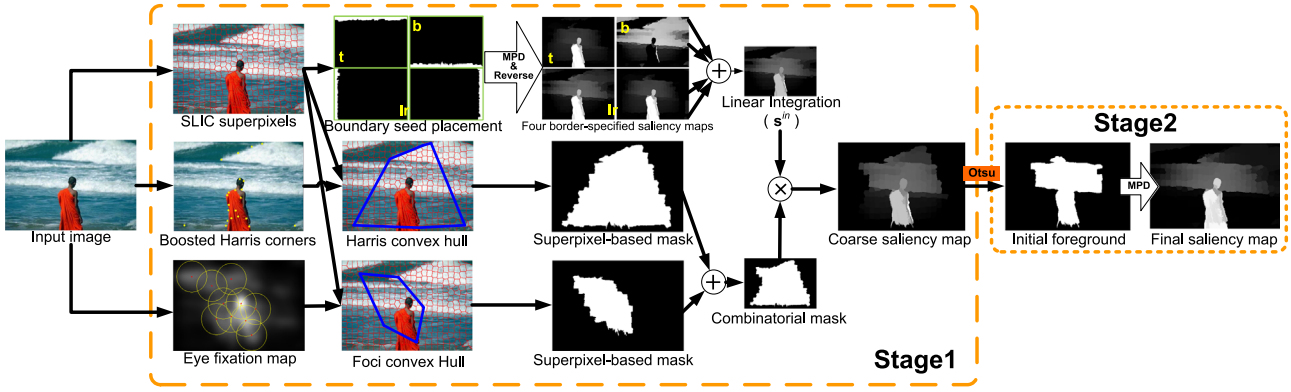


Fig. 4. The block diagram of the proposed two-stage saliency detection scheme MPDS that utilizes MPD. In the top line, "t,b,l,r" are short for "top, bottom, left, right", respectively; In the middle line, boosted Harris corners are shown as yellow dots. After removing corners closed to image boundary, there remain 22 corners; In the third line, yellow circles indicate the range of focus of attention [48] and their centers (red dots) are foci locations. Operation "+" stands for superpixel-wise "or" and "×" stands for superpixel-wise multiplication. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

In (6), the first term is related to the fitness that the diffused energy **s** fits to the original seed **y** [29,31]. The second term is the penalty to enforce the manifold smoothness assumption so that the diffused energy **s** varies smoothly on the manifold [31,29], or, nodes connected by large weights w_{ij} should have similar diffusion labels. The third term is the penalty to enforce the local reconstruction assumption where the diffusion energy is assumed to maintain on the same manifold as in the initial feature space (i. e. sharing the same reconstruction weights). The underly assumption is that there exists a local linear mapping from the feature space to the diffused energy space, $s_i = \mathbf{c}^T\mathbf{f}_i + b$, where **c** and b are the vector and bias in the linear model. If $\mathbf{f}_i = \sum_{j,j\sim i} a_{ij}\mathbf{f}_j$, then $s_i = \sum_{j,j\sim i} a_{ij}s_j$ holds. Such a linear model assumption is similar to that in [45] though the latter is used globally, and also in image matting [46] and filtering [47]. In contrast, the assumption here is used differently as the penalty to enforce the local reconstruction assumption on the manifold-based diffusion and jointly employed with the other two terms to preserve the manifold structure. Furthermore, the proposed manifold reconstruction method is related to the linear neighborhood propagation [33] in semi-supervised learning.

Recall a_{ij}/w_{ij} between non-connected nodes are zeros. Equation (6) then can be rewritten equivalently in a matrix form as

$$\arg\min_{\mathbf{s}} \mu(\mathbf{s}-\mathbf{y})^T\mathbf{K}(\mathbf{s}-\mathbf{y}) + \mathbf{s}^T\mathbf{L}\mathbf{s} + \lambda(\mathbf{s}-\mathbf{A}\mathbf{s})^T(\mathbf{s}-\mathbf{A}\mathbf{s}) \tag{7}$$

where $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the Laplacian of the graph, **D** is the diagonal degree matrix of **W**, and **K** is an $n \times n$ diagonal weighting matrix (usually set to **I** or **D**. We chose $\mathbf{K} = \mathbf{D}$). By taking the derivative on (7) and setting it equal to zero, the following solution for the diffused energy is obtained:

$$\mathbf{s} = [\mu\mathbf{K} + \mathbf{L} + \lambda(\mathbf{I}-\mathbf{A})^T(\mathbf{I}-\mathbf{A})]^{-1}\mathbf{K}\mathbf{y} \tag{8}$$

Comparing with (1), it is easy to see that $\mathbf{A}^* = [\mu\mathbf{K} + \lambda(\mathbf{I}-\mathbf{A})^T(\mathbf{I}-\mathbf{A}) + \mathbf{L}]^{-1}\mathbf{K}$. Noting that with $\mu > 0$ and $\lambda \geq 0$, $\mathbf{A}^* = \mu\mathbf{K} + \lambda(\mathbf{I}-\mathbf{A})^T(\mathbf{I}-\mathbf{A}) + \mathbf{L}$ is invertible since both **L** and $(\mathbf{I}-\mathbf{A})^T(\mathbf{I}-\mathbf{A})$ are semi-positive definite, and **K** is positive definite. Hence, to this end, we obtain (8) as the closed-form solution to the diffusion process and can be used for the subsequent saliency detection.

3.4. Manifold-preserving diffusion-based saliency (MPDS)

In this section, a two-stage detection scheme is proposed, referred to as *manifold-preserving diffusion-based saliency (MPDS)*, to leverage the proposed MPD for saliency detection. Boundary prior, Harris convex hull, and foci convex hull are incorporated during the coarse saliency map estimation. The block diagram of the scheme is shown in Fig. 4. For the boundary prior, by specifying each image border as the background seeds, MPD is used to perform diffusion and generate four intermediate saliency maps. After the linear integration of four maps, the obtained map is incorporated with the binary mask generated by the Harris convex hull and the foci convex hull (see the "Combinatorial mask" in Fig. 4). More specifically, two binary masks are first generated separately from the Harris convex hull and the foci convex hull. Masking values of superpixels inside each hull are set to 1, and those of superpixels outside each hull are set 0. Then, the two binary masks are combined by superpixel-wise "or" operation (denoted as "+" in Fig. 4). Saliency values of superpixels outside this combined mask are cropped by superpixel-wise multiplication (denoted as "×" in Fig. 4), resulting in a coarse saliency map in the first stage. In the second stage, we use Ostu's method [49] to adaptively select an initial foreground region that tends to include salient objects. MPD is then utilized again for diffusion from the initial foreground region, resulting in a final saliency map

that emphasizes an entire salient object. Below, details of the three parallel processes in Fig. 4 are described.

