Journal of Visual Communication and Image Representation 50 (2018) 227–236

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

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

Saliency integration driven by similar images☆

Jingru Ren[a,b], Zhi Liu[a,b,*], Xiaofei Zhou[a,b], Guangling Sun[b], Cong Bai[c]

[a] Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai 200444, China
[b] School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China
[c] College of Computer Science, Zhejiang University of Technology, Hangzhou 310023, China

ARTICLE INFO

Keywords:
Saliency integration
Saliency propagation
Similar image
Saliency model

ABSTRACT

This paper proposes a saliency integration approach via the use of similar images to elevate saliency detection performance. Given the input image, a group of similar images are first retrieved, and meanwhile, the corresponding multiple saliency maps of the input image are generated by using existing saliency models. Then, the saliency fusion map is generated by using an adaptive fusion method to integrate such saliency maps, for which the fusion weights are measured by the corresponding similarity between each similar image and the input image. Next, an inter-image graph, for each pair of input image and similar image, is constructed to propagate the confident saliency values from the similar image to the input image, yielding the saliency propagation map. Finally, the saliency fusion map and the saliency propagation map are integrated to obtain the final saliency map. Experimental results on two public datasets demonstrate that the proposed approach achieves the better saliency detection performance compared to the existing saliency models and other saliency integration approaches.

1. Introduction

In the past several years, a large number of saliency models have been proposed to identify the visual objects of interest in natural scenes automatically. Inspired by the mechanism of human visual attention, saliency detection has been performed on color images [1–9], RGBD images [10,11], videos [12–14], and a group of images which have the same salient objects [15–18]. Saliency detection plays an important role in many multimedia applications, such as human fixation prediction [19,20], salient object segmentation [21,22], image classification [23], content-based image retrieval [24], image/video retargeting [25–27], image/video compression [28,29], and so forth.

The performance of saliency models usually varies with different images. Motivated by this phenomenon, some works [30–33] investigated the integration of different saliency maps, which are generated by some existing saliency models, to improve saliency detection performance. Borji et al. [30] utilized some simple integration approaches to make combinations of different saliency maps and verified the validity of combinations to improve saliency detection performance. In [31], a data-driven saliency aggregation approach, which exploits the conditional random field (CRF) framework, synthetically considers the interaction among pixels, the performance gap among individual saliency models, and the dependency of saliency models on individual image. Besides, in [32], several unsupervised and supervised learning based integration schemes are performed to investigate whether the aggregation of different saliency maps can improve the performance of fixation prediction or not. Furthermore, in [33], a quality assessment model is exploited to estimate the quality scores of different saliency maps for effective integration.

Nevertheless, it could be insufficient for effective saliency detection only based on the input image, which could be challenging for individual saliency models. Therefore, it is a natural idea to study how we can promote saliency detection performance with the help of other images, such as similar images. In recent years, some saliency models [34–36] with the aid of similar images have been proposed. The similar images, retrieved from an image dataset, share similar objects or background scenes with the input image. The similar object/background regions, which have not been highlighted/suppressed in the input image, could be effectively highlighted/suppressed in similar images. In [34], the saliency values of similar images retrieved from internet collections are propagated to obtain correspondence saliency maps. An approach of saliency transfer is proposed in [35], where the transitional saliency scores are generated by warping the annotations of similar images onto the input image according to the computed dense correspondences. In [36], Nguyen et al. presented an error-aware saliency fusion approach, in which the error-aware coefficients are

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author at: Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai 200444, China.
E-mail address: liuzhisjtu@163.com (Z. Liu).

https://doi.org/10.1016/j.jvcir.2017.12.002
Received 30 June 2017; Received in revised form 2 November 2017; Accepted 9 December 2017
Available online 11 December 2017
1047-3203/

computed based on similar images.

Motivated by the aforementioned analysis, in this paper, we propose a saliency integration approach driven by similar images. Firstly, we retrieve a group of similar images for the input image, and utilize the existing saliency models to generate multiple saliency maps for the input image and its similar images. Secondly, an adaptive fusion method is exploited to integrate the corresponding saliency maps and generate the saliency fusion map for the input image. Thirdly, the confident saliency values are propagated from similar images to the input image via an inter-image propagation method to generate the saliency propagation map. Finally, the saliency fusion map and the saliency propagation map are integrated together to obtain the final saliency map. Particularly, the similar images not only determine the adaptive fusion weights of saliency maps generated by different saliency models, but also directly propagate reliable object and background information to the input image. Overall, the main contributions of this paper are summarized as follows:

(1) We propose a saliency integration approach driven by similar images. The proposed approach utilizes adaptive fusion of multiple saliency maps and inter-image propagation to effectively improve saliency detection performance.
(2) Different from the previous approach [36], the proposed adaptive fusion method is used to generate the saliency fusion map and the adaptive fusion weights for multiple saliency maps are measured by the similarity computed between the input image and each similar image.
(3) The proposed inter-image propagation method constructs an inter-image graph to propagate reliable information from each similar image to the input image, yielding the saliency propagation map, which is an effective complement to the saliency fusion map.

