



Co-saliency detection via inter and intra saliency propagation

Chenjie Ge^a, Keren Fu^a, Fanghui Liu^a, Li Bai^b, Jie Yang^{a,*}

^a Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China

^b University of Nottingham, UK



ARTICLE INFO

Article history:

Received 6 August 2015

Received in revised form

19 March 2016

Accepted 20 March 2016

Available online 24 March 2016

Keywords:

Co-saliency detection

Inter-saliency propagation

Intra-saliency propagation

Fusion

ABSTRACT

The goal of salient object detection from an image is to extract the regions which capture the attention of the human visual system more than other regions of the image. In this paper a novel method is presented for detecting salient objects from a set of images, known as co-saliency detection. We treat co-saliency detection as a two-stage saliency propagation problem. The first inter-saliency propagation stage utilizes the similarity between a pair of images to discover common properties of the images with the help of a single image saliency map. With the pairwise co-salient foreground cue maps obtained, the second intra-saliency propagation stage refines pairwise saliency detection using a graph-based method combining both foreground and background cues. A new fusion strategy is then used to obtain the co-saliency detection results. Finally an integrated multi-scale scheme is employed to obtain pixel-level co-saliency maps. The proposed method makes use of existing saliency detection models for co-saliency detection and is not overly sensitive to the initial saliency model selected. Extensive experiments on three benchmark databases show the superiority of the proposed co-saliency model against the state-of-the-art methods both subjectively and objectively.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Visual saliency of the human or machine visual system serves as a filter for selecting a certain subset of visual information for further processing. The results of saliency detection are saliency maps which enhance salient objects while suppressing background objects. Saliency detection has applications in many fields including objects segmentation [1,2], content based image editing [3–5], image retrieval [6], image compression [7] and video summarization [8–10]. Existing saliency detection models however mainly detect salient objects from a single image [11–14], and the information of similar salient objects in a sequence of images is not exploited.

The concept of co-saliency has been proposed to select common salient objects from a sequence of images [15–21]. Co-saliency detection satisfies the following two properties: (1) co-salient regions should be salient regions in the image and (2) all co-salient regions from different images should share similar characteristics. Because co-saliency maps highlight similar foreground objects, they naturally can be used in many applications such as object co-segmentation [22], co-recognition [23] and common pattern discovery [24].

Most of current co-saliency models [16–19] split the co-salient

object detection problem into single-image saliency detection and multi-image saliency detection, to discover what the salient object is within each image and how frequently the salient object occurs across the images. Various features are used such as texture descriptors [16], corresponding feature [17], and color histograms [18,19] to solve the co-saliency detection problem. Such low-level features are not enough to describe the properties of co-salient objects, leading to unsatisfactory co-saliency detection results. To overcome the difficulties, recently Li et al. [21] directly use single image saliency maps to find co-salient objects based on the fact that co-salient objects in an image sequence should also be salient in each image. In their method, single image saliency maps are exploited to highlight salient objects through stage-one manifold ranking. Thereafter, the foreground of each image is probed to find similar regions in other images through stage-two manifold ranking. The problem with the method is that it is unable to highlight the co-salient object and suppress the background information simultaneously. The co-saliency detection result is easily affected by inaccurate foreground maps obtained by stage-one manifold ranking. If the foreground and background have similar colors, background regions will be highlighted as well, leading to unsatisfactory detection results.

In this paper we propose a novel saliency propagation framework to fulfill the co-saliency detection task. Under this framework co-salient foreground and background cues are separately acquired to enhance the co-salient object. A preliminary conference version of our work has appeared in [25]. Unlike the

* Corresponding author.

E-mail address: jieyang@sjtu.edu.cn (J. Yang).

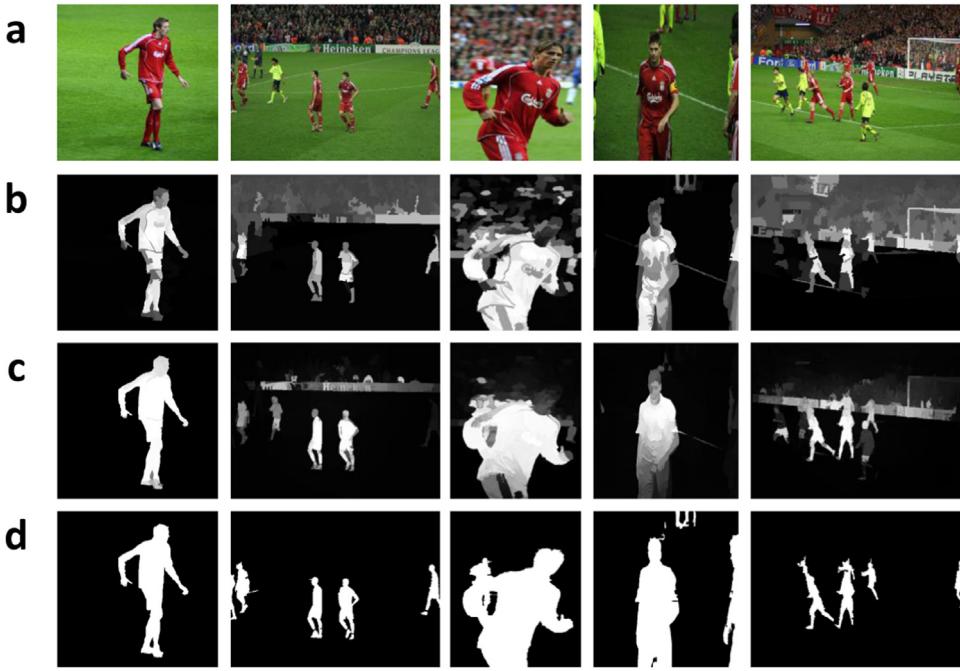


Fig. 1. The performance of [21] and ours. (a) source images; (b) results of [21]; (c) results of our method; (d) ground truth.

method described in [21], the proposed inter-saliency propagation uses superpixel similarity between two images to obtain the foreground cue with the help of a single image saliency map. This is more effective for characterizing objects and preserving edges than pixel-level ranking. Background cue is obtained independently from each image considering both the background connectivity and saliency mask, followed by a graph-based intra-saliency propagation combining both foreground and background cues with a new edge constraint. As shown in Fig. 1, the red soccer players are co-salient objects among all the five images and the proposed model surpasses [21] in both foreground enhancement and background suppression.

The main contributions of this paper are as follows:

- A novel inter-saliency propagation method is proposed to transmit saliency values between two images to find the co-salient foreground cue.
- A new intra-saliency propagation method is presented which simultaneously highlights co-salient objects, suppresses background information and smooths saliency values with edge constraint.
- A new fusion strategy is proposed combining the intra-saliency propagation maps to obtain the final co-saliency map, adaptively weighed by a rough co-saliency approximation.

The paper is organized as follows. We describe the related work in Section 2. The proposed approach is presented in Section 3. Experimental results are shown in Section 4 and finally we conclude in Section 5.

2. Related work

Saliency detection methods can be divided into two major categories: human fixation prediction [26,27] and salient object detection [28–30,11,31–35]. Fixation prediction models usually simulate how human attention focus on an image. This produces a saliency distribution map based on the focus points of the eye.

Salient object detection models focus on detecting salient regions of the whole image. The outcome saliency map reflects the saliency probability of each pixel, which can be regarded as a soft segmentation of salient objects.

Saliency detection from a single image has been an active research area for decades. Itti et al. [36] propose a neural network based saliency model integrating three features and multiple scales. Zhai et al. [37] introduce the image histograms in luminance channel to measure saliency. Pixel-level saliency is computed by the luminance contrast. Hou et al. [28] utilize the spectral residual in the amplitude spectrum of the Fourier transform to compute saliency. Guo et al. [38] propose that the phase spectrum of Fourier transform instead of the amplitude transform is the key in obtaining the location of salient areas. Achanta et al. [39] propose to define pixel saliency using color difference from the image average color, which is equivalent to combining center-surround differences from different scales. Goferman et al. [5] use local and global features to estimate the patch saliency in multi-scales. Visual organizational rules and high-level factors are also considered to enhance the saliency map. Cheng et al. [29] propose histogram and region contrast methods to enhance salient objects. In [30], Perrazzi et al. propose to combine the uniqueness of locally constrained regions and the spatial distribution of regions. Fu et al. [11] present a comprehensive salient object detection system taking advantage of both color contrast and color distribution. Most of the methods mentioned above are simple to implement but cannot always highlight all pixels on the salient objects when the background is complex. By considering the connectivity of background regions, Wei et al. [31] use the shortest path towards the boundary to define the saliency value for each region. Such boundary prior is based on the assumption that boundary parts of an image will most likely belong to the background. Later the graph-based manifold ranking [32] is proposed to compute region saliency according to its relevance to boundary patches. Yan et al. [33] present a hierarchical saliency method which merge regions according to user-defined scales to eliminate small-size distractors. A novel learning based method [34] exploits a random forest regressor to map multiple features to the saliency value of the

region. The fundamental is to learn the weights of discriminative features from each segmented region. In [35], Zhu et al. propose a robust geodesic distance assisted boundary measure, followed by an optimization procedure to highlight the contrast between background and foreground.

Co-saliency models usually start with detecting salient objects in a pair of images. In [40], the local structure changes caused by salient objects are utilized to obtain the co-saliency, and the constraint is that the two images need to have highly similar backgrounds. In [15], the joint information provided by the image pair is used under a pre-attentive scheme, meaning that co-salient object regions are detected before focused attention on individual objects. In [16], co-saliency is formulated as a combination of single-image saliency maps using the three available saliency models, and the co-multilayer graph based multi-image saliency. Co-saliency detection is not limited to image pairs. In [17], cluster-based co-saliency model is proposed by fusing contrast, spatial and corresponding cues. On the basis of hierarchical segmentation, Liu et al. [18] use the regional similarity measure to generate co-salient maps. Object prior and pixel-level refinement [19] which measures color-spatial similarity between pixel and region are also used to achieve better result. Cao et al. [20] generate the co-saliency map through adaptively weighed saliency maps with rank constraint. As the common salient object appears in most images, in [21], the two-stage efficient manifold ranking strategy aims to detect repetitive salient objects. Similar work on co-saliency

detection is co-segmentation which aims to segment the common regions from multiple images [22,41,42]. Co-segmentation needs several object-like proposals obtained in advance, and the segmentation process usually proceeds in supervised or weakly supervised manner. Different from co-saliency detection, the task of co-segmentation is to extract certain common objects no matter they are salient or not. In practice, co-saliency may serve as an pre-processing [22] step for co-segmentation by taking co-saliency maps as object-like proposals.

3. The proposed method

The proposed co-saliency model for salient object detection produces saliency maps containing the co-salient regions. Our novelty lies in proposing a two-stage saliency propagation framework, which is inter-saliency propagation and intra-saliency propagation respectively, to detect the co-salient objects in the image group. New techniques are used in co-salient foreground cue selection, background information exclusion and saliency smoothing. At last a new fusion strategy is proposed to guarantee satisfying results. The framework of the proposed model is illustrated in Fig. 2. It consists of four main steps: pre-processing, inter-saliency propagation, intra-saliency propagation and co-saliency integration. Details of the proposed model are given below.