Boundary prior: Image borders are usually considered as the background. For four image borders, four seed vectors $\mathbf{y}^t$, $\mathbf{y}^b$, $\mathbf{y}^l$, $\mathbf{y}^r$ are specified, respectively (where the superscript "t, b, l, r" indicates "top, bottom, left, right", respectively). Taking the case of the upper border as an example, the component in $\mathbf{y}^t$ is set to 1 if the corresponding suerpixel touches a margin of 5-pixels near upper image border, and is set to 0 otherwise. In a similar way, $\mathbf{y}^b, \mathbf{y}^l, \mathbf{y}^r$ can be set. Let $\mathbf{A}^*$ be the propagation matrix (see (8)), an intermediate saliency map that is based on the boundary prior is defined as

$$\mathbf{s}^{in} = \sum_{o \in \{t,d,l,r\}} \{1 - \mathcal{N}(\mathbf{A}^* \mathbf{y}^o)\} = 4 - \sum_{o \in \{t,d,l,r\}} \mathcal{N}(\mathbf{A}^* \mathbf{y}^o) \tag{9}$$

where $\mathcal{N}(\cdot)$ is a normalization operator that scales a map to [0,1], $1-\mathcal{N}(\cdot)$ reverses the map as the diffused energy indicates the likelihood to be the background. Noting that the integration of four maps is motivated by [26], however, we use the linear integration instead of multiplication. The rationale is that multiplication can easily cause an object heavily suppressed if it touches any image border. Intuitively, the linear integration is more conservative, since it links with the phenomenon that *the fewer borders that an object touches, the more salient the object is*.

Harris convex hull: Boosted Harris corners are used in the scheme to generate the Harris convex hull in each given image. We use 30 corners to specify a coarse coverage of objects. After removing corners that are fairly close to image borders (again in a margin of 5 pixels near image borders), a convex hull is generated to enclose all corner points. Although Harris convex hull prior was used in [37,27,50], the Harris convex hull is employed in this paper for a different purpose for excluding potential background.

Foci convex hull: We also employ a foci convex hull based on an attention map predicted by some eye-fixation model (e. g. [48,3]), in order to compensate the possible failure of Harris convex hull for covering parts of an object. Our motivation of exploiting the attention model is that it tends to be selective and often detects salient corners/edges of object [1]. Thereby a convex hull that encloses all foci points has a high probability to cover a salient object. We choose the standard Itti's model [48] to produce an attention map. A typical ten foci points are sampled from the attention map similar to that in [48], where the foci radius is set to 1/6 of the minimum dimension of image (to emulate the inhibition-of-return behavior of human eyes). A foci convex hull is then generated similarly as the Harris convex hull (Fig. 4). The foci convex hull is used jointly with the Harris convex hull to cover salient objects robustly, meanwhile exclude irrelevant background.

4. Experimental results and performance evaluations

4.1. Setup

We use $n=400$ superpixels for an input image. There are three important parameters in our MPD scheme, i. e. μ, λ and initial σ_0, which are empirically set. First, σ_0 is conservatively set to 20 for un-normalized CIELab color space, which we find suitable for most images. The terminating conditions in Section 3.3.2 then can be reached usually in fewer than 100 iterations. μ for the data term can be set to a small number [29,30], and we use $\mu = 0.01$. A relatively complicated parameter is λ, which controls the power of local reconstruction penalty. In practice, we use $\lambda = 0.1$ for all tests.

4.2. Evaluation metrics, methods for comparison, and datasets

The performance evaluation is conducted using three metrics:

(a) *Precision–recall curve* is defined as (the same as in [12,13,15,26,21,1]):

$$\text{Precision}(T) = \frac{|M(T) \cap G|}{|M(T)|}, \ \text{Recall}(T) = \frac{|M(T) \cap G|}{|G|} \tag{10}$$

where $M(T)$ is the binary mask obtained by directly thresholding a saliency map (denoted as S_{map}) using threshold T, G is the ground truth map having the same size as S_{map}, and $|\cdot|$ is the total area of a mask.

(b) *Mean Absolute Error (MAE)* is defined as (the same as in [22,15]):

$$MAE = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} |S_{map}(x,y) - G(x,y)| \tag{11}$$

where $S_{map}(x,y)$ is the saliency value and $G(x,y)$ is the ground truth value at the spatial location (x,y), W and H are the width and height of the map S_{map}, respectively.

(c) *Weighted F-measure* F_β^w is adopted from the recently proposed definition in [51]:

$$F_\beta^w = \frac{(1+\beta^2)\text{Precision}^w \times \text{Recall}^w}{\beta^2 \times \text{Precision}^w + \text{Recall}^w} \tag{12}$$

where Precision^w and Recall^w are the weighted precision and recall.

The main difference between (10) and (12) is that Precision^w and Recall^w in (12) can directly compare a non-binary map against a binary ground truth without thresholding. As demonstrated in [51], the weighted F-measure (F_β^w) gives more reliable evaluation that meets the human perception. Fig. 5 shows the evaluation using the metric in (12).

Visual observation on the results in Fig. 5 shows that the proposed method has generated much better performance than that from GMR [26]. This evidence is clearly reflected on F_β^w, i. e. 0.957 versus 0.539. However, in the precision–recall curves, the difference is small. In some places, the curve of GMR even has a slightly