The rest of this paper is organized as follows. Section 2 details the proposed saliency integration approach. Experimental results and analysis are given in Section 3, and conclusions are presented in Section 4.

2. Proposed saliency integration approach

An overview of the proposed saliency integration approach is illustrated in Fig. 1. The proposed approach is described in the following subsections: Section 2.1 briefly introduces how to retrieve the similar images; Section 2.2 presents the adaptive fusion method to generate the saliency fusion map; Section 2.3 elaborates the inter-image propagation method to obtain the saliency propagation map and the final saliency map.

2.1. Retrieve similar images

Given the input image, a group of similar images which have the ground truths of salient objects are first collected based on the work [34], in which the similarity between the input image and the candidate image is computed by using the chi-square distance of Gist descriptor and the weighted color histogram. Specifically, the Gist descriptor, which includes a set of perceptual dimensions such as naturalness, openness, roughness, expansion and ruggedness, is widely used in image retrieval. The weighted color histogram strengthens the color information of salient objects, thus the retrieved images share a more similar salient object with the input image. We select a total of M images with the highest similarity as shown in Fig. 1(d). In our work, M is set to 6, a moderate value for retrieving similar images. Specifically, for the input image, the average of multiple saliency maps is used to weight its color histogram, and for each candidate image, the corresponding ground truth is used to weight its color histogram. The similarity is defined as $Sim(\mathbf{I},\mathbf{R}^m)$, where $\mathbf{I}$ is the input image and $\mathbf{R}^m(m = 1,2,...,M)$ denotes the mth retrieved similar image.

2.2. Adaptive fusion

Although each saliency model may show different performances on different images, it shows similar performance on similar images, *i.e.* using the same saliency model, the saliency maps of similar images exhibit similar quality. For the above reason, we propose an adaptive fusion method driven by similar images to integrate the multiple saliency maps of the input image and generate the saliency fusion map. Specifically, for the adaptive fusion, each group of linear summation coefficients is computed by minimizing the difference between the fused saliency map and the ground truth of one similar image. In this way, we can obtain M groups of linear summation coefficients, and the similarity between each similar image and the input image is exploited to weight the corresponding group of linear summation coefficients.

Let $\mathbf{S}_n^m$ denotes the saliency map (as shown in Fig. 1(e)) of the similar image $\mathbf{R}^m$ generated by the nth existing saliency model, where $n = 1,2,...,N$ and N is the number of existing saliency models. Let w_n^m denote the linear summation coefficient of $\mathbf{S}_n^m$. Let denote the ground truth of $\mathbf{R}^m$ as shown in Fig. 1(c). The objective function for minimization is defined as follows:

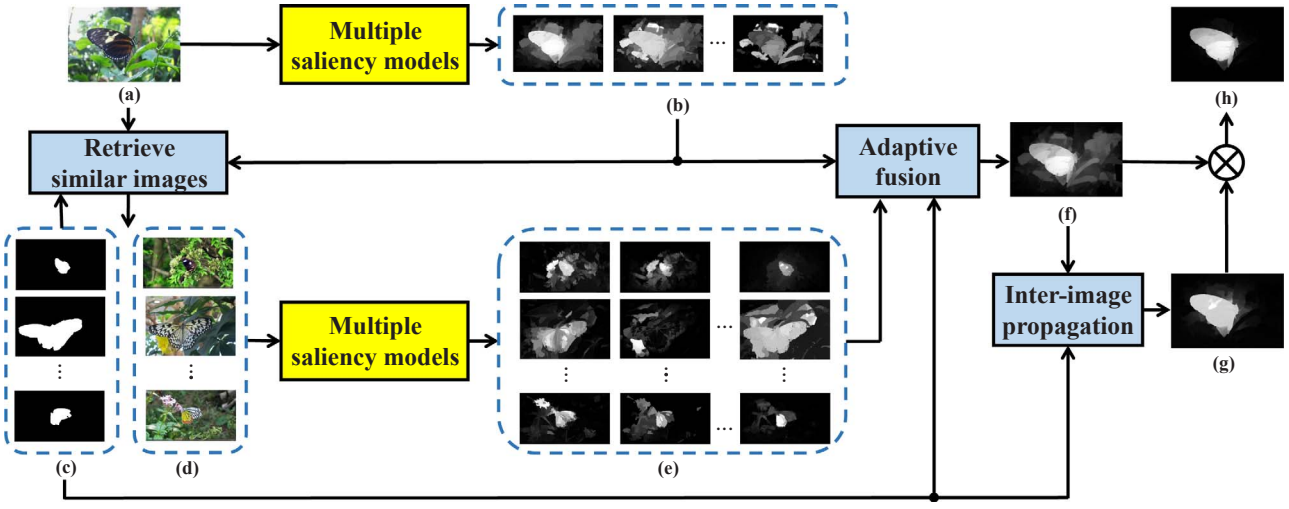


Fig. 1. Overview of the proposed saliency integration approach. (a) Input image; (b) multiple saliency maps of the input image; (c) ground truths of similar images; (d) similar images; (e) multiple saliency maps of similar images; (f) saliency fusion map; (g) saliency propagation map; (h) final saliency map.