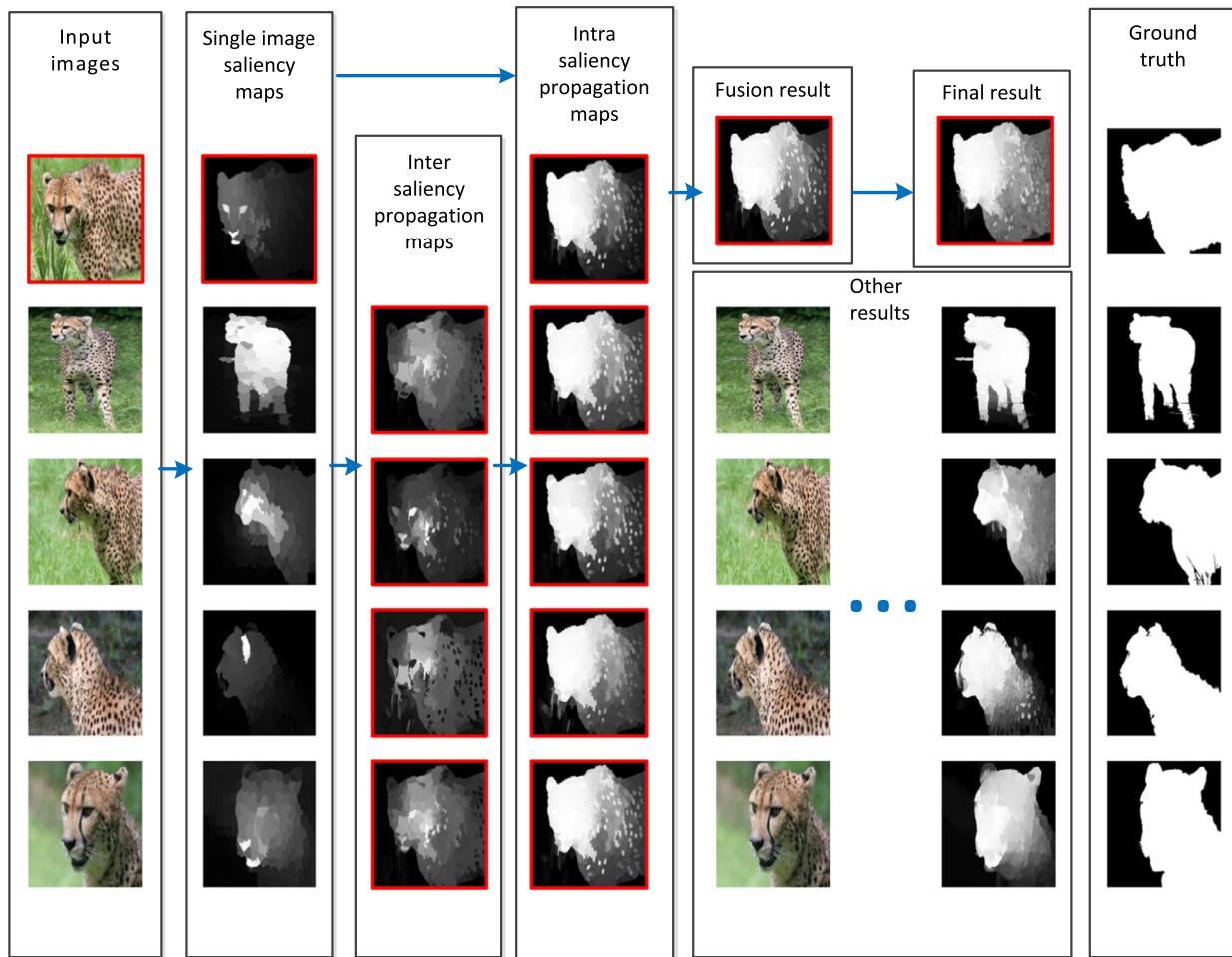


Fig. 2. The framework of the proposed co-saliency model. *Second column:* results after pre-processing. *Third column:* inter-saliency propagation maps of the first image guided by different group members. The intermediates of the first image are shown as an example, in red boxes. Final co-saliency detection results of other group members are also displayed.

3.1. Pre-processing

Each image in a given image set $\{I_m\}_{m=1}^M$ is first segmented into K_m superpixels [43]. Each superpixel of the m_{th} image is labelled as i ($i = 1, \dots, K_m$), its mean color in LAB space is denoted as c_i^m , and its mean 2D position vector is denoted as p_i^m . As the proposed model depends on the single image saliency map, we adopt the initial saliency value $S_o^m(i)$ of i_{th} superpixel from m_{th} image generated by superpixel based method [32]. Other state-of-the-art saliency models can also be applied, as the final co-saliency performance does not heavily rely on the initial model. As shown in Fig. 2, superpixel-based single image saliency maps fail to highlight the whole body of the salient cheetahs because of the boundary prior used in [32]. Therefore, a new co-saliency detection strategy based on two-stage saliency propagation is proposed to highlight the co-salient object of all images.

3.2. Inter-saliency propagation

We propose to obtain pairwise co-salient foreground cue using a novel inter-saliency propagation method, which transmits the saliency value from the guiding image to the guided one according to their color similarity. Each superpixel in the guided image is compared separately with all the superpixels of the guiding image. Taking the m_{th} image as example, if it is to be guided by the n_{th} image, its inter-saliency propagation by the n_{th} ($n = 1, \dots, M, n \neq m$) image is formulated as follows:

$$S_{raw}^{n \rightarrow m}(i) = g_m(i) * \frac{\sum_{j=1}^{K_n} \exp(-\alpha \|c_i^m - c_j^n\|_2) S_o^n(j)}{\sum_{j=1}^{K_n} \exp(-\alpha \|c_i^m - c_j^n\|_2)} \quad (1)$$

$$g_m(i) = \exp(-\|p_i^m - p_c^m\|_2^2 / \sigma_1^2) \quad (2)$$

where $\exp(-\alpha \|c_i^m - c_j^n\|_2)$ denotes the color similarity between superpixel i of m_{th} image and superpixel j of n_{th} image, and α is empirically chosen as 10. $g_m(i)$ is a center bias term highlighting superpixels locating near the center of an image, and p_c^m represented the center position of the m_{th} image. The standard deviation σ_1 is set to one-third of the longest dimension of the input image. Even if there is more than one salient object in the image, the center bias will work because salient objects are less likely to locate at all the four image corners. The rationale is that photographers tend to locate the object at an image center and human fixation has much higher probability to fall onto the center area of an image. By searching all the superpixels in the n_{th} image, the m_{th} image's raw saliency guided by the n_{th} image is decided by color similarity weighed by initial saliency values. This simple but effective procedure propagates the initial saliency values from the guiding image to the guided image, enabling similar colored superpixels of different images to have the similar saliency values. If the m_{th} image has the same salient object as the n_{th} image, the common salient object will be highlighted in the raw inter-saliency propagation map $S_{raw}^{n \rightarrow m}$.

Geodesic distance is a more continuous measure [44,45] to define how well any two superpixels are spatially connected compared to traditional Euclidean distance. As the above color similarity based propagation may be sensitive to color fluctuation even within the homogeneous region, a geodesic distance based smoothing strategy is adopted to refine the raw intra-saliency propagation maps. We connect all adjacent superpixels to create an undirected graph and assign the edge value as the distance between adjacent superpixels in LAB space. The geodesic weight $w_{geo}^m(i, j)$ is then defined as [44,45] to give two superpixels a large weight if their geodesic distance is small. We present the

smoothed inter-saliency propagation map as:

$$S_{inter}^{n \rightarrow m}(i) = \frac{\sum_{j=1}^{K_m} w_{geo}^m(i, j) S_{raw}^{n \rightarrow m}(j)}{\sum_{j=1}^{K_m} w_{geo}^m(i, j)} \quad (3)$$

Each image in the image group take turns to guide other images to generate the inter-saliency propagation maps. For example, if the m_{th} image is selected as the guided one, it will get $M - 1$ inter-saliency propagation maps guided by other group members denoted as $S_{inter}^{n \rightarrow m}$ ($n = 1, \dots, M, n \neq m$). The inter-saliency propagation is like the retrieving process, once the guided image contains the common salient object, most of its inter-saliency propagation maps will highlight the same objects because most other guiding images contain the same salient object. As shown in Fig. 2, the cheetahs are the co-salient objects according to the five images, but single image saliency model fails to highlight the whole body of the cheetah. After inter-saliency propagation between image pairs, the co-salient cheetah in the first image has roughly emerged, indicating that the proposed inter-saliency propagation effectively transmit saliency values between the co-salient objects.

3.3. Intra-saliency propagation

Fig. 2 shows that inter-saliency propagation maps are not quite satisfactory. The reasons are that inter-saliency propagation is able to highlight the common salient object but lacks the ability of fully suppressing background. In the inter-saliency propagation section, we mainly focused on what the co-salient object is. Since each image contains not only the co-salient object but also its own background information, we propose to obtain the background cue according to boundary connectivity prior and saliency mask. Boundary connectivity prior [35] measures how heavily a superpixel is cropped on the image boundary. The prior $P_{con}^m(i)$ is close to 1 if the superpixel i is deemed as background and is close to 0 if deemed as belonging to an object:

$$P_{con}^m(i) = 1 - \exp\left(-\frac{Ratio_m^2(i)}{2\sigma_2^2}\right) \quad (4)$$

$$Ratio_m(i) = \frac{\sum_{j=1}^{K_m} w_{geo}^m(i, j) \delta(j \in \text{boundary})}{\sqrt{\sum_{j=1}^{K_m} w_{geo}^m(i, j)}} \quad (5)$$

where $\sigma_2 = 1$ is set as [35] and $\delta(\cdot) = 1$ if the superpixel j is on the image boundary and 0 otherwise. For the initial saliency map of the m_{th} image S_o^m , an OTSU [46] threshold th_m is employed to distinguish the foreground and background, and the saliency mask is obtained as follows:

$$Mask_m(i) = \begin{cases} 1 & S_o^m(i) \leq th_m \\ 0 & S_o^m(i) > th_m \end{cases} \quad (6)$$

The background cue for the m_{th} image is then constructed by combining the boundary connectivity prior and saliency mask:

$$w_{bg}^m = P_{con}^m * Mask_m \quad (7)$$

We improved the boundary connectivity prior by filtering it with a mask derived from the initial saliency map. Note that the boundary connectivity prior may wrongly emphasize the foreground with a large value sometimes if the foreground region is cropped on the image boundary. The saliency mask lets the salient region remain foreground in the background cue (by assigning it as 0), which makes the background cue better characterize the backgroundness.

Unlike traditional co-saliency models [17–21], the proposed intra-saliency propagation first takes both co-salient foreground

cue and background cue of each image into consideration to highlight co-salient object, suppress background information and smooth saliency values simultaneously. Propagated saliency values are obtained by minimizing a graph-based regularization energy function [35], which is formulated as follows:

$$E(s) = \sum_{i=1}^{K_m} w_i^{bg} s_i^2 + \sum_{i=1}^{K_m} w_i^{fg} (s_i - 1)^2 + \sum_{ij} w_{ij} (s_i - s_j)^2 \quad (8)$$

The three terms of the energy function are for the background cue, co-salient foreground cue and smoothness constraints respectively. The background term encourages superpixel i with large background cue w_i^{bg} ($w_i^{bg} = w_{bg}^m(i)$) to take a small value (close to 0). The inter-saliency propagation map is taken as the co-salient foreground cue ($w_i^{fg} = S_{inter}^{n \rightarrow m}(i)$), and the foreground term encourages superpixel i with large co-salient foreground cue w_i^{fg} to take a large value (close to 1). The smoothness term encourages continuous saliency values. s_i and s_j are the saliency values of superpixel i and j .