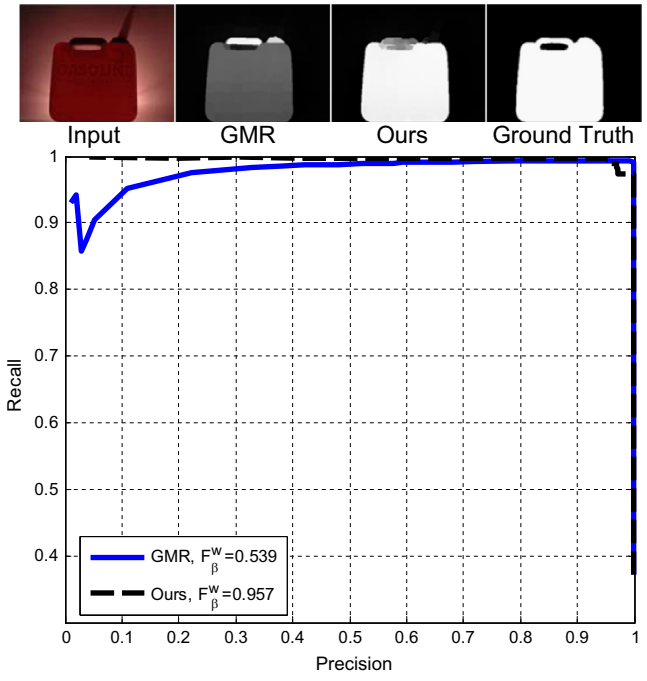


Fig. 5. An example showing deficiency of the precision–recall curve. Though our method produces a saliency map much closer to the binary ground truth than GMR [26], the precision–recall curve hardly reflects such fact. In contrast, a large difference on F_β^w is observed.

better performance than our method (see the right corner of the curve in Fig. 5). Therefore, we use F_β^w as a complementary measure to (10). We use the evaluation code from ⟨http://cgm.technion.ac.il/Computer-Graphics-Multimedia/Software/FGEval/⟩, where $\beta^2 = 1$ is set without bias between the precision and the recall. Although $\beta^2 = 0.3$ is suggested in [12] to weigh the precision more than the recall, it is shown in [51] that the overall evaluation of a saliency detection method should dedicate to applications. Viewing that in applications such as using saliency detection to guide object detection, the recall appears more important than the precision. In addition, the reason for setting β^2 exactly to 0.3 is not clearly stated in [12]. Therefore, we decide to set $\beta^2 = 1$ without bias between the precision and recall. Note that $\beta^2 = 1$ is also used in [52] for evaluation on saliency detection. Among the above three metrics, a high precision–recall curve, low MAE, and high F_β^w indicate a good saliency model.

We compare MPDS with 8 recent state-of-the-art methods which output real-valued saliency maps including: CB (Context-Based) [14], GS (Geodesic Saliency) [25], HS (Hierarchical Saliency) [21], PCA [38], GC (Global Cue) [22], DSR (Dense and Sparse Reconstruction) [39], GMR (Graph-based Manifold Ranking) [26], PISA (Pixelwise Image Saliency by Aggregation). Note that we adopt the public implementations from the original authors for all the methods and all saliency maps are scaled into the range [0, 1] for the unified evaluation. Although other methods exist, most of them are not as good as the above selected methods.

Tests and comparisons were conducted on five commonly used benchmark datasets, including MSRA-1000 [12] (1000 images), SOD [53] (300 images), SED1 [54] (100 images), SED2 [54] (100 images), and ECSSD [21] (1000 images with complex and texture scenes).

4.3. Validation of MPD

Firstly, we validate the impact of λ. The quantitative evaluation on MSRA-1000 and ECSSD by varying λ from 0 to 100 is shown in Fig. 6. One can see that incorporating the local reconstruction penalty is helpful for improving the precision–recall curve meanwhile maintaining similar F_β^w and MAE. However too large λ (e. g. 10 and 100) in turn causes performance degenerated, which are reflected on all three criteria. This is because such strong assumption on linear model may be violated in some cases. Note the linear model may be contradictory to the purpose of uniform enhancement. For example, for two adjacent but distinctive regions both of which we want to highlight uniformly (i. e. have saliency values close to 1), the linear model may cause counteraction due to discrepancy after the same linear mapping. On the other hand, such linear model can be helpful for distinguishing object and background as long as they present different features, i. e. $\mathbf{f}_i$. Fig. 7 shows such changes visually. By considering local reconstruction, large areas on objects tend to be well highlighted since the saliency of a node (i. e. superpixel) which lies inside can be better reconstructed due to large amount of similar neighbors. From both $\lambda = 0$ and 100, complementary effect can be observed. Since $\lambda = 0.1$ is found generating fairly good results in Figs. 6 and 7, we use it for all the tests.

We also show MPD is better than GMR [26] when applied to saliency detection. To validate this, we compare MPD with GMR under the same configuration, namely by applying MPD and GMR to the same graph structure used in this paper and also to the same seeds. We conducted two experiments. The one was by diffusion from four image borders followed by linear integration, yielding to $\mathbf{s}^{in}$ in the top line in Fig. 4. The other is by diffusion from given ground truth masks, which provides an intuitive estimation of performance *upper bound*. Fig. 8 shows the diffusion evaluation on MSRA-1000 and ECSSD. It is obvious that MPD achieves better performance in terms of all three criteria. The difference on F_β^w is even more significant than that on the precision–recall curves. The reason for this has been demonstrated in Section 4.2. The better results on diffusion from ground truth reveal that the proposed MPD guarantees a higher upper bound (6.7% gain on MSRA and 8.3% gain on ECSSD in terms of F_β^w) than GMR, which allows more potential space for future improvement. Recall that the major differences between MPD and GMR lie In: (1) MPD estimates $\mathbf{A}$ and $\mathbf{W}$ adaptively whereas GMR uses manually determined σ for $\mathbf{W}$; (2) MPD applies the additional local reconstruction assumption as compared to GMR. Another phenomenon indicated in Fig. 8 is that compared to dataset MSRA-1000, the margin between the detection performance and the upper bound on dataset ECSSD is much larger. This implies ECSSD is much harder than MSRA-1000 for saliency detection, and there is more space for further improvement.

4.4. Validation of individual components

To evaluate the effectiveness of using Harris convex hull and foci convex hull, experiments were conducted on MSRA-1000 dataset by ablating these two components from our full implementation in Fig. 4. Results are shown in Fig. 9. One can see from Fig. 9 that the performance without both hull priors degenerates in a noticeable margin. Incorporating either hull prior leads to a performance boost. Additionally, removing Harris convex hull influences the overall performance more than removing the foci