$$\mathbf{w}^{m*} = \underset{\mathbf{w}^m}{\operatorname{argmin}} \left\| \sum_{n=1}^{N} w_n^m \mathbf{S}_n^m - \mathbf{G}^m \right\|^2, \quad \text{s. t. } w_n^m \geqslant 0, \tag{1}$$

where $\mathbf{w}^m = [w_1^m, w_2^m, ..., w_N^m]$ represents all the linear summation coefficients for similar image $\mathbf{R}^m$. The objective function can be considered as the non-negative quadratic programming problem and we use the active set method [37] to solve it.

For different similar images of the input image, we can obtain many groups of linear summation coefficients with a little difference. In order to attain the more reasonable fusion weights for multiple saliency maps, we distinguish the confidence of the linear summation coefficients from different similar images according to the similarity. In other words, if one similar image is more similar with the input image, the contribution from this similar image to saliency fusion is more trusted. So, we incorporate the similarity measure, $Sim(\mathbf{I}, \mathbf{R}^m)$, between the input image and each similar image $\mathbf{R}^m$, to obtain the adaptive fusion weights $\theta = [\theta_1, \theta_2, ..., \theta_N]$ as follows:

$$\theta = \frac{1}{M} \sum_{m=1}^{M} [Sim(\mathbf{I}, \mathbf{R}^m) \cdot \mathbf{w}^{m*}]. \tag{2}$$

The adaptive fusion weights consider all the M groups of linear summation coefficients. Then, for the input image $\mathbf{I}$, the saliency fusion map $\mathbf{F}_{fusion}$ by using the adaptive fusion weights is generated as follows:

$$\mathbf{F}_{fusion} = \sum_{n=1}^{N} (\theta_n \cdot \mathbf{S}_n^I), \tag{3}$$

where $\mathbf{S}_n^I$ denotes the saliency map generated by using the nth existing saliency model for the input image $\mathbf{I}$.

The saliency fusion map, as shown in Fig. 1(f), presents the better saliency detection performance, comparing with the saliency maps generated by using the existing saliency models, as shown in Fig. 1(b). However, the saliency fusion map has its limitation owing to the linear summation scheme of adaptive fusion. Specifically, for some regions, which are falsely highlighted by all the existing saliency models, the saliency fusion map also cannot effectively suppress them, such as the leaves around the butterfly. Therefore, a complementary procedure in the following subsection is proposed to compensate the limitation of saliency fusion map and further promote saliency detection performance.

2.3. Inter-image propagation

The saliency fusion maps are not accurate enough, while the similar images with ground truths locate salient objects correctly. It comes to us that similar images can be used to refine the saliency fusion maps by means of inter-image saliency propagation. In [7], an algorithm of saliency propagation is performed on the single image via manifold ranking. Inspired by this, we propose an inter-image propagation method to propagate the confident saliency values of similar images to the saliency fusion map of the input image.

Firstly, the input image and its corresponding similar images are segmented to superpixels at multiple scales by using the SLIC algorithm [38], as shown in Fig. 2(a) and (b). We denote the number of superpixels at the lth scale as $k_l (l = 1,2,...,L)$. The number of scales, L, is set to 8, and the number of superpixels at each scale, $k_l (l = 1,2,...,L)$, is set to 150, 200, 250, 300, 350, 400, 450 and 500, respectively. Then, we construct a close-loop graph $G(V,E)$, where V represents a group of nodes, namely superpixels in our work, and E denotes a group of undirected edges connecting all nodes according to the node connection rules as follows:

(1) Each node is connected to its nearby nodes because nearby nodes usually share similar features and approximate saliency values. Specially, the nearby nodes not only refer to the spatially adjacent neighbors, but also include the nodes which are neighbors of the spatially adjacent neighbors mentioned above. These two kinds of nodes are connected to each node using green lines and yellow lines, respectively, in Fig. 2(c).
(2) All nodes at image borders are connected together, as shown in Fig. 2(c) and (d), using cyan lines. Thus, we construct a close-loop graph, which can shorten the geodesic distance of similar superpixels and facilitate them to subsequently obtain similar saliency values. Moreover, connecting nodes at image borders also means connecting potential background seeds.
(3) The potential object seeds in the input image and the similar image are connected, as shown in Fig. 2(c) and (d), using blue lines. The potential object seeds in the similar image are indicated by the ground truth, while those in the input image are determined by the corresponding saliency fusion map $\mathbf{F}_{fusion}$. Specifically, a superpixel, of which the mean saliency value is greater than the mean saliency value of the entire saliency fusion map $\mathbf{F}_{fusion}$, is regarded as a potential object seed. The potential object seeds are surrounded by carmine lines in Fig. 2(c) and (d). It should be noted that Fig. 2(c) and (d) just shows a small part of connections in order to illustrate the connection rules clearly.