We propose a new edge cue to compute the above smoothness weight w_{ij} . First an undirected graph is constructed not only linking the adjacent superpixels, but also linking neighbor's neighbor. Besides, arbitrary boundary superpixels are connected with each other. Such kind of close-loop graph in Fig. 3(a) can significantly smooth saliency values because the distance of similar superpixels is reduced. Traditionally the graph edge w_{ij} is defined based on only color similarities [32]. We argue that only the color information is unlikely to distinguish the object from the background if they have similar color but distinguishable textures. Thus a new edge cue is proposed, which is complementary to the color cue in such challenging cases. The newly defined graph smoothness weight combining both color and edge cues is presented as follows:

$$w_{ij} = \exp(-\lambda_c \|c_i^m - c_j^m\|_2) * \exp(-\lambda_e |\max_{i' \in \bar{ij}} f_{i'}|) \quad (9)$$

where \bar{ij} is the straight line connecting the centers of the superpixel i and j on the image plane and i' is a pixel on \bar{ij} . We utilize an edge detector [47,48] to get an edge map as illustrated in Fig. 3(b). The edge map displays high contrast in the object contour to the surrounding area, which is helpful to preserve the object edge. $f_{i'}$ is the edge magnitude of i' derived from the edge map. The edge cue indicates that if strong intervening boundaries or edges exist between two superpixels, their edge cue will be small, in other words they have less possibility to belong to the same object. λ_c

and λ_e control the similarity of color cue and contribution of the edge cue respectively, and they are both empirically set to 10. By simply multiplying the color cue with the edge cue, saliency value will be smoothed only between superpixels sharing similar color and texture. Eq. (8) can be rewritten in the matrix form as follows to minimize the energy cost:

$$x^T W_b x + (x - \mathbf{1})^T W_f (x - \mathbf{1}) + x^T (D - W) x \quad (10)$$

where W_f and W_b are the diagonal matrices with diagonal entries be w_i^{fg} and w_i^{bg} , indicating the co-salient foreground cue and background cue respectively, and $\mathbf{1}$ denotes the column vector consisting of 1, the number of which is the same as the superpixels of the m_{th} image. W is the smoothing matrix composed of w_{ij} from (9) and D is the degree matrix of W . x is the column vector with elements of the intra-saliency propagation values to be computed, which is equal to $S_{intra}^{n \rightarrow m}$. The three terms of (10) are for the purpose of highlighting co-salient object, suppressing background and smoothing. By differentiating x the closed-form solution is achieved as follows:

$$S_{intra}^{n \rightarrow m} = (W_b + W_f + D - W)^{-1} W_f \cdot \mathbf{1} \quad (11)$$

As illustrated in Fig. 3, background cue map and co-salient foreground cue map are complementary to each other. The background cue in Fig. 3(c) is effective in describing the backgroundness though some parts of the cheetah body is wrongly selected as background because these parts are overly connected to the image boundary. The co-salient foreground cue in Fig. 3(c) is good at separating the foreground from background but the whole cheetah is not highlighted uniformly. After the intra-saliency propagation, the cheetah has been highlighted, showing the merits of combining both the foreground and background cues to pop out co-salient object.

As for the guided image itself (the m_{th} image), we directly take its original saliency value as the foreground cue, and the background cue is computed as above. Its intra-saliency propagation result is depicted as $S_{intra}^{n \rightarrow m}$. Then we get all five intra-saliency propagation maps displayed in Fig. 2. Slight differences exist among them because they are guided by different group members. Intra-saliency propagation maps show the refined results of pairwise guidance, which means that if the m_{th} image contains the co-salient object, it will be highlighted in most intra-saliency propagation maps ($S_{intra}^{n \rightarrow m}$ and $S_{intra}^{n \rightarrow m}$, $n = 1, \dots, M, n \neq m$).

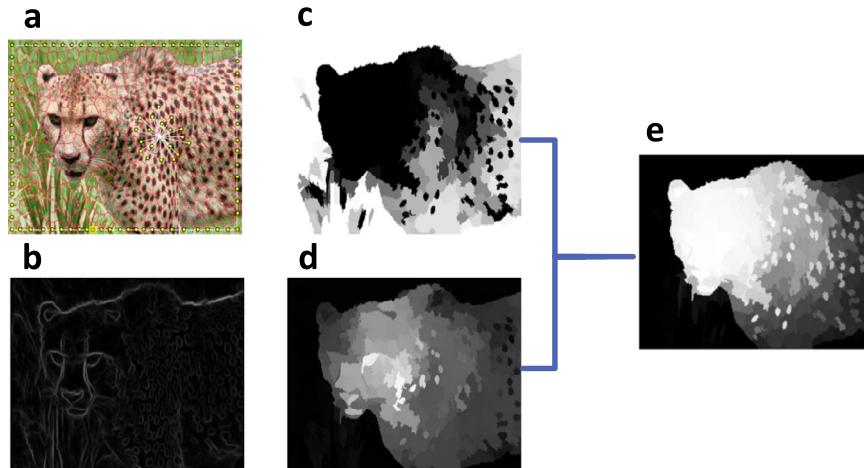


Fig. 3. Graph construction and intra-propagation components. (a) Superpixel segmentation and graph segmentation; (b) edge map; (c) background cue map; (d) co-salient foreground cue map; (e) intra-saliency propagation map.

3.4. Co-saliency integration

In this paper our algorithm is based on the principle that an object is deemed co-salient only when it is salient in most intra-saliency propagation maps. We will introduce a new fusion strategy to combine all these intra-saliency propagation maps. The simplest way of fusion is to average all these intra-saliency propagation maps, but it may bring in the negative impact of some inaccurate intra-saliency propagation map highlighting some unco-salient parts, such as some region salient but not co-salient in the image set, or other meaningless background clutter. To mitigate such limitation and consider the intrinsic relationship of these components as suggested in [20] instead of using the fixed combination weights, we propose the following novel fusion strategy.

It is our belief that the average map of all intra-saliency propagation maps is a rough approximation of co-saliency. For the superpixel i in the m_{th} image, if it appears to be salient in most of the intra-saliency propagation maps (S_{intra}^m and $S_{intra}^{n \rightarrow m}$, $n = 1, \dots, M, n \neq m$), it will be regarded as co-salient and its co-saliency is obtained by the fusion strategy. We sum up the saliency values of superpixel i from all the intra-saliency propagation maps weighed by the deviation from the average saliency. The fusion favors the guided intra-saliency propagation results close to the average saliency and penalizes those inaccurate ones far away from the average saliency, and it is formulated as follows:

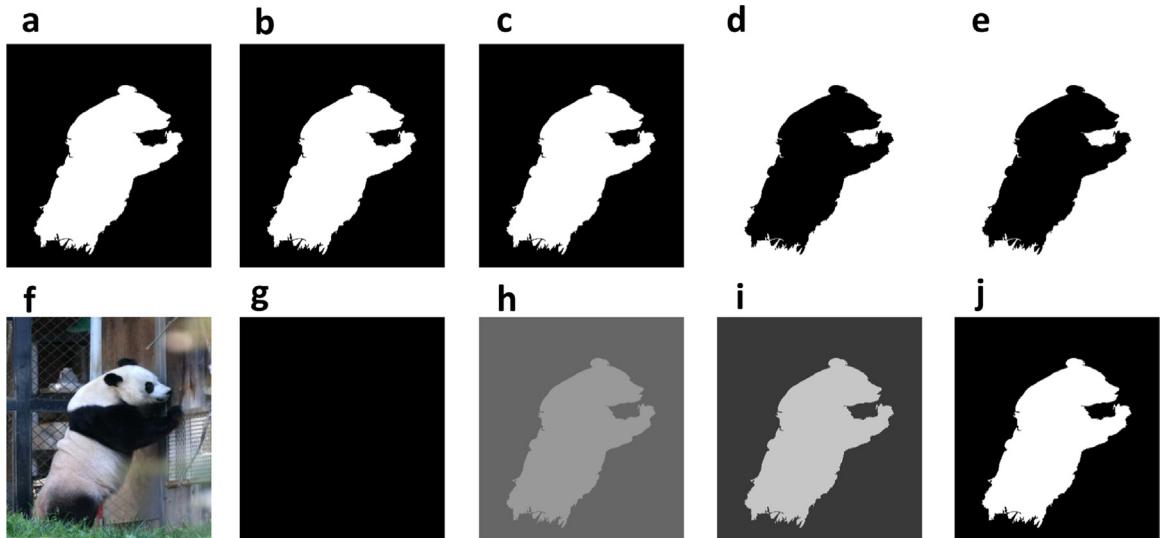


Fig. 4. An example showing the benefits of our fusion strategy. (a–c) Accurate intra-saliency propagation components; (d and e) inaccurate intra-saliency propagation components; (f) original image; (g) fusion result by multiplying all components; (h) fusion result by averaging all components; (i) our fusion result; and (j) ground truth.

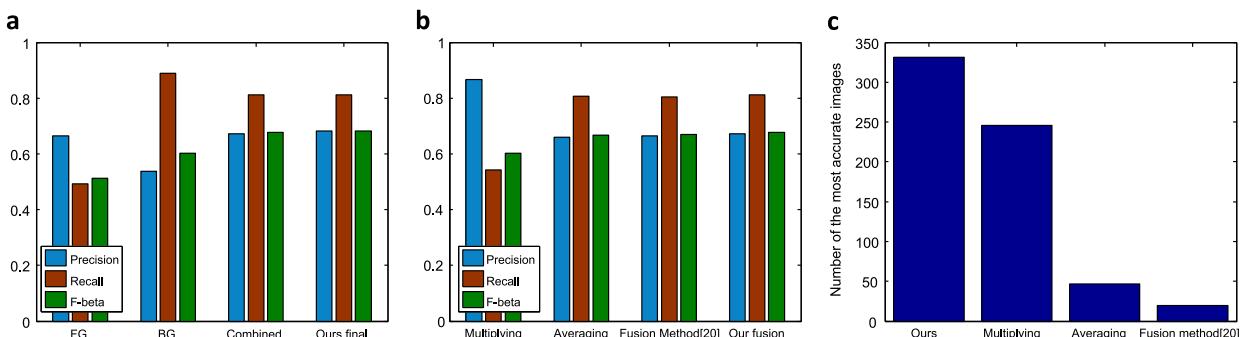


Fig. 5. Comparison in each step of the proposed model on icoseg database. (a) F_β^W comparisons of our propagation based foreground, background, combined results and our final results; (b) F_β^W comparisons of various fusion methods including multiplying, averaging, fusion method [20] and our fusion strategy; (c) number of the most accurate images among various fusion methods including multiplying, averaging, fusion method [20] and our fusion strategy.