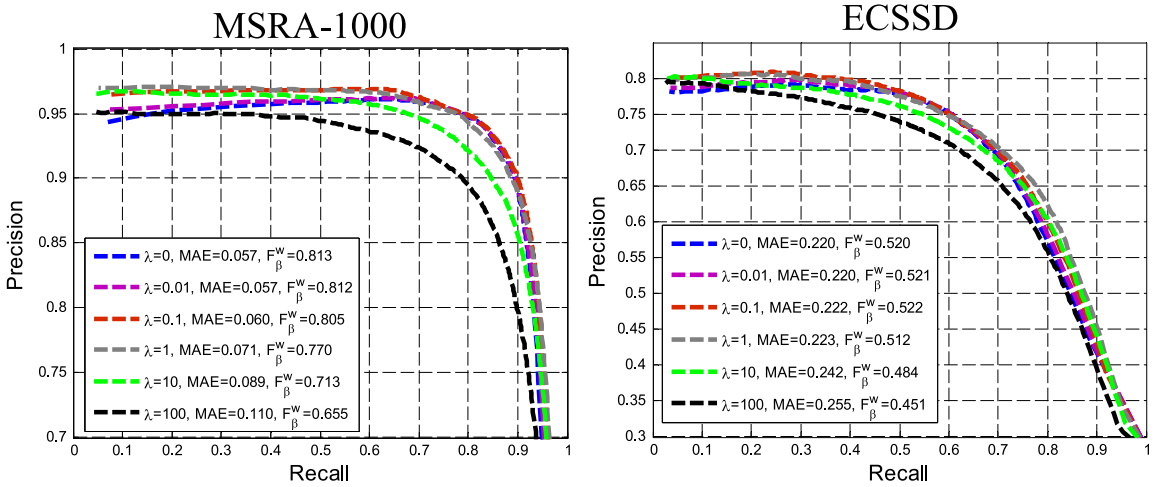


Fig. 6. Quantitative evaluation on MSRA-1000 (left) and ECSSD datasets (right) by tuning λ from 0 to 100.

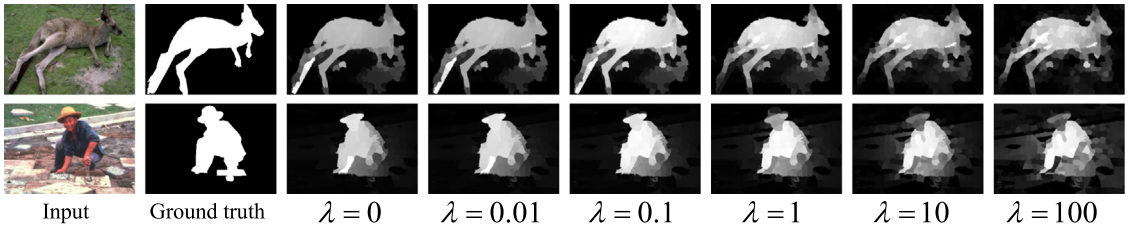


Fig. 7. Visual impact of parameter λ. Moderately incorporating λ can emphasize an entire salient object and avoid over-suppressing in cluttered background.

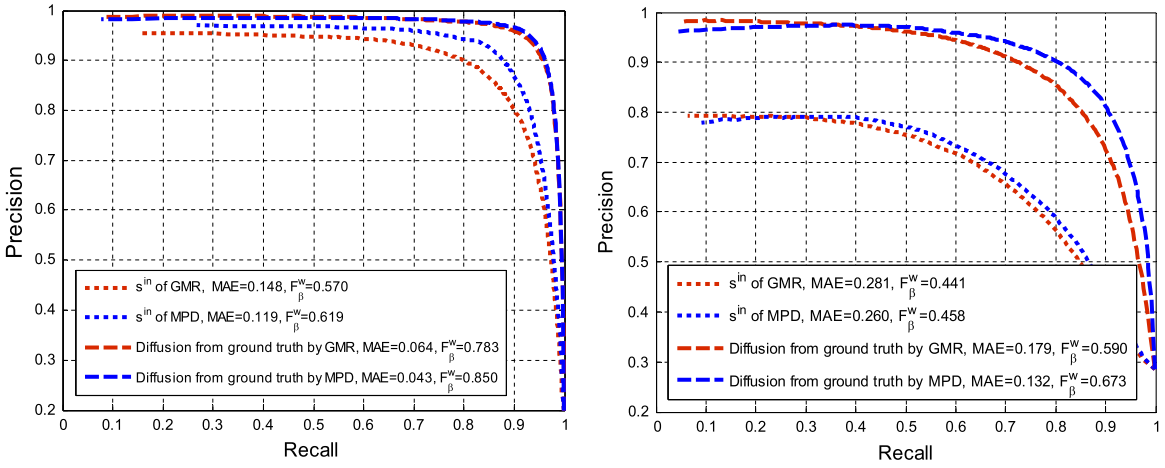


Fig. 8. Diffusion evaluation between MPD and GMR [26] on MSRA-1000 (left) and ECSSD (right) datasets.

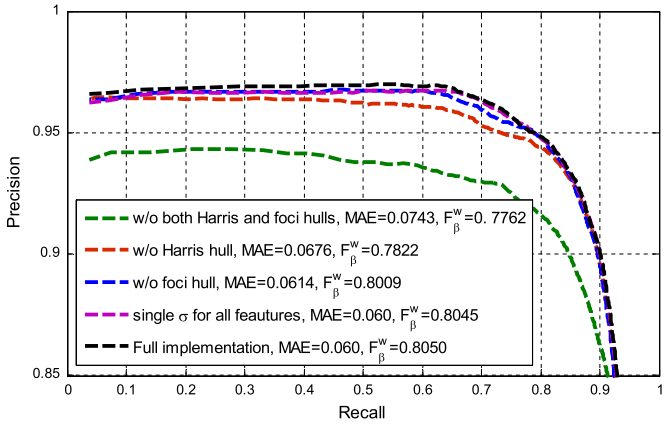


Fig. 9. Validation of Harris convex hull and foci convex hull on MSRA-1000 dataset. The performance by using a single σ for all feature dimensions is also compared.

convex hull, indicating that Harris hull is more informative. Further incorporating foci convex hull with Harris convex hull as shown in Fig. 4 slightly enhances the final performance.

Besides, we also compared using a single σ for all feature channels to using different σ, i. e. $\{\sigma = \sigma_1, \ldots, \sigma_d\}$. Note that if a single σ is used in (4), the solution to (5) can also be obtained through gradient descent. As shown in Fig. 9, allowing σ to vary (i. e. the full implementation) leads to only marginal improvement comparing to using a single σ, though the difference between these two cases is expected to be large. Intuitively $\{\sigma = \sigma_1, \ldots, \sigma_d\}$ enables weighing different feature channels adaptively. We find the factors causing this phenomenon are twofold. Firstly, on some images the both cases will lead to similar σ when gradient descent stops. For example in Fig. 3, the result of using a single σ is 15.87, and it is similar to $\{\sigma = 12.92, 15.55, 16.94\}$. Secondly, the two-stage scheme in Fig. 4 could also reduce the overall impact of σ.