Based on the above three rules, a graph is constructed at each scale. Specifically, at the lth scale, the weights of edges are defined by an affinity matrix, $\mathbf{A} = [a_{ij}]_{2k_l \times 2k_l}$, and $\mathbf{D} = diag\{d_{11}, ..., d_{2k_l 2k_l}\}$ denotes the degree matrix, where $d_{ii} = \sum_{j=1}^{2k_l} a_{ij}$. Since the close-loop graph is constructed for two images, the dimension of the affinity matrix is set to the total number of superpixels in the two images, *i.e.*, $2k_l$. If the ith node is connected with the jth node, their affinity is defined as follows:

$$a_{ij} = \exp\left(-\frac{||c_i - c_j||}{\beta^2}\right), \tag{4}$$

where c_i and c_j denote the average color of pixels covered by the ith node and the jth node, respectively, in the *Lab* color space. β^2 is used to control the strength of weight between a pair of nodes, and is set to 0.1 in our work. If there is no connection between the ith node and the jth node, a_{ij} is set to 0.

Using each similar image $\mathbf{R}^m$, the saliency propagation map at the lth scale is defined as follows:

$$\mathbf{F}_{prop}^{m,l} = (\mathbf{D} - \alpha \mathbf{A})^{-1} y, \tag{5}$$

where $y = [y_1, y_2, ..., y_{2k_l}]^T$ denotes the initial indication vector. In our work, we use the nodes belonging to the potential object seeds in the similar image $\mathbf{R}^m$ as the labeled nodes, since they are generated by the ground truth. The other nodes including potential foreground seeds in the input image are considered as the unlabeled nodes. The parameter α is set to 0.99 according to [7]. The propagation process is to deliver the labeled nodes to the unlabeled nodes according to their relevance. Inevitably, some noise is introduced to result in somewhat inaccurate propagation result. Following [33], the graph cut [39] based refinement is exploited to eliminate noise and improve the spatial coherence of saliency map.

Based on all similar images, a total of M saliency propagation maps are generated and the final saliency propagation map $\mathbf{F}_{prop}$ is defined as follows:

$$\mathbf{F}_{prop} = \frac{1}{L \times M} \sum_{l=1}^{L} \sum_{m=1}^{M} [Sim(\mathbf{I}, \mathbf{R}^m) \cdot \mathbf{F}_{prop}^{m,l}]. \tag{6}$$

For the example in Fig. 1(a), the final saliency propagation map is shown in Fig. 1(g). It can be seen from Fig. 1(g) that background regions are suppressed more effectively and salient object regions are highlighted more uniformly compared to the saliency fusion map. To take advantage of both saliency fusion map and saliency propagation map, we integrate them to obtain the final saliency map as follows:

$$\mathbf{F}_{final} = \mathbf{F}_{fusion} \otimes \mathbf{F}_{prop}, \tag{7}$$

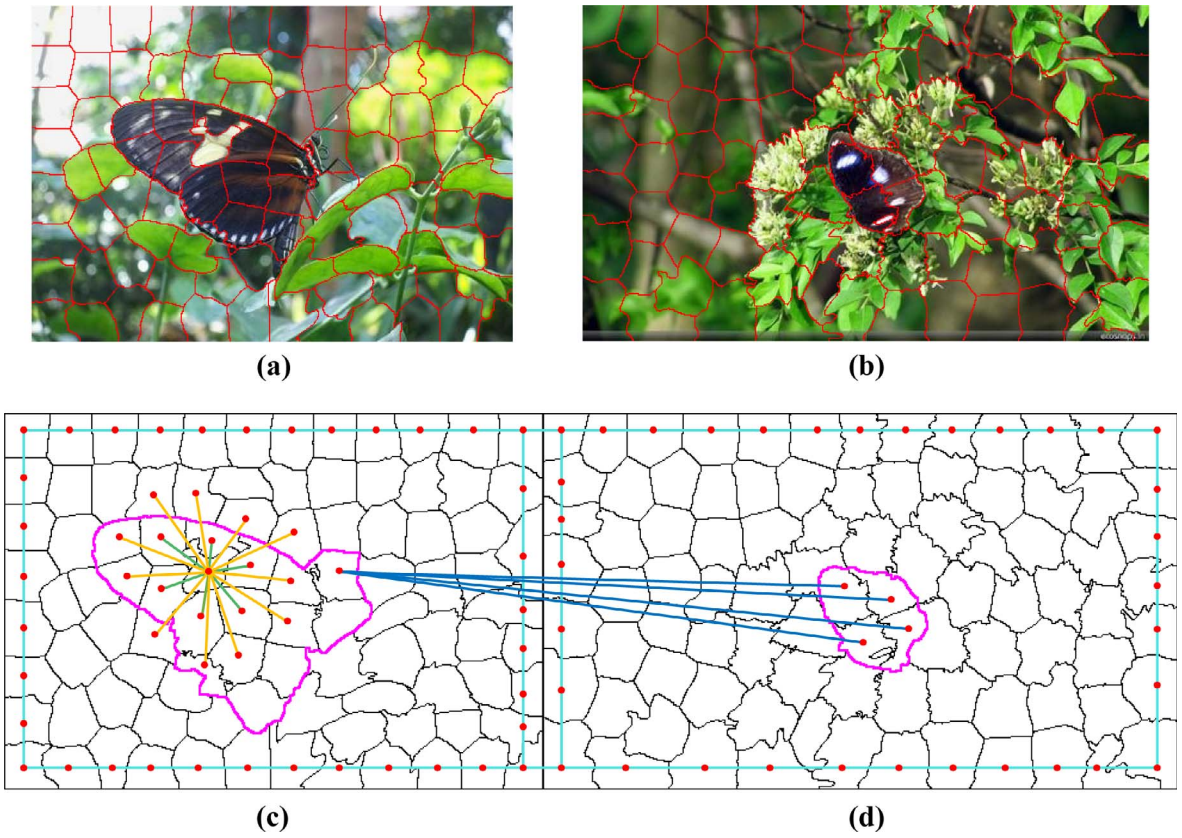
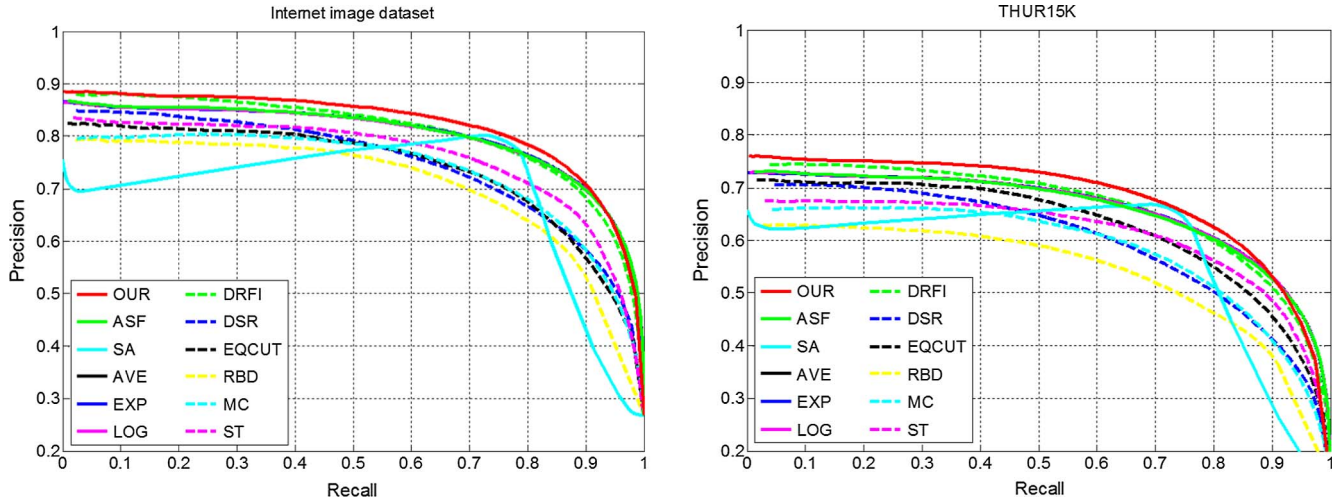


Fig. 2. Illustration of the close-loop graph. (a) The input image overlaid with the segmentation result corresponding to (c); (b) a similar image overlaid with the segmentation result corresponding to (d); (c) and (d) some connection examples for illustrating the proposed node connection rules.

Table 1
Comparison of average F-measure, average F_β^ω-measure and average MAE with six saliency models in *group1*, five integration approaches on two datasets.

Dataset	Metric	OUR	ASF	SA	AVE	EXP	LOG	DRFI	DSR	EQCUT	MC	RBD	ST
Internet image dataset	F-measure	.766	.753	.757	.756	.750	.756	.755	.670	.684	.659	.667	.710
	F_β^ω – measure	.641	.569	.494	.543	.551	.539	.562	.555	.519	.484	.533	.532
	MAE	.165	.172	.196	.190	.191	.190	.180	.197	.196	.210	.199	.198
THUR15K	F-measure	.640	.594	.633	.608	.616	.603	.612	.562	.582	.439	.494	.565
	F_β^ω – measure	.522	.392	.316	.391	.408	.384	.411	.425	.433	.323	.405	.386
	MAE	.116	.171	.183	.157	.147	.163	.147	.142	.137	.184	.148	.179



(a) Internet image dataset (b) THUR15K dataset

Fig. 3. Comparison of precision-recall (PR) curves with six saliency models in *group1*, five integration approaches on two datasets.