$$S_{fusion}^m(i) = \frac{\sum_{n=1, n \neq m}^M w_n S_{intra}^{n \rightarrow m}(i) + w_m S_{intra}^m(i)}{\sum_{n=1}^M w_n} \quad (12)$$

$$w_n = \begin{cases} \exp\left(-\frac{(S_{intra}^{n \rightarrow m}(i) - S_{mean}(i))^2}{\sigma_m^2}\right) & n \neq m \\ \exp\left(-\frac{(S_{intra}^m(i) - S_{mean}(i))^2}{\sigma_m^2}\right) & n = m \end{cases} \quad (13)$$

$$S_{mean}(i) = \frac{1}{M} \left(\sum_{n=1, n \neq m}^M S_{intra}^{n \rightarrow m}(i) + S_{intra}^m(i) \right) \quad (14)$$

where σ_m is set to the standard deviation of $S_{intra}^m(i)$ and $S_{intra}^{n \rightarrow m}(i)$ ($n = 1, \dots, M, n \neq m$). As can be seen from the fusion result Fig. 1, the whole cheetah stands out after the fusion step. A more extreme example showing the benefits of our fusion is in Fig. 4. Fig. 4(a–e) are the components remaining to be fused, (a–c) are the absolutely accurate components which are also ground truth, and (d–e) are totally wrong components which are the inverse of ground truth. As shown in Fig. 4(g), multiplying all the components may lose some foreground parts if some components fail to highlight the salient object, and averaging all these components is

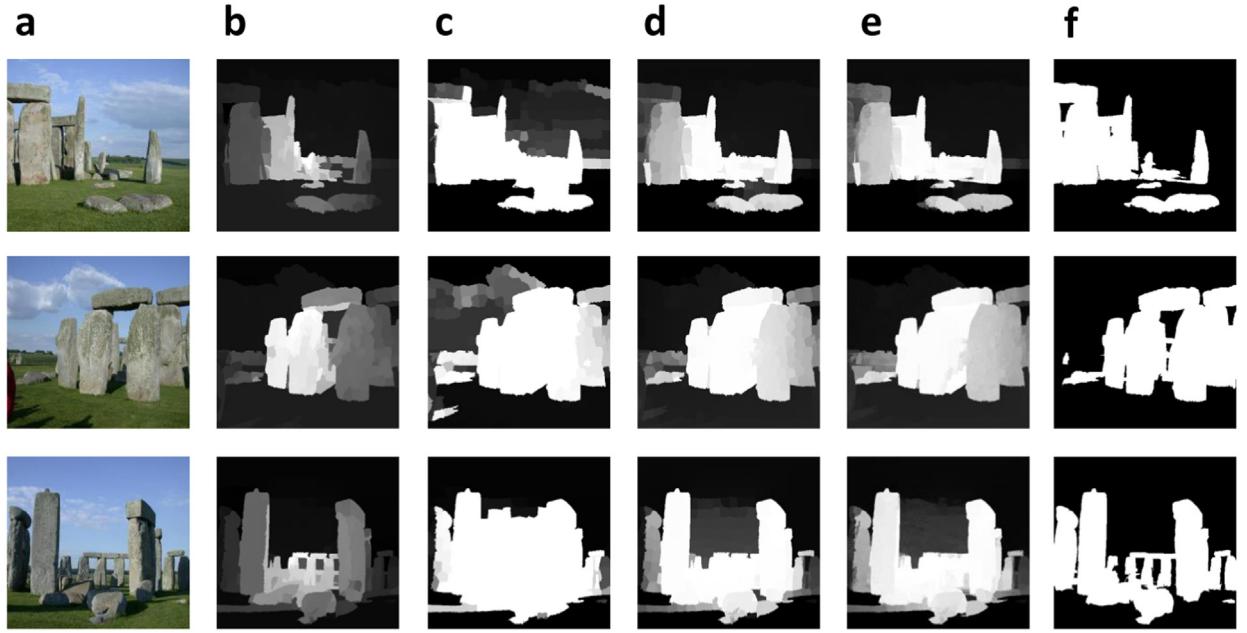


Fig. 6. Visual exhibition of our propagation based foreground, background, combined results and our final results. (a) Original images; (b) foreground-induced maps; (c) background-induced maps; (d) fusion maps, (e) our final maps, and (f) ground truth.

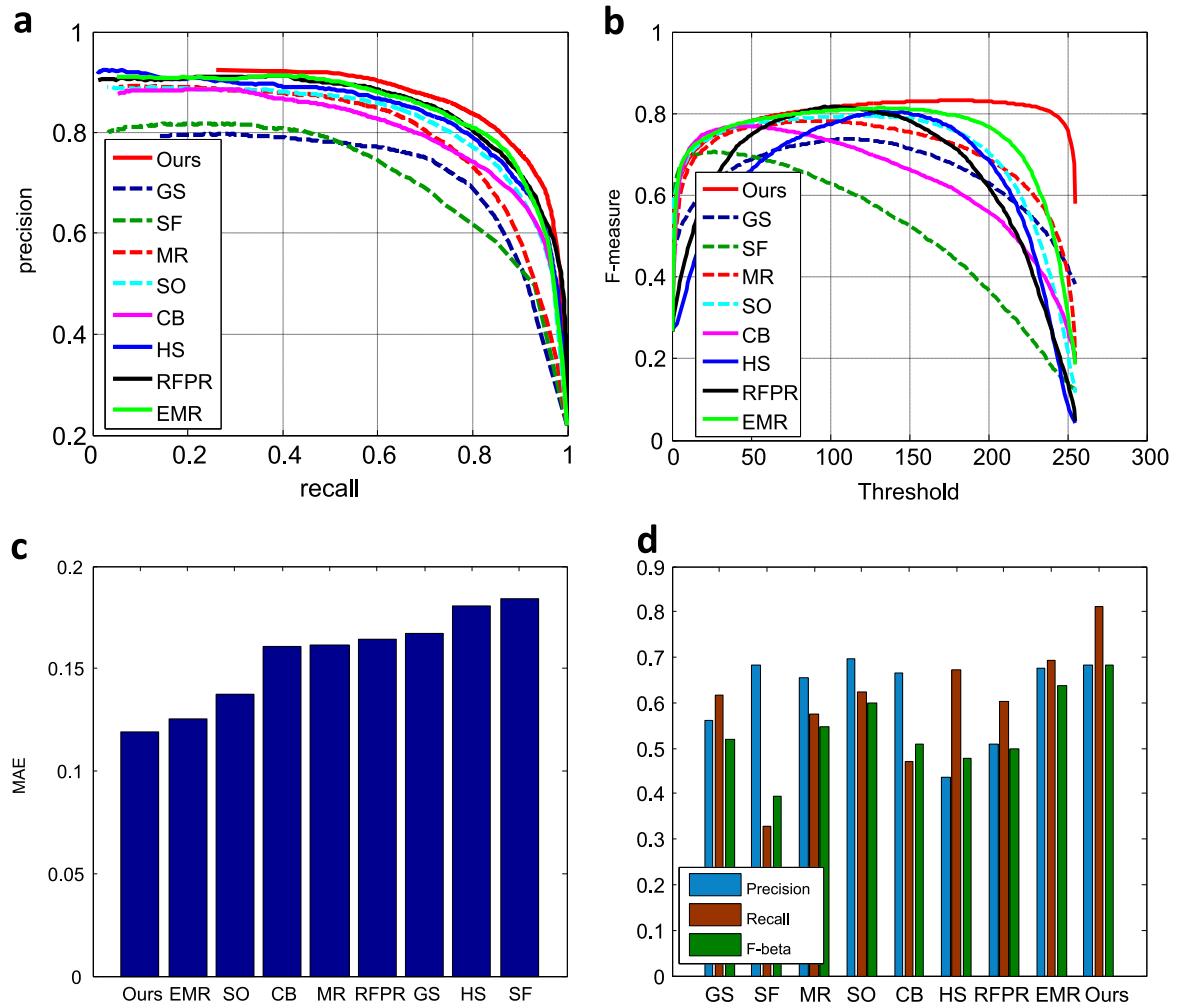


Fig. 7. Objective comparison of the proposed model with eight state-of-the-art saliency detection models on icoseg database. (a) PR curve comparison; (b) F -measure curve comparison; (c) MAE comparison; (d) weighed precision, recall and F -beta comparison.

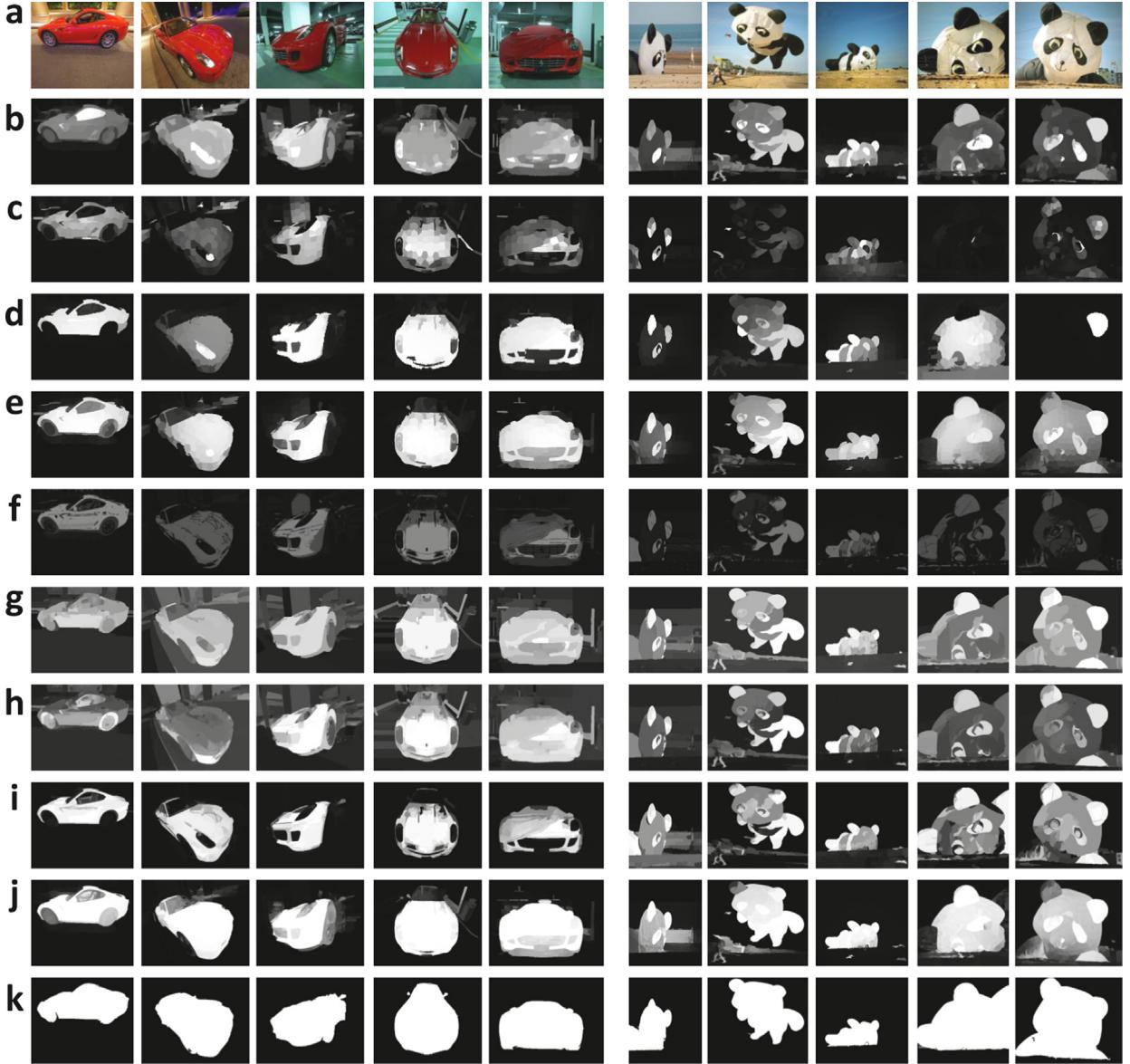


Fig. 8. Subjective comparisons of the proposed model with eight state-of-the-art saliency detection models on icoseg database. (a) Original images; (b) GS [31]; (c) SF [30]; (d) MR [32]; (e) SO [35]; (f) CB [17]; (g) HS [18]; (h)RFPR [19]; (i) EMR [21]; (j) Ours; and (k) Ground truth.

less good at suppressing the effect of inaccurate components. Our fusion strategy reaches the closest result to ground truth in both foreground rendering and background exclusion.