4.5. Comparisons to state-of-the-art methods on saliency detection

The performance of the proposed saliency detection scheme is further examined by comparing with 8 recent state-of-the-art methods. Fig. 10 shows the results from the quantitative comparisons. To facilitate comparisons, MAE values are sorted in descending order and F_β^w are sorted in ascending order. Several interesting observations can be found from Fig. 10. Firstly, from precision–recall curves, one can see that our method, namely MPDS, achieves comparable precision to the existing methods under the same recall. Noticeable improvement over the HS, DSR, and PISA methods can be observed on the datasets MSRA-1000 and SED1, whereas on the rest datasets, DSR, HS, GMR, PISA, and the proposed MPDS perform similarly.

Secondly, since precision–recall curves do not reflect the highlight level of a whole object (see Fig. 5), the proposed method consistently achieves much smaller MAE and higher F_β^w over most of the compared methods, indicating that MPDS is more capable of enhancing a salient object holistically. This is mainly due to the proposed diffusion method which preserves manifold structure by employing penalty terms to enforce smoothing and local reconstruction assumptions. Fig. 11 shows visual results from the proposed method and from these 8 methods. From Fig. 11, one can see that the proposed saliency detection scheme can effectively suppress the background clutter and uniformly enhance the foreground objects. Comparing with the remaining methods such as GMR and PISA, results from the proposed method are visually closer to the ground truth and maintains much clearer object boundaries.

Some intermediate results from individual steps of the proposed scheme are shown in Fig. 12. Observing Fig. 12(f), i. e., the results of integrating four boundary specified saliency maps in the first stage, the proposed method is shown to be able to coarsely highlight the salient regions in an image. Subsequent combination with two convex hull priors and Ostu's thresholding (Fig. 12(i) and (j)) provide more informative initialization to the second stage.

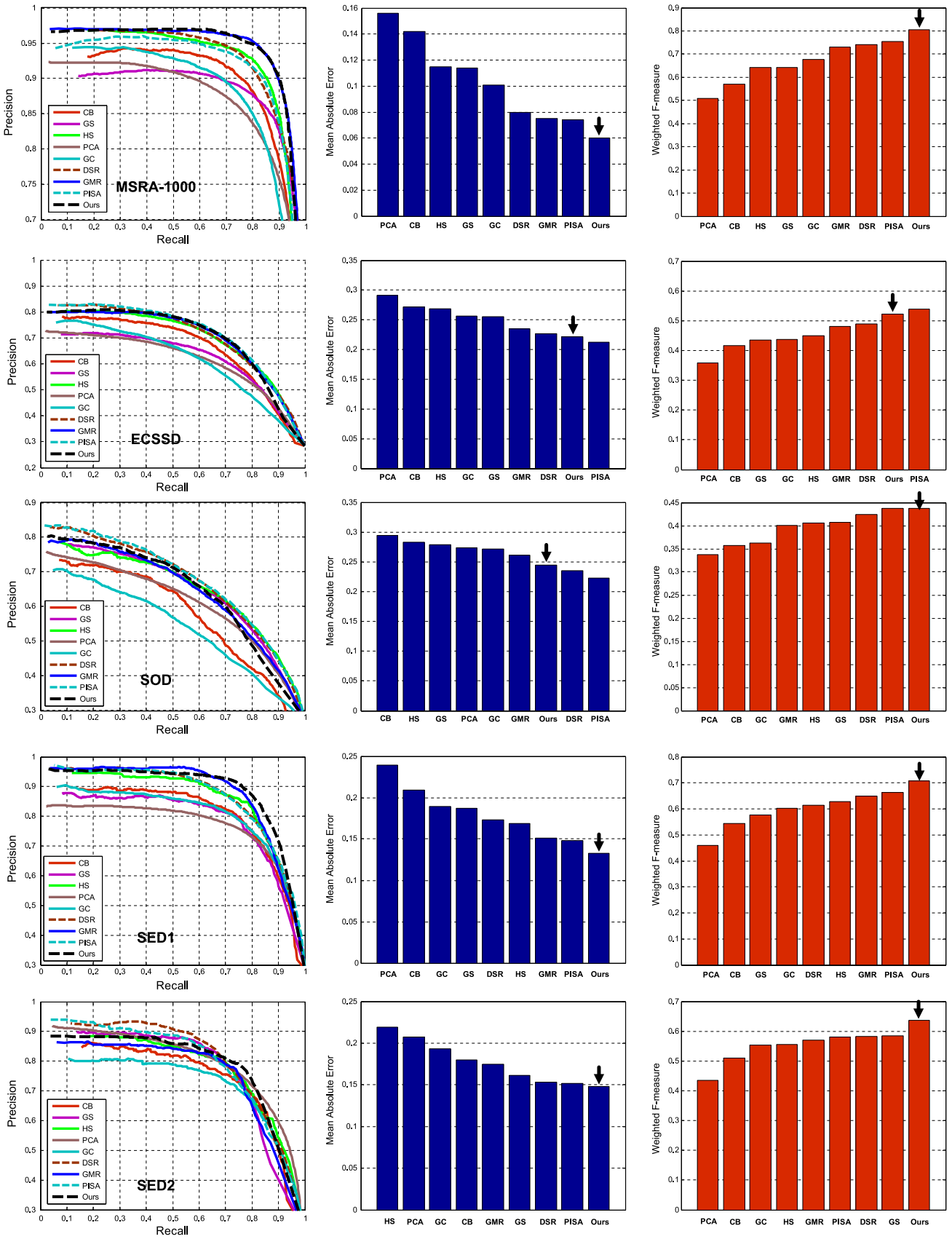


Fig. 10. Quantitative evaluation by precision–recall curves (left), mean absolute error (MAE, middle), weighted F-measure (F_β^w, right) on five benchmark datasets. From top to bottom are MSRA-1000, ECSSD, SOD, SED1, and SED2. The proposed method is highlighted by black arrows in the middle and right sub-figures.