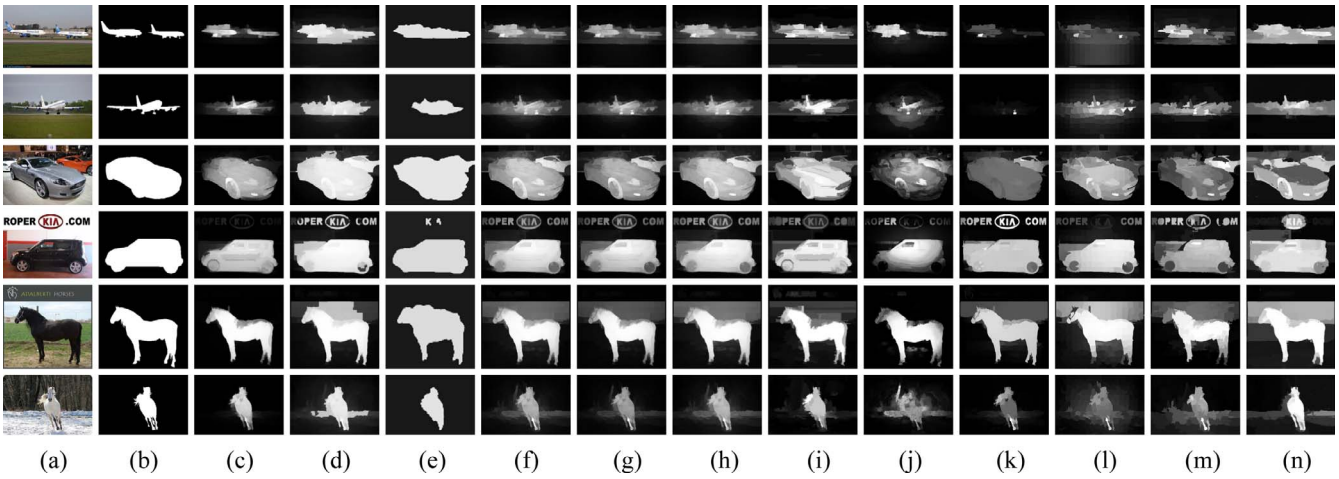


Fig. 4. Visual comparison of saliency maps on the Internet image dataset. (a) Images; (b) ground truths; saliency maps generated by using (c) our saliency integration approach, five saliency integration approaches including (d) ASF, (e) SA, (f) AVE, (g) EXP and (h) LOG, and six saliency models in *group1* including (i) DRFI, (j) DSR, (k) EQCUT, (l) MC, (m) RBD and (n) ST.