Finally the multi-scale scheme [49] is used to refine the superpixel saliency values to pixel-level and handle the scale problem. Superpixels are generated at N_l different scales initially for subsequent processing, and in this paper we choose the 200, 300, 400 superpixel segments. After getting the final superpixel-level co-saliency results at different scales, we use the weighted summation to obtain the pixel-level saliency value as follows. The weights are determined by how similar a pixel is to the superpixel containing it in LAB color space:

$$S_{final}(z) = \frac{\sum_{l=1}^{N_l} w_{zn^{(l)}} S_{n^{(l)}}}{\sum_{l=1}^{N_l} w_{zn^{(l)}}} \quad (15)$$

$$w_{zn^{(l)}} = \frac{1}{\|f_z - c_{n^{(l)}}\|_2} \quad (16)$$

where $S_{n^{(l)}}$ is the co-saliency result at scale l , f_z is the LAB feature of pixel z , $n^{(l)}$ denotes the label of the superpixel containing pixel z at scale l and $c_{n^{(l)}}$ is its corresponding LAB feature. The weight $w_{zn^{(l)}}$ denotes the similarity between pixel z and its corresponding superpixel $n^{(l)}$. The final cheetah shown in Fig. 1 is more uniformly highlighted than the previous superpixel level, and the final co-saliency detection results of other images from the same image set are also listed in Fig. 1.

4. Experimental results

We test our co-saliency model on three benchmark databases, i.e., the CP database [16] which contains 105 image pairs, the MSRC database [50,51] which contains 14 object classes with

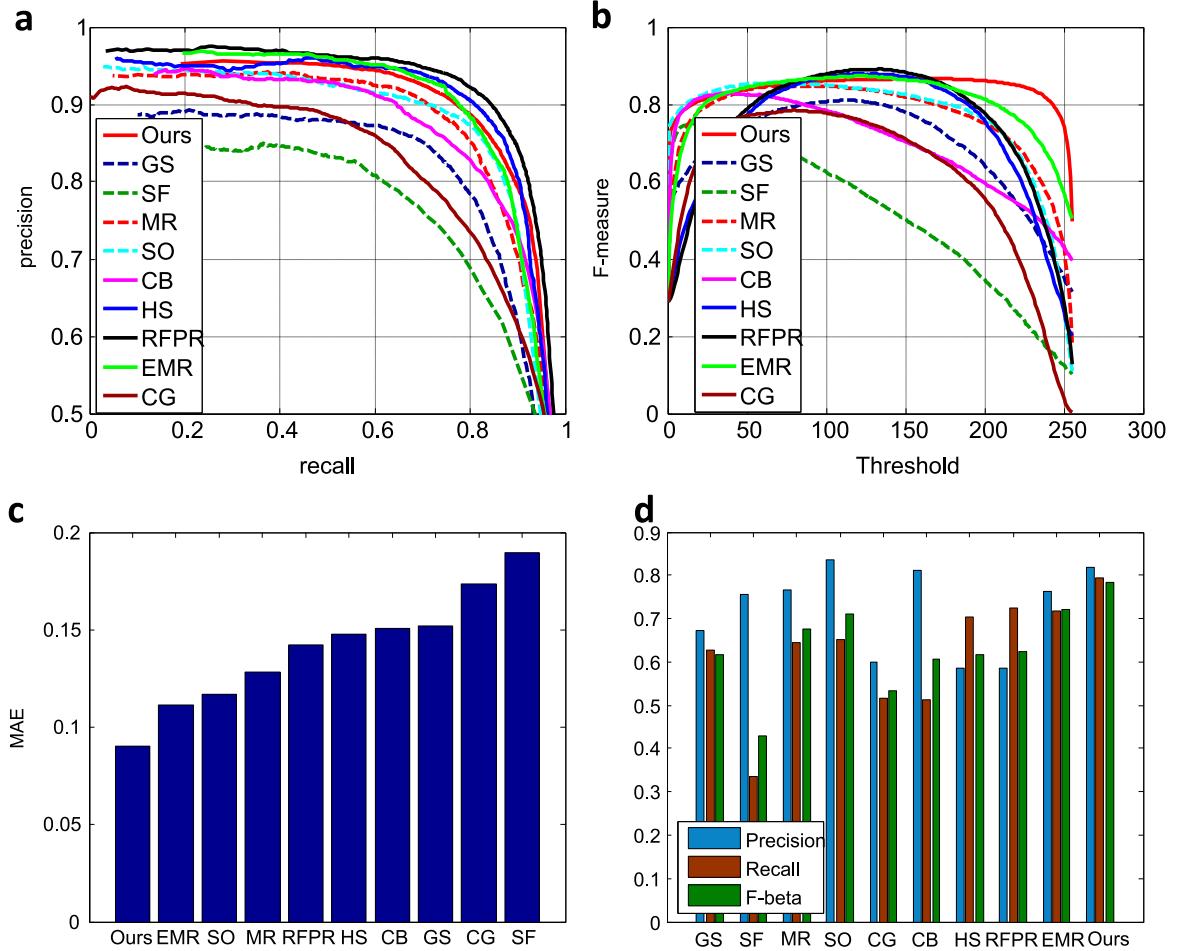


Fig. 9. Objective comparisons of the proposed model with nine state-of-the-art saliency detection models on CP database. (a) PR curve comparison; (b) F-measure curve comparison; (c) MAE comparison; (d) weighed precision, recall and F-beta comparison.

about 30 images per class, and the iCoseg database [52] which contains 643 images from 38 object classes. For quantitative comparison, we employ four evaluation metrics: (a) Precision-Recall (PR) curve [29,30], (b) *F*-measure curve [21], (c) Mean absolute error (MAE) [30], (d) Weighed *F*-measure (F_β^w) [53]. Given a threshold T , the *precision* and *recall* rates of a certain saliency detection model are defined as

$$\text{Precision}(T) = \frac{1}{N} \sum_{i=1}^N \frac{|SM_i(T) \cap G_i|}{|SM_i(T)|} \quad (17)$$

$$\text{Recall}(T) = \frac{1}{N} \sum_{i=1}^N \frac{|SM_i(T) \cap G_i|}{|G_i|} \quad (18)$$

where $SM_i(T)$ is a binary saliency map on threshold T , G_i denotes ground truth, $| \cdot |$ is the binary mask's area and N is the number of images in a database. The PR curve is obtained by plotting the *precision* and *recall* rates when T varies from the range $[0, 255]$. Similarly *F*-measure is defined as

$$F\text{-measure}(T) = \frac{(1 + \beta^2) * \text{Precision}(T) * \text{Recall}(T)}{\beta^2 * \text{Precision}(T) + \text{Recall}(T)} \quad (19)$$

where β^2 is set as 0.3 [29,30] to emphasize the precision, and *F*-measure curve is obtained by plotting $F\text{-measure}(T)$ and T within the range $[0, 255]$. Mean absolute error measures the difference between the continuous saliency map S and ground truth G :

$$MAE = \frac{1}{W * H} \sum_{x=1}^W \sum_{y=1}^H |S(x, y) - G(x, y)| \quad (20)$$

where W and H denote the width and the height of the saliency or ground truth map. The MAE in the whole database is the average of MAE of each image.

Recently Margolin et al. [53] reveal that the interpolation flaw, dependency flaw and equal-importance flaw exist in the traditional PR curve and *F*-measure. Instead, they propose the weighed precision Precision^w , weighed recall Recall^w and weighed *F*-measure F_β^w to better evaluate the performance of saliency map. Note that $\beta^2 = 1$ is set without bias between precision and recall. We also adopt this metric in our experiment and for more details please see [53].

4.1. The role of each step of the proposed model

The foreground-induced maps are obtained by the inter-saliency propagation step and the fusion strategy without the intra-saliency propagation step considering the background cue. We also generate the background-induced maps using only the inverse of the background cue in the intra-saliency propagation step. The fusion maps are the results of our method including inter-saliency propagation, intra-saliency propagation and co-saliency integration step without multi-scale scheme. Finally, a multi-scale strategy is utilized to produce final results.

Fig. 5(a) shows that combining background and foreground cues has achieved a larger F_β^w than using background or