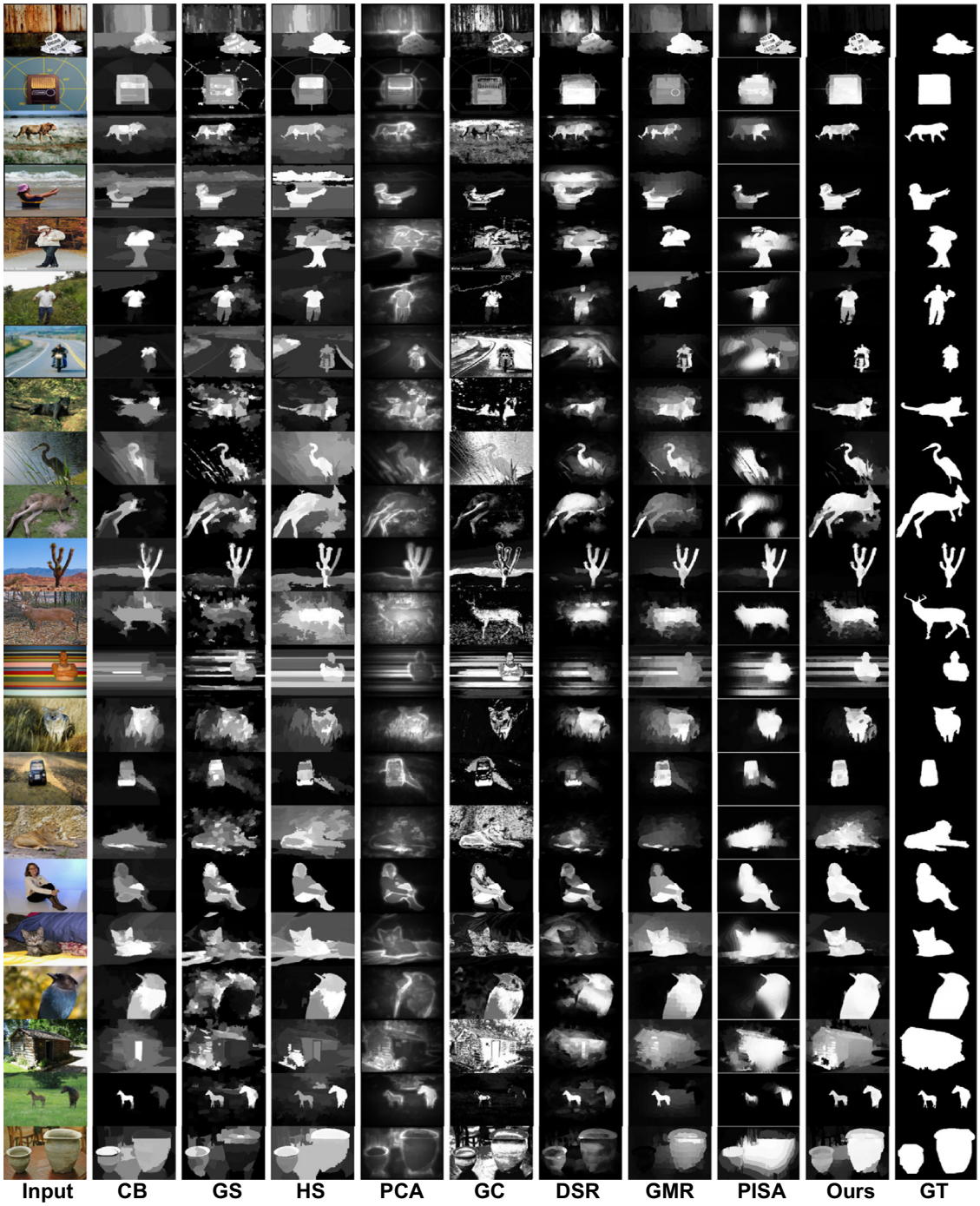


Fig. 11. Visual comparisons. The abbreviations of the methods are listed in Section 4.2.

Background regions with low saliency are discarded. Consequently, good results of final saliency maps (Fig. 12(k)) from the proposed method are observed.

5. Conclusion

In this paper, the proposed novel diffusion scheme called manifold-preserving diffusion (MPD), which uses penalty terms for jointly enforcing the manifold smoothness and the local reconstruction assumptions, has been tested and shown to be effective. The affinity matrix and the reconstruction matrix of the graph in MPD are both determined adaptively. This reduces the required manually turning parameters and enhances the adaptability of diffusion on different images. Experimentally, MPD performs better than manifold ranking when using image boundary as seeds and guarantees a higher performance upper bound. This is probably attributed to introducing penalty terms in both assumptions and also to adaptive weight construction. The proposed two-stage detection scheme (MPDS) by utilizing MPD is also tested. By integrating boundary prior, Harris convex hull and foci convex hull in the scheme, the proposed MPDS maintains comparable performance on precision–recall curves, meanwhile reaches the lowest mean absolute error and the highest weighted F-measure when compared to 8 recent state-of-the-art saliency models. Further works include exploiting MPD in other computer