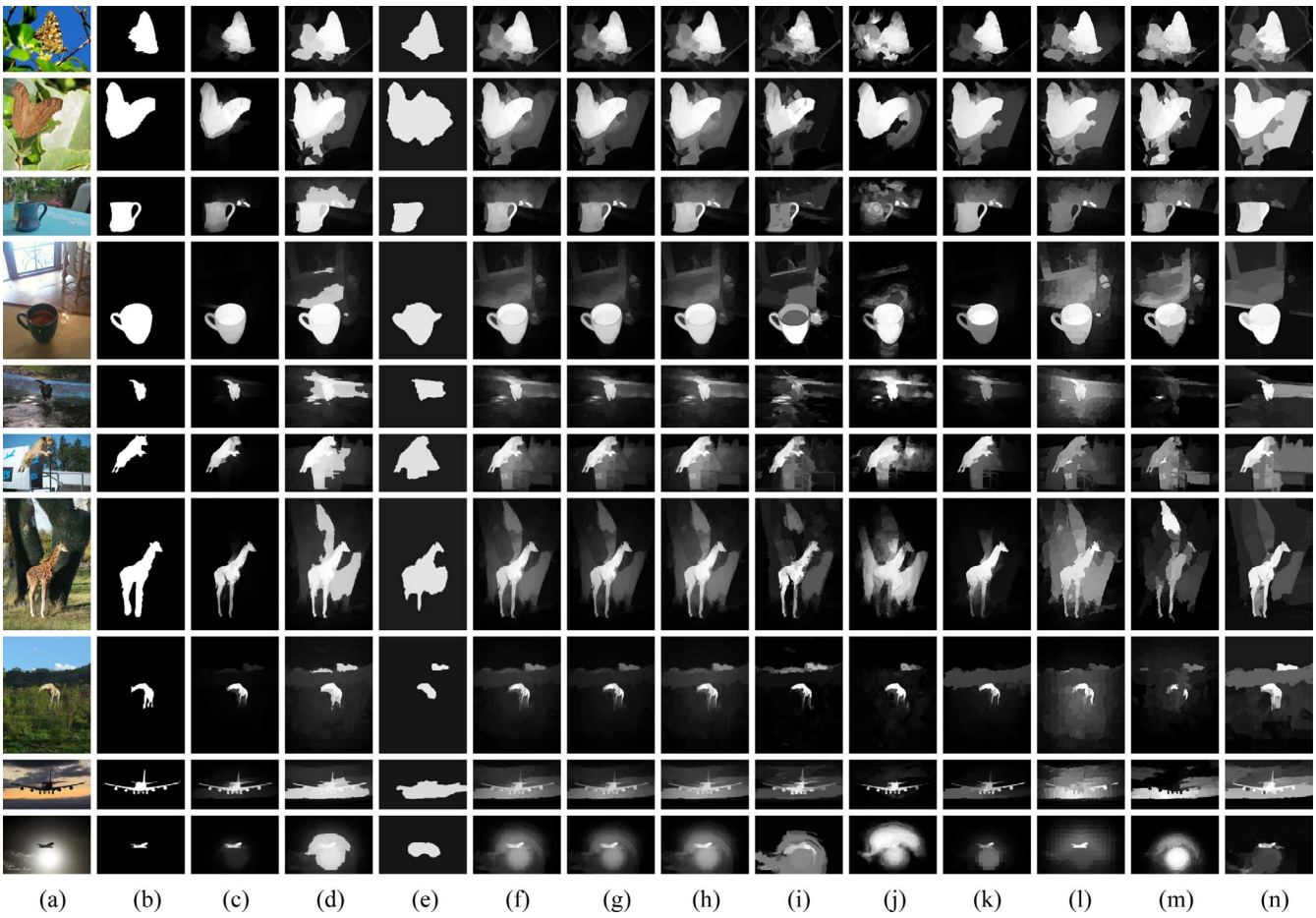


Fig. 5. Visual comparison of saliency maps on the THUR15K dataset. (a) Images; (b) ground truths; saliency maps generated by using (c) our saliency integration approach, five saliency integration approaches including (d) ASF, (e) SA, (f) AVE, (g) EXP and (h) LOG, and six saliency models in *group1* including (i) DRFI, (j) DSR, (k) EQCUT, (l) MC, (m) RBD and (n) ST.

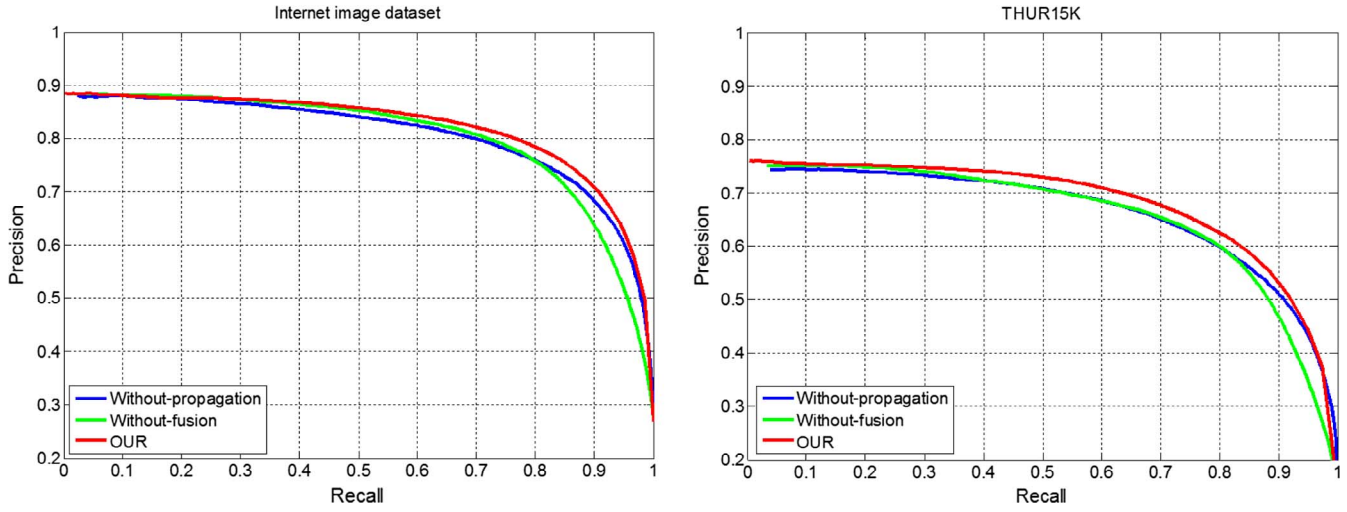
where the pixel-wise multiplication operation $\otimes$ is exploited to better suppress background regions and meanwhile preserve salient object regions highlighted in both saliency fusion map and saliency propagation map. The final saliency map, as shown in Fig. 1(h), can better highlight the salient object more uniformly with well-defined boundaries and suppress background more effectively compared to the corresponding saliency fusion map and saliency propagation map.