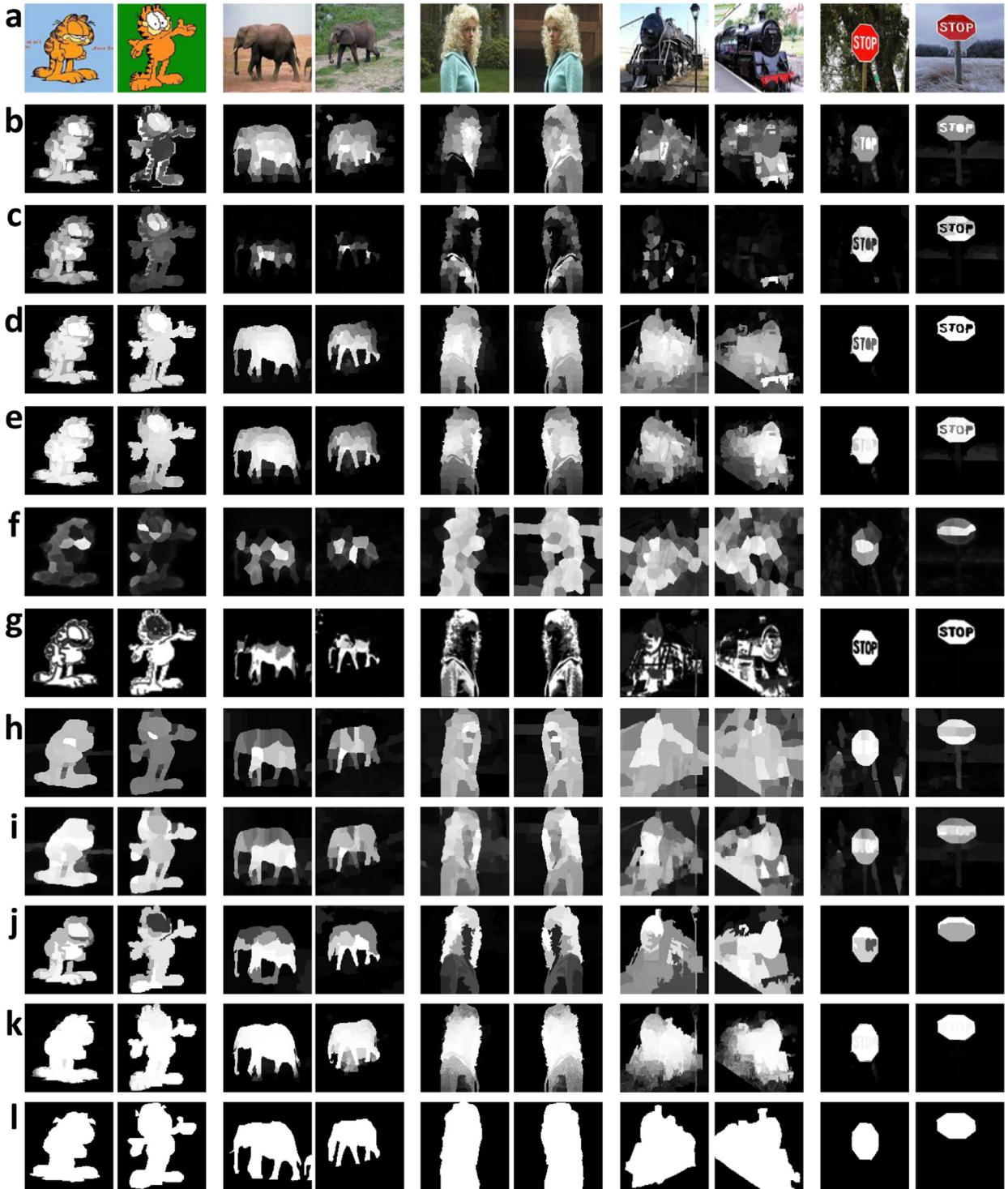


Fig. 10. Subjective comparisons of the proposed model with nine state-of-the-art saliency detection models on CP database. (a) Original images; (b) GS [31]; (c) SF [30]; (d) MR [32]; (e) SO [35]; (f) CG [16]; (g) CB [17]; (h) HS [18]; (i) RFPR [19]; (j) EMR [21]; (k) Ours; and (l) ground truth.

foreground alone. The multi-scale refinement slightly improves F_{β}^W compared to the fusion maps. Fig. 6(b) shows that the proposed inter-saliency propagation step is able to accurately highlight the stones although salient parts are not uniformly bright, showing the weakness in dealing with color fluctuations among co-salient objects. Fig. 6(c) shows that the background-induced map succeeds in excluding most background parts but some of them still have high saliency values. Fig. 6(d) and (e) demonstrate that the proposed model gives the best performance because we treat the

foreground and background cues separately. The intra-saliency propagation with edge constraint and novel fusion strategy effectively combine the previous two cues and generate accurate co-saliency maps.

Finally, we compare the proposed fusion strategy with other fusion strategies quantitatively in Fig. 5(b) and (c). Besides the multiplying and averaging scheme, fusion results are also generated by low rank fusion framework [20], where the matrix jointing the feature histograms of intra-saliency propagation results appears low rank and self-adaptive weights are used for

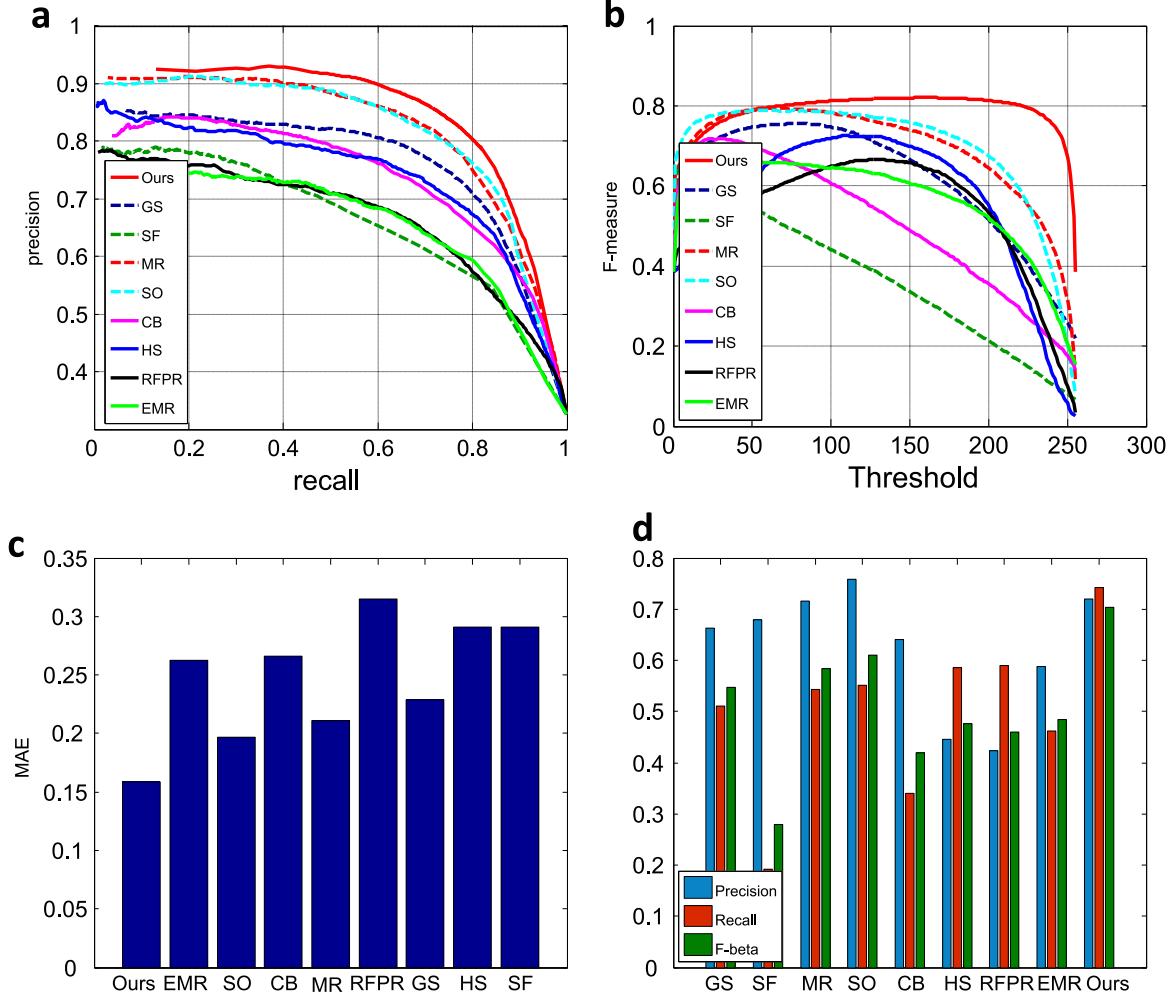


Fig. 11. Objective comparison of the proposed model with eight state-of-the-art saliency detection models on MSRC database. (a) PR curve comparison; (b) F-measure curve comparison; (c) MAE comparison; (d) weighed precision, recall and F-beta comparison.

combination. In Fig. 5(b), F_β^w of each method is 0.602, 0.667, 0.669, 0.677 respectively, and our fusion method outperforms three other fusion methods on this metric. What's more, we count the number of the most accurate images among four fusion methods. MAE of each image based on four fusion methods is calculated and the most accurate image means the one with the lowest MAE. As shown in Fig. 5(c), our fusion method achieves the maximum number of accurate results.

4.2. Evaluation on the iCoseg database

We generate the co-saliency maps using the proposed model on the iCoseg database, and all the 643 saliency maps are obtained. The proposed model is compared, both subjectively and objectively, with eight state-of-the-art saliency detection models including four single image saliency models GS [31], SF [30], MR [32], SO [35] and four co-saliency models CB [17], HS [18], RFPR [19], and EMR [21]. The original EMR is tested on a subset of iCoseg database from their paper. We use the EMR detection results of all 643 images in the iCoseg database in our experiment.

In Fig. 7(a), we compare the proposed model with other eight models using the PR curve metric. The proposed model obtained a higher accuracy than all the others models. Fig. 7(b) shows that the F-measure of the proposed model is higher than any other model. As shown in Fig. 7(c), the MAE value of the proposed model reaches the minimum among all the models compared, which

indicates that the proposed model is the closest to ground truth. Fig. 7(d) shows that the F_β^w of the proposed model is the best, followed by a relatively high $Precision^w$ and the highest $Recall^w$.

We also compare the proposed model visually with other eight models on some test images. The proposed model gives the best performance showing brighter foreground and darker background than others. For the model GS, each superpixel's saliency value is based on its geodesic distance to the image boundary. As a result, GS may be sensitive to object texture and background clutter, as shown in the 2nd and 7th examples of row b. The color contrast and distribution used in SF only highlight some part of the co-saliency object visually in Fig. 8. The two graph-based method MR and SO suppress background effectively because of the boundary prior but are still weak in highlighting the whole foreground object. So the four single image saliency models do not perform well on the database because they do not utilize the intrinsic relationship of the images. The co-saliency model CB generates the final co-saliency map by multiplying the contrast, spatial and correspondence cues. It achieves the high accuracy at the cost of incomplete salient object rendering and relatively low saliency values. The HS and REPR models use object priors but some background areas are still highlighted especially when the background is complex (see row g and h in Fig. 8). As shown in row i of Fig. 8, the EMR results lose some parts of the red car and the pandas are not highlighted uniformly because color fluctuates within these foreground objects. Among all these models, our co-