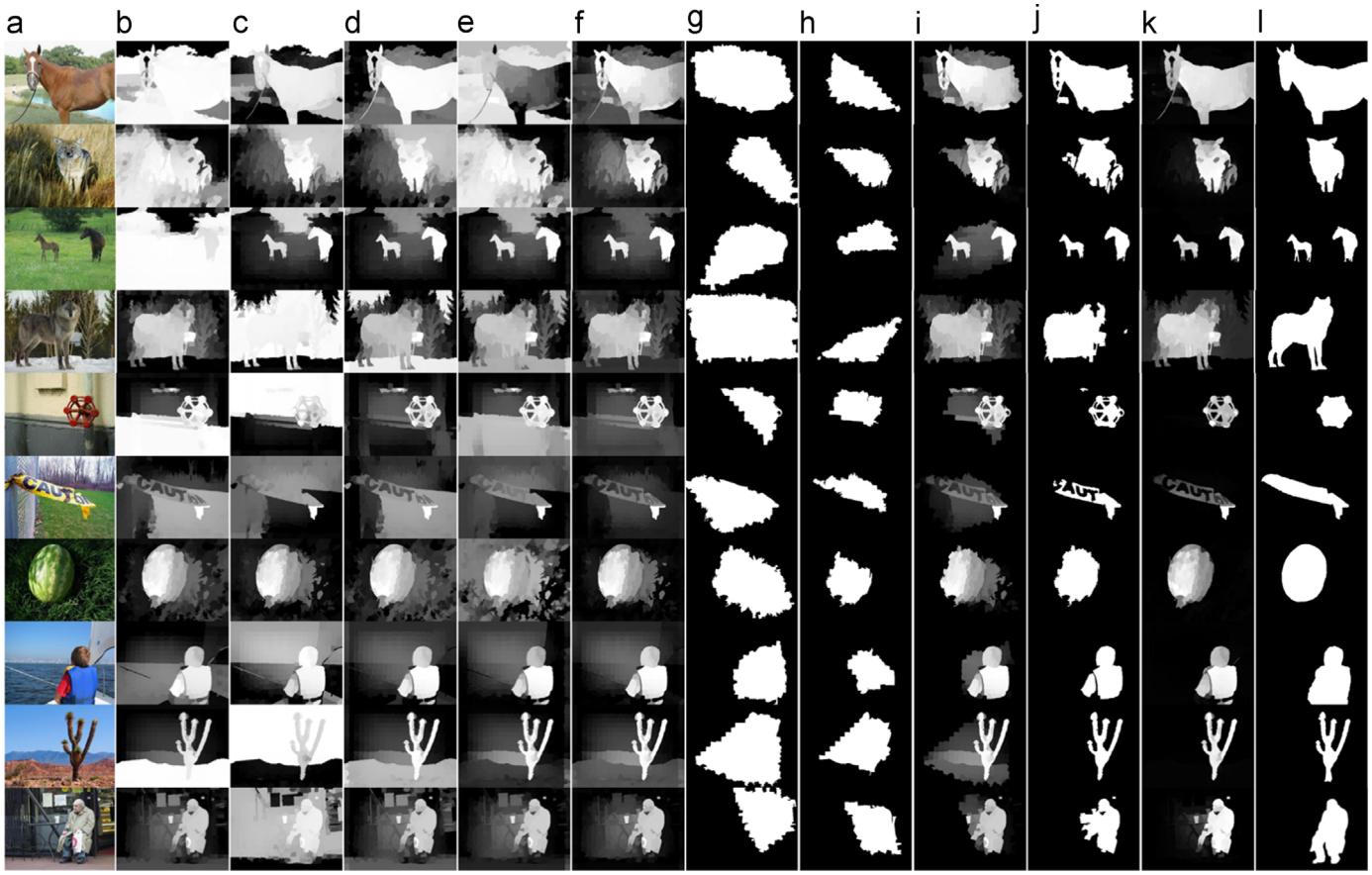


Fig. 12. Visual comparisons of individual steps in our method. (a) Input image. (b)–(e) Diffusion from top, down, left and right image borders, respectively and then reverse. (f) Linear integration ($\mathbf{s}^{in}$). (g) Binary mask by Harris convex hull. (h) Binary mask by foci convex hull. (i) Coarse saliency map of stage 1. (j) Initial foreground map. (k) Final saliency map of our saliency detection scheme after stage 2. (l) Ground truth.

vision areas that call for diffusion, and other cost functions for adaptive edge weights.

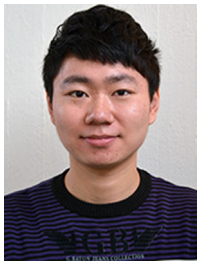
Acknowledgment

This research is partly supported by National Science Foundation, China (No: 61273258) and 973 Plan, China (No: 2015CB856004).

References

[1] A. Borji, D. Sihite, L. Itti, Salient object detection: a benchmark, In: European Conference on Computer Vision (ECCV), 2012.
[2] A. Borji, L. Itti, State-of-the-art in visual attention modeling, IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) 35 (1) (2013) 185–207.
[3] X. Hou, L. Zhang, Saliency detection: a spectral residual approach, In: CVPR, 2007.
[4] E. Rahtu, J. Kannala, M. Salo, J. Heikkila, Segmenting Salient Objects from Images and Videos, In: ECCV, 2010.
[5] L. Wang, J. Xue, N. Zheng, G. Hua, Automatic Salient Object Extraction with Contextual Cue, In: ICCV, 2011.
[6] F. Stentiford, Attention based auto image cropping, In: Workshop on Computational Attention and Applications, ICVS, 2007.
[7] L. Marchesotti, et al., A framework for visual saliency detection with applications to image thumbnailing, In: ICCV, 2009.
[8] Y. Ding, X. Jing, J. Yu, Importance filtering for image retargeting, In: CVPR, 2011.
[9] S. Goferman, et al., Context-aware saliency detection, In: CVPR, 2010.
[10] T. Chen, M. Cheng, et al., Sketch2photo: internet image montage, ACM Trans. Graph 28 (5) (2006) 1–10.
[11] Y. Zhai, M. Shah, Visual attention detection in video sequences using spatiotemporal cues, ACM Multimed. (2006) 815–824.
[12] R. Achanta, S. Hemami, F. Estrada, S. Süsstrunk, Frequency-tuned salient region detection, In: CVPR, 2009.
[13] M. Cheng, G. Zhang, N. Mitra, X. Huang, S. Hu, Global contrast based salient region detection, In: CVPR, 2011.
[14] H. Jiang, J. Wang, et al., Automatic salient object segmentation based on context and shape prior, In: BMVC, 2011.
[15] F. Perazzi, P. Krahenbul, et al., Saliency filters: contrast based filtering for salient region detection, In: CVPR, 2012.
[16] K. Wang, L. Lin, J. Lu, C. Li, K. Shi, Pisa: pixelwise image saliency by aggregating complementary appearance contrast measures with edge-preserving coherence, IEEE Trans. Image Process. (IP) 24 (10) (2015) 3019–3033.
[17] H. Jiang, et al., Salient object detection: a discriminative regional feature integration approach, In: CVPR, 2013.
[18] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, Learning to detect a salient object, TPAMI 33 (2) (2011) 353–367.
[19] L. Mai, Y. Niu, F. Liu, Saliency aggregation: a data-driven approach, In: CVPR, 2013.
[20] Y. Lu, W. Zhang, H. Lu, X. Xue, Salient object detection using concavity context, In: IEEE International Conference on Computer Vision (ICCV), 2011.
[21] Q. Yan, et al., Hierarchical saliency detection, In: CVPR, 2013.
[22] M. Cheng, J. Warrell, et al., Efficient salient region detection with soft image abstraction, In: ICCV, 2013.
[23] J. Lafferty, A. McCallum, F. Pereira, Submodular salient region detection, In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.
[24] V. Gopalakrishnan, et al., Random walks on graphs for salient object detection in images, TIP 19 (12) (2010) 3232–3242.
[25] Y. Wei, F. Wen, W. Zhu, J. Sun, Geodesic saliency using background priors, In: ECCV, 2012.
[26] C. Yang, L. Zhang, et al., Saliency detection via graph-based manifold ranking, In: CVPR, 2013.
[27] K. Fu, C. Gong, I. Gu, J. Yang, Geodesic saliency propagation for image salient region detection, In: ICIP, 2013.
[28] S. Lu, V. Mahadevan, et al., Learning optimal seeds for diffusion-based salient object detection, In: CVPR, 2014.
[29] D. Zhou, et al., Learning with local and global consistency, In: NIPS, 2003.
[30] D. Zhou, et al., Ranking on data manifolds, In: NIPS, 2004.
[31] X. Zhu, A. Goldberg, Introduction to semi-supervised learning, Synthesis lectures on artificial intelligence and machine learning 3 (1) (2009) 1–130.
[32] S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.