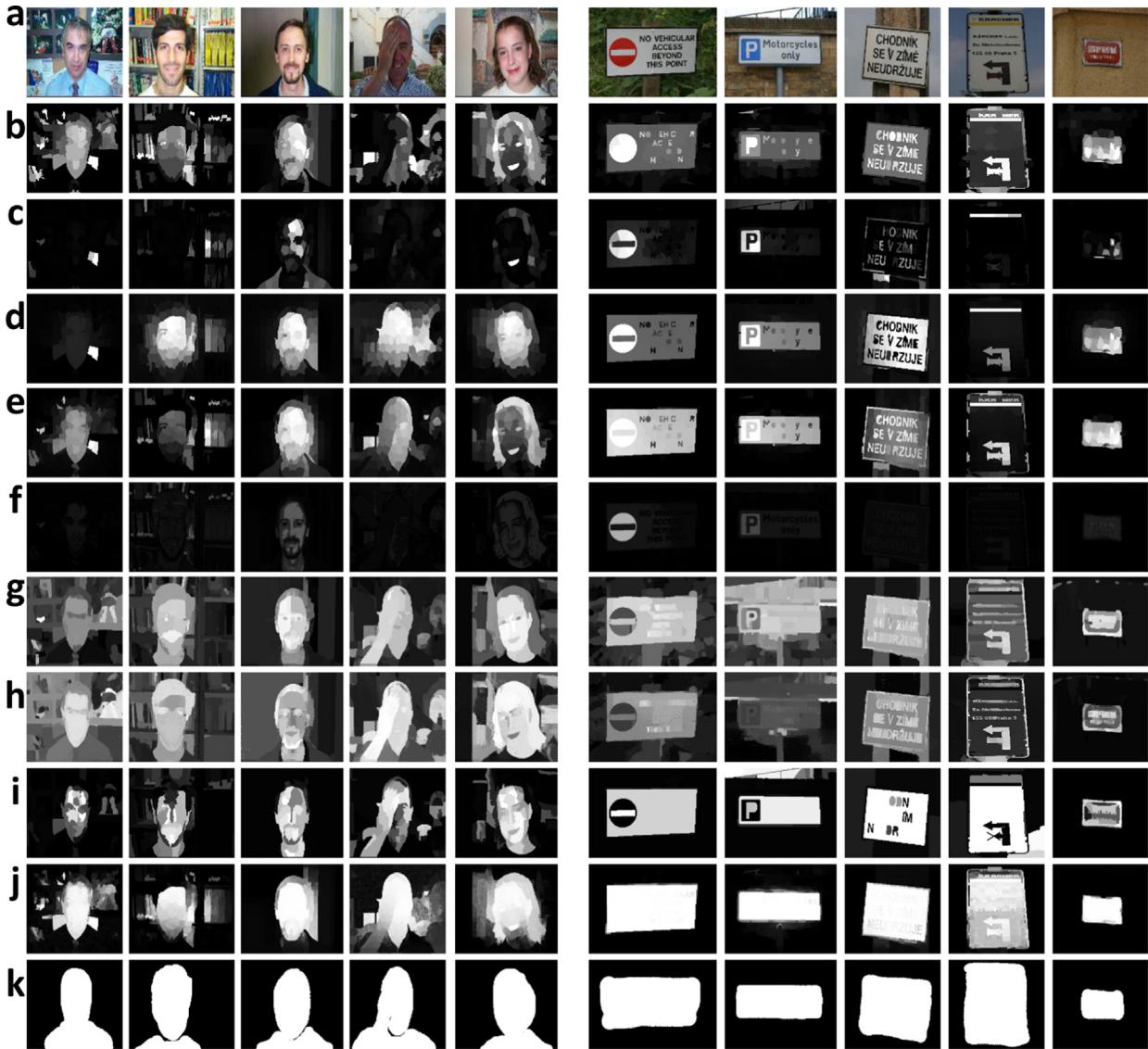


Fig. 12. Subjective comparisons of the proposed model with eight state-of-the-art saliency detection models on MSRC database. (a) Original images; (b) GS [31]; (c) SF [30]; (d) MR [32]; (e) SO [35]; (f) CB [17]; (g) HS [18]; (h) RFPR [19]; (i) EMR [21]; (j) Ours; and (k) Ground truth.

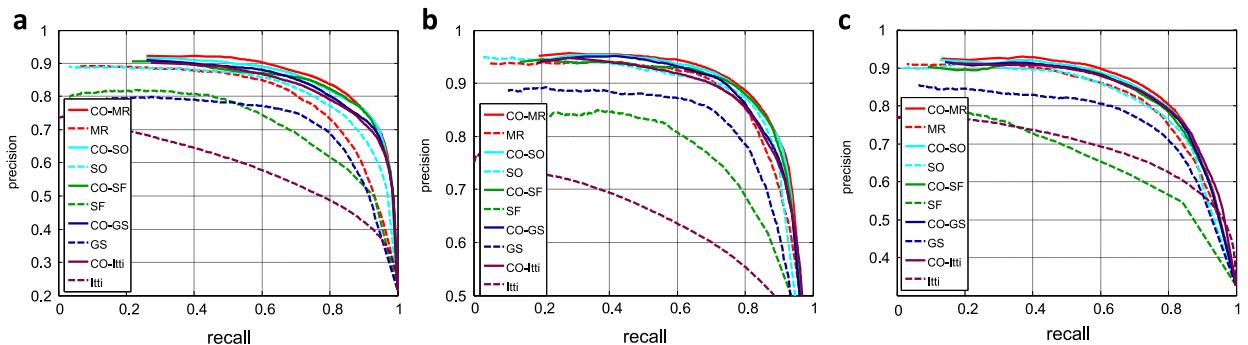


Fig. 13. PR curves of the existing saliency models and their integration with the proposed co-saliency model on three databases. (a) On iCoseg database; (b) On CP database; (c) On MSRC database.

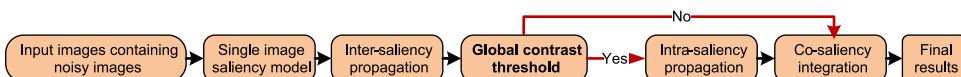


Fig. 14. The block diagram of the modified model in order to deal with noisy images.

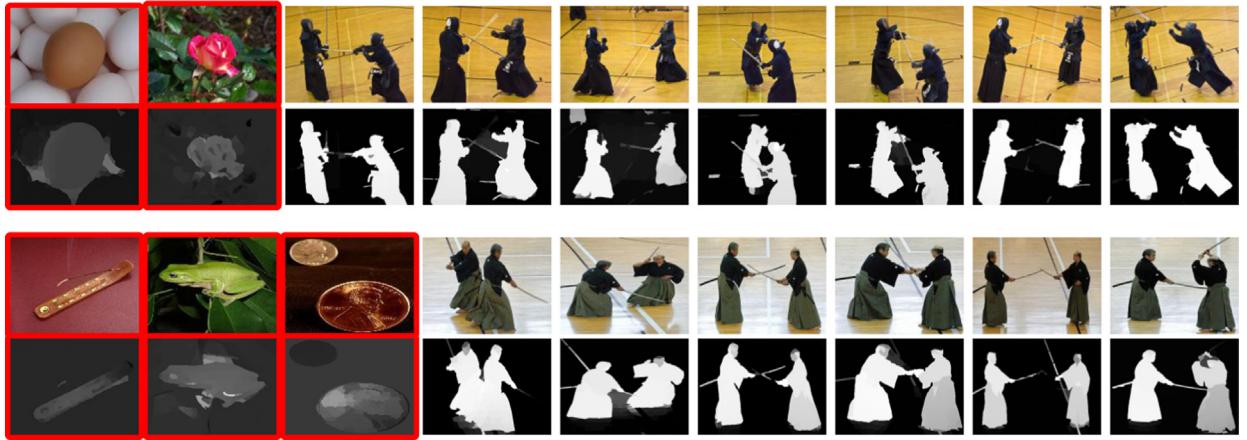


Fig. 15. Examples on the image group containing noisy images.

Table 1
Average running time per image (seconds).

Models	CG [16]	CB [17]	HS [18]	RFPR [19]	EMR [21]	Ours
CP	472.40	1.74	15.73	16.48	0.63	1.28
iCoseg	N/A	4.18	103.38	109.30	4.92	4.21
MSRC	N/A	3.02	94.21	86.37	4.81	3.49

saliency detection results are the closest to ground truth.

4.3. Evaluation on CP database

The CP database [16] contains 105 image pairs and 210 images in total. Besides the above eight models GS [31], SF [30], MR [32], SO [35], CB [17], HS [18], RFPR [19], EMR [21], another co-saliency model CG [16] designed for image pairs is joined in comparison. In Fig. 9(a), our PR curve is a little faded against RFPR but is still comparable to HS and EMR. From Fig. 9(b), we conclude that our F -measure is more stable than others. Fig. 9(c) and (d) reveal that the proposed model achieves the least MAE and the highest F_p^w . In summary, the proposed co-saliency model does not perform the best on the PR curve metric because the model is derived from pairwise saliency propagation but the number of images for propagation in the CP database is limited. On the other three metrics, the proposed model surpasses all the others. It is observed in Fig. 10 that the proposed model performs better visually on the five example image pairs.

4.4. Evaluation on the MSRC database

The MSRC database is a bit challenging, where the common objects have different colors. The proposed model is compared both subjectively and objectively with the above eight models GS [31], SF [30], MR [32], SO [35], CB [17], HS [18], RFPR [19], EMR [21]. Although the proposed method mainly uses the color similarity to measure the common objects in different images, the intra-saliency propagation step further enhances the salient object of each image, and that is why the proposed model outperforms the color-based co-saliency models [17–19,21]. Fig. 11 shows that the proposed model surpasses all the other models on all the four evaluation metrics. Besides, the proposed method gets the closest results to ground truth as shown in the two sets of example images in Fig. 12.

4.5. Integration with state-of-the-art single image saliency models

The proposed method uses a single image saliency model MR [32] to produce the initial saliency values, and the resulting co-saliency model is named as CO-MR. We also embed other models like SO [35], SF [30], GS [31] and Itti [36] into the proposed co-saliency model named as CO-SO, CO-SF, CO-GS and CO-Itti. As shown in Fig. 13, all these co-saliency models outperform their initial saliency models on both databases. We observe that CO-MR, CO-SO, CO-SF and CO-GS all achieve comparable performance to the state-of-the-art co-saliency models such as HS, RFPR and EMR. We even embed Itti [36], which is the oldest saliency model generating less satisfying detection results, into the proposed co-saliency model and CO-Itti achieves the state-of-the-art performance. The proposed model is not too sensitive to the initial saliency model. The reason behind this is that our inter-saliency propagation strategy propagates the initial saliency value from the guiding image to the guided one. As long as part of the salient object is highlighted by the initial saliency model, the inter-saliency propagation map will recover the whole object based on color similarity. Furthermore, our intra-saliency propagation step takes advantage of characters of each individual image such as boundary connectivity and edge constraint, making the proposed model robust to background clutter.

4.6. Analysis on the image group containing noisy images

It is important for a co-saliency model to detect the common salient objects not only in the standard database but also in the image group containing noisy images, since such occasion is more common in practical applications. Noisy images mean those which are irrelevant with most images in the whole image group, and they are highlighted with red boxes in Fig. 15.

The proposed model can detect the common salient objects from such kind of image group with a small modification as shown in Fig. 14. After the inter-saliency propagation step, inter-saliency propagation maps which are relevant to noisy images are called noisy intermediates and they will not have any object highlighted, because the objects in noisy images are quite different from the ones in other ordinary images. Thereafter, we compute the global contrast value which denotes the gap between the maximum and minimum values of an inter-saliency propagation map. A simple threshold, which is set to $0.5 * (\max - \min)$ in practice (\max and \min denote the maximum and minimum values of all the inter-saliency propagation maps), can be used to filter out the noisy intermediates. Such images will not go through the intra-saliency

propagation step unlike the other images, as the intra-saliency propagation step will always increase the contrast between the saliency object and the background region. Finally co-saliency integration is used to obtain the co-saliency detection result. It is observed in Fig. 15 that the common salient objects are all popped out while the irrelevant objects are suppressed in the noisy images.

4.7. Computation cost

Table 1 reports the average running time of all the co-saliency models used for comparison. Experiments are implemented on a PC with Intel i3-4130 3.4 GHz CPU and 4GB RAM. From **Table 1** the proposed model is comparable to the most efficient model on both databases, because we convert the co-saliency detection into the pairwise propagation problem, which reduces the computation cost compared to fine segmentation [18,19]. The most time-consuming step is the edge detection [47] and multi-scale superpixel segmentation [43], taking about 2 s and 1.8 s respectively for a typical 300*400 image. Note that the average image size in CP database is about 128*128, so the time cost of above two steps is much lower.

5. Conclusion

This paper presents a novel and efficient method for co-saliency detection. Unlike the existing co-saliency models, the proposed method first considers co-saliency detection as a two-stage saliency propagation problem which uses a single image saliency model to propagate pairwise saliency values. Co-salient foreground cue based on pairwise similarity comparison is obtained in the inter-saliency propagation step. Saliency mask is used to improve the background cue, and it is then integrated with co-salient foreground cue by a graph based regularization with a new edge constraint. Finally the obtained intra-saliency propagation maps are combined using a novel fusion strategy. Multi-scale superpixel segmentation is then used to obtain high quality saliency maps. Both qualitative and quantitative evaluations have shown that the proposed method has outperformed the state-of-the-art models.

Acknowledgments

This research is partly supported by NSFC, China (No. 61273258) and 863 Plan, China (No. 2015AA042308).

Appendix A. Supplementary data

Supplementary data associated with this paper can be found in the online version at <http://dx.doi.org/10.1016/j.image.2016.03.005>.

References

- [1] E. Rahtu, J. Kannala, M. Salo, J. Heikkilä, Segmenting salient objects from images and videos, in: Computer Vision—ECCV 2010, Springer, Crete, Greece, 2010, pp. 366–379.
- [2] L. Wang, J. Xue, N. Zheng, G. Hua, Automatic salient object extraction with contextual cue, in: 2011 IEEE International Conference on Computer Vision (ICCV), IEEE, Barcelona, Spain, 2011, pp. 105–112.
- [3] L. Marchesotti, C. Cifarelli, G. Csurka, A framework for visual saliency detection with applications to image thumbnailing, in: 2009 IEEE 12th International Conference on Computer Vision, IEEE, Kyoto, Japan, 2009, pp. 2232–2239.
- [4] Y. Ding, J. Xiao, J. Yu, Importance filtering for image retargeting, in: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Colorado Springs, CO, USA, 2011, pp. 89–96.
- [5] S. Goferman, L. Zelnik-Manor, A. Tal, Context-aware saliency detection, *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (10) (2012) 1915–1926.
- [6] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, S.-M. Hu, Sketch2photo: internet image montage, in: ACM Transactions on Graphics (TOG), vol. 28, ACM, 2009, p. 124.
- [7] L. Itti, Automatic foveation for video compression using a neurobiological model of visual attention, *IEEE Trans. Image Process.* 13 (10) (2004) 1304–1318.
- [8] Y.-F. Ma, X.-S. Hua, L. Lu, H.-J. Zhang, A generic framework of user attention model and its application in video summarization, *IEEE Trans. Multimed.* 7 (5) (2005) 907–919.
- [9] N. Ejaz, I. Mehmood, S.W. Baik, Efficient visual attention based framework for extracting key frames from videos, *Signal Process.: Image Commun.* 28 (1) (2013) 34–44.
- [10] G. Evangelopoulos, A. Zlatintsi, A. Potamianos, P. Maragos, K. Rapantzikos, G. Skoumas, Y. Avrithis, Multimodal saliency and fusion for movie summarization based on aural, visual, and textual attention, *IEEE Trans. Multimed.* 15 (7) (2013) 1553–1568.
- [11] K. Fu, C. Gong, J. Yang, Y. Zhou, I.Y.-H. Gu, Superpixel based color contrast and color distribution driven salient object detection, *Signal Process.: Image Commun.* 28 (10) (2013) 1448–1463.
- [12] L. Zhou, K. Fu, Y. Li, Y. Qiao, X. He, J. Yang, Bayesian salient object detection based on saliency driven clustering, *Signal Process.: Image Commun.* 29 (3) (2014) 434–447.
- [13] W. Wang, D. Cai, X. Xu, A.W.-C. Liew, Visual saliency detection based on region descriptors and prior knowledge, *Signal Process.: Image Commun.* 29 (3) (2014) 424–433.
- [14] Z. Liu, W. Zou, O. Le Meur, Saliency tree: a novel saliency detection framework, *IEEE Trans. Image Process.* 23 (5) (2014) 1937–1952.
- [15] H.-T. Chen, Preattentive co-saliency detection, in: 2010 17th IEEE International Conference on Image Processing (ICIP), IEEE, Hong Kong, 2010, pp. 1117–1120.
- [16] H. Li, K.N. Ngan, A co-saliency model of image pairs, *IEEE Trans. Image Process.* 20 (12) (2011) 3365–3375.
- [17] H. Fu, X. Cao, Z. Tu, Cluster-based co-saliency detection, *IEEE Trans. Image Process.* 22 (10) (2013) 3766–3778.
- [18] Z. Liu, W. Zou, L. Li, L. Shen, O. Le Meur, Co-saliency detection based on hierarchical segmentation, *IEEE Signal Process. Lett.* 21 (2014) 88–92.
- [19] L. Li, Z. Liu, W. Zou, X. Zhang, et al., Co-saliency detection based on region-level fusion and pixel-level refinement, in: International Conference on Multimedia Expo (ICME), 2014, pp. 1–6.
- [20] X. Cao, Z. Tao, B. Zhang, H. Fu, W. Feng, Self-adaptively weighted co-saliency detection via rank constraint, *IEEE Trans. Image Process.* 23 (9) (2014) 4175–4186.
- [21] Y. Li, K. Fu, Z. Liu, J. Yang, Efficient saliency-model-guided visual co-saliency detection, *IEEE Signal Process. Lett.* 22 (2015) 588–592.
- [22] K.-Y. Chang, T.-L. Liu, S.-H. Lai, From co-saliency to co-segmentation: an efficient and fully unsupervised energy minimization model, in: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Colorado Springs, CO, USA, 2011, pp. 2129–2136.
- [23] S.-M. Hu, T. Chen, K. Xu, M.-M. Cheng, R.R. Martin, Internet visual media processing: a survey with graphics and vision applications, *Vis. Comput.* 29 (5) (2013) 393–405.
- [24] J. Yuan, Y. Wu, Spatial random partition for common visual pattern discovery, in: IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007, IEEE, Rio de Janeiro, Brazil, 2007, pp. 1–8.
- [25] C. Ge, K. Fu, Y. Li, J. Yang, P. Shi, L. Bai, Co-saliency detection via similarity-based saliency propagation, in: 2015 IEEE International Conference on Image Processing (ICIP), IEEE, Quebec City, Canada, 2015.
- [26] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, in: Advances in Neural Information Processing Systems, 2006, pp. 545–552.
- [27] T. Judd, K. Ehinger, F. Durand, A. Torralba, Learning to predict where humans look, in: 2009 IEEE 12th International Conference on Computer Vision, IEEE, Kyoto, Japan, 2009, pp. 2106–2113.
- [28] X. Hou, L. Zhang, Saliency detection: a spectral residual approach, in: IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07, IEEE, Minneapolis, MN, USA, 2007, pp. 1–8.
- [29] M.-M. Cheng, G.-X. Zhang, N.J. Mitra, X. Huang, S.-M. Hu, Global contrast based salient region detection, in: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Colorado Springs, CO, USA, 2011, pp. 409–416.
- [30] F. Perazzi, P. Krahenbuhl, Y. Pritch, A. Hornung, Saliency filters: contrast based filtering for salient region detection, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Providence, RI, USA, 2012, pp. 733–740.
- [31] Y. Wei, F. Wen, W. Zhu, J. Sun, Geodesic saliency using background priors, in: Computer Vision—ECCV 2012, Springer, Florence, Italy, 2012, pp. 29–42.
- [32] C. Yang, L. Zhang, H. Lu, X. Ruan, M.-H. Yang, Saliency detection via graph-based manifold ranking, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Portland, OR, USA, 2013, pp. 3166–3173.
- [33] Q. Yan, L. Xu, J. Shi, J. Jia, Hierarchical saliency detection, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2013, pp. 1155–1162.
- [34] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, S. Li, Salient object detection: a discriminative regional feature integration approach, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Portland, OR, USA, 2013, pp. 2083–2090.

- [35] W. Zhu, S. Liang, Y. Wei, J. Sun, Saliency optimization from robust background detection, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Columbus, OH, USA, 2014, pp. 2814–2821.
- [36] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, *IEEE Trans. Pattern Anal. Mach. Intell.* 20 (11) (1998) 1254–1259.
- [37] Y. Zhai, M. Shah, Visual attention detection in video sequences using spatio-temporal cues, in: Proceedings of the 14th annual ACM international conference on Multimedia, ACM, Santa Barbara, CA, USA, 2006, pp. 815–824.
- [38] C. Guo, Q. Ma, L. Zhang, Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008. cvpr 2008, IEEE, Anchorage, AL, USA, 2008, pp. 1–8.
- [39] R. Achanta, S. Hemami, F. Estrada, S. Süsstrunk, Frequency-tuned salient region detection, in: IEEE Conference on Computer Vision and Pattern Recognition, 2009, CVPR 2009, IEEE, Miami, FL, USA, 2009, pp. 1597–1604.
- [40] D.E. Jacobs, D.B. Goldman, E. Shechtman, Cosaliency: Where people look when comparing images, in: Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology, ACM, New York, NY, USA, 2010, pp. 219–228.
- [41] G. Kim, E.P. Xing, L. Fei-Fei, T. Kanade, Distributed cosegmentation via submodular optimization on anisotropic diffusion, in: 2011 IEEE International Conference on Computer Vision (ICCV), IEEE, Barcelona, Spain, 2011, pp. 169–176.
- [42] F. Meng, H. Li, G. Liu, K.N. Ngan, Object co-segmentation based on shortest path algorithm and saliency model, *IEEE Trans. Multimed.* 14 (5) (2012) 1429–1441.
- [43] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, Slic Superpixels, Technical Report, 2010.
- [44] K. Fu, C. Gong, I.Y. Gu, J. Yang, Geodesic Saliency Propagation for Image Salient Region Detection, in: ICIP, 2013, pp. 3278–3282.
- [45] L. Zhao, S. Liang, Y. Wei, J. Jia, Size and location matter: A new baseline for salient object detection, in: Computer Vision—ACCV 2014, Springer, Singapore, 2015, pp. 578–592.
- [46] N. Otsu, A threshold selection method from gray-level histograms, *Automatica* 11 (285–296) (1975) 23–27.
- [47] P. Dollár, C. L. Zitnick, Structured forests for fast edge detection, in: 2013 IEEE International Conference on Computer Vision (ICCV), IEEE, Sydney, Australia, 2013, pp. 1841–1848.
- [48] K. Fu, C. Gong, I. Gu, J. Yang, Normalized cut-based saliency detection by adaptive multi-level region merging, *IEEE Trans. Image Process.* 24 (12) (2015) 5671–5683.
- [49] X. Li, H. Lu, L. Zhang, X. Ruan, M.-H. Yang, Saliency detection via dense and sparse reconstruction, in: 2013 IEEE International Conference on Computer Vision (ICCV), IEEE, Sydney, Australia, 2013, pp. 2976–2983.
- [50] J. Shotton, J. Winn, C. Rother, A. Criminisi, Textronboost: joint appearance, shape and context modeling for multi-class object recognition and segmentation, in: Computer Vision—ECCV 2006, Springer, Graz, Austria, 2006, pp. 1–15.
- [51] M. Rubinstein, A. Joulin, J. Kopf, C. Liu, Unsupervised joint object discovery and segmentation in internet images, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Portland, OR, USA, 2013, pp. 1939–1946.
- [52] D. Batra, A. Kowdle, D. Parikh, J. Luo, T. Chen, Icoseg: Interactive co-segmentation with intelligent scribble guidance, in: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, San Francisco, CA, USA, 2010, pp. 3169–3176.
- [53] R. Margolin, L. Zelnik-Manor, A. Tal, How to evaluate foreground maps, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Columbus, OH, USA, 2014, pp. 248–255.