[33] J. Wang, F. Wang, C. Zhang, et al., Linear neighborhood propagation and its applications, IEEE Trans. Pattern Anal. Mach. Intell. 31 (9) (2009) 1600–1615.
[34] J. Wang, H. Lu, X. Li, N. Tong, W. Lei, Saliency detection via background and foreground seed selection, Neurocomputing 152 (2015) 359–368.
[35] X. Wang, L. Zhang, L. Lin, Z. Liang, W. Zuo, Deep joint task learning for generic object extraction, In: NIPS, 2014.
[36] X. Shen, Y. Wu, A unified approach to salient object detection via low rank matrix recovery, In: CVPR, 2012.
[37] Y. Xie, H. Lu, Visual saliency detection based on Bayesian model, In: ICIP, 2011.
[38] R. Margolin, et al., What makes a patch distinct, In: CVPR, 2013.
[39] X. Li, H. Lu, L. Zhang, X. Ruan, M. Yang, Saliency detection via dense and sparse reconstruction, In: ICCV, 2013.
[40] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Susstrunk, Slic superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) 34 (11) (2012) 2274–2282.
[41] A. Gretton, K. Borgwardt, M. Rasch, et al., A kernel method for the two-sample-problem, In: NIPS, 2006.
[42] L. Zelnik-Manor, P. Perona, Self-tuning spectral clustering, In: NIPS, 2004.
[43] M. Karasuyama, H. Mamitsuka, Manifold-based similarity adaptation for label propagation, In: NIPS, 2013.
[44] U. von Luxburg, A tutorial on spectral clustering, Stat. Comput. 17 (4) (2007) 395–416.
[45] J. Kim, D. Han, Y. Tai, et al., Salient region detection via high-dimensional color transform, In: CVPR, 2014.
[46] A. Levin, D. Lischinski, Y. Weiss, A closed-form solution to natural image matting, IEEE Trans. Pattern Anal. Mach. Intell. 30 (2) (2008) 228–242.
[47] K. He, J. Sun, X. Tang, Guided image filtering, IEEE Trans. Pattern Anal. Mach. Intell. 35 (6) (2013) 1397–1409.
[48] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) 20 (11) (1998) 1254–1259.
[49] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. 9 (1) (1979) 62–66.
[50] C. Yang, L. Zhang, H. Lu, Graph-regularized saliency detection with convex-hull-based center prior, Signal Process. Lett. 20 (7) (2013) 647–648.
[51] R. Margolin, L. Zelnik-Manor, A. Tal, How to evaluate foreground maps, In: CVPR, 2014.
[52] Z. Liu, W. Zou, O. Meur, Saliency tree: a novel saliency detection framework, IEEE Trans. Image Process. 23 (5) (2014) 1937–1952.
[53] V. Movahedi, J. Elder, Design and perceptual validation of performance measures for salient object segmentation, In: IEEE Computer Society Workshop on Perceptual Organization in Computer Vision, 2010.
[54] S. Alpert, M. Galun, et al., Image segmentation by probabilistic bottom-up aggregation and cue integration, In: CVPR, 2007.



Keren Fu received his bachelor degree from Huazhong University of Science and Technology (HUST), China in 2011. Currently he is a dual doctorial degree student at the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University (SJTU), and Department of Signals and Systems, Chalmers University of Technology, Sweden, under the supervision of Prof. Jie Yang and Prof. Irene Yu-Hua Gu. In 2014, he was awarded the Degree of Licentiate of Engineering from Chalmers University of Technology, Sweden. His research areas are computer vision and image/video modeling, with applications to, e. g. visual saliency detection, object tracking and detection, traffic analysis, machine learning.

Irene Y.H. Gu received the Ph.D. degree in electrical engineering from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 1992. From 1992 to 1996, she was a Research Fellow at Philips Research Institute IPO, Eindhoven, The Netherlands, and post Dr. at Staffordshire University, Staffordshire, U. K., and Lecturer at the University of Birmingham, Birmingham, U.K. Since 1996, she has been with the Department of Signals and Systems, Chalmers University of Technology, Göteborg, Sweden, where she is currently a full Professor. Her research interests include statistical image and video processing, object tracking and video surveillance, pattern classification, and signal processing with applications to electric power systems. Dr. Gu was an Associate Editor for the IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, and Part B: Cybernetics from 2000 to 2005. She was the Chair-elect of the IEEE Swedish Signal Processing Chapter from 2002 to 2004. She has been an Associate Editor with the EURASIP Journal on Advances in Signal Processing since 2005, and with the Editorial board of the Journal of Ambient Intelligence and Smart Environments since 2011.

Chen Gong received his bachelor degree from East China University of Science and Technology (ECUST) in 2010. Currently he is a dual doctorial degree student at the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University (SJTU), and the Centre for Quantum Computation & Intelligent Systems, University of Technology, Sydney (UTS), under the supervision of Prof. Jie Yang and Prof. Dacheng Tao. His research interests mainly include machine learning, data mining and learning-based vision problems. He has published 21 technical papers at prominent journals and conferences such as IEEE T-NNLS, IEEE T-CYB, CVPR, AAAI, ICME.

Jie Yang received his Ph.D. degree from the Department of Computer Science, Hamburg University, Germany, in 1994. Currently, he is a professor at the Institute of Image Processing and Pattern recognition, Shanghai Jiao Tong University, China. He has led many research projects (e. g., National Science Foundation, 863 National High Tech. Plan), had one book published in Germany, and authored more than 200 journal papers. His major research interests are object detection and recognition, data fusion and data mining, and medical image processing.