3. Experimental results

3.1. Experimental setting

3.1.1. Dataset

To verify the effectiveness of the proposed saliency integration approach driven by similar images, we evaluate its performance on a total of 2488 images from the Internet image dataset [40] and a total of 6233



(a) Internet image dataset (b) THUR15K dataset

Fig. 6. Comparison of precision-recall (PR) curves with two variants on two datasets.

Table 2
Comparison of average F-measure, average F_β^ω-measure and average MAE with two variants on two datasets (the best performance is marked with boldface).

Datasets	Metric	OUR	Without-propagation	Without-fusion
Internet image dataset	F-measure	**0.766**	0.759	0.753
	F_β^ω-measure	**0.641**	0.557	0.589
	MAE	**0.165**	0.183	0.168
THUR15K	F-measure	**0.640**	0.627	0.611
	F_β^ω-measure	**0.522**	0.413	0.443
	MAE	**0.116**	0.150	0.144

images from the THUR15K dataset [24], in which all images are manually annotated with pixel-wise binary ground truths. The Internet image dataset has three classes of objects covering airplane, car and horse. The THUR15K dataset contains five classes of objects including butterfly, dog jump, coffee mug, giraffe and plane. The images in both datasets were downloaded from the internet and marked by a keyword, thus they fit to the proposed approach.

3.1.2. Evaluation metrics

We evaluate all saliency maps using four metrics including precision-recall (PR) curve, F-measure, weighted F_β^ω-measure [41] and mean absolute error (MAE). The PR curve is an important metric to measure saliency detection performance. We binarize the saliency map using each integer threshold from 0 to 255, and calculate the precision value and recall value to plot the PR curve with recall as x-coordinate and precision as y-coordinate. In order to calculate the F-measure, the binary object mask is first obtained by adaptively thresholding each saliency map using [42], the precision and recall are then calculated by comparing the binary object mask with the ground truth, and the F-measure combining the precision and recall with β^2 as the balance factor is defined as follows:

$$F_\beta = \frac{(1+\beta^2)Precision \times Recall}{\beta^2 Precision + Recall}. \tag{8}$$

The F_β^ω-measure generalizes the F-measure to a unified evaluation for binary or non-binary map, and thus can directly evaluate the saliency map with the similar definition as follows:

$$F_\beta^\omega = \frac{(1+\beta^2)Precision^\omega \times Recall^\omega}{\beta^2 Precision^\omega + Recall^\omega}, \tag{9}$$

where $Precision^\omega$ and $Recall^\omega$ (namely weighted precision and weighted recall) are computed by the extended basic quantities including true positive, true negative, false positive and false negative, which are weighted according to the pixels' location and neighborhood [41]. The parameter β^2 in both F-measure and F_β^ω-measure is set to 0.3 to indicate more importance of precision than recall according to [43]. The MAE computes the difference at pixel level between the saliency map S and the ground truth G, and is defined as follows:

$$MAE = \frac{1}{W \times H}\sum_{x=1}^{W}\sum_{y=1}^{H}|S(x,y)-G(x,y)|, \tag{10}$$

where W and H are the width and the height, respectively, of the saliency map S.

3.1.3. Compared integration approaches

We compare our results with the existing individual saliency models, which are used for our saliency integration, and the three integration approaches in [30]:

$$p(x|S_1,S_2,\ldots,S_N) \propto \frac{1}{Z}\sum_{i=1}^{N}\zeta(p(x|S_n)), \tag{11}$$

where $\zeta(x)$ takes the function: (1) $\zeta(x)=x$; (2) $\zeta(x)=\exp(x)$; (3) $\zeta(x)=-1/\log(x)$. The three integration approaches are denoted as "AVE", "EXP" and "LOG" in order. Furthermore, we also compare with two machine learning based saliency fusion approaches, SA [31] and ASF [33], which are trained on 3000 images from the MSRA10K dataset [44].

3.1.4. Overview of experiments

We performed a set of experiments to sufficiently evaluate the performance of our saliency integration approach and compare with other saliency integration approaches. We use the six existing saliency models with the highest performance in the light of the benchmark [43], *i.e.*, DRFI [1], DSR [2], EQCUT [3], MC [4], RBD [5] and ST [6], denoted as *group1*, to generate multiple saliency maps. Notably, we adopt the idea of cross-validation for the purpose of splitting the testing dataset and retrieval dataset. Specifically, we first divide each class of the Internet images dataset and the THUR15K into three average portions randomly. Then we treat each one portion as the testing dataset in turn and the remaining two portions as the retrieval dataset, from which the similar images are retrieved. Finally, we obtain the saliency integration results of all images in each dataset. The quantitative comparison and qualitative comparison are shown in s 3.2 and 3.3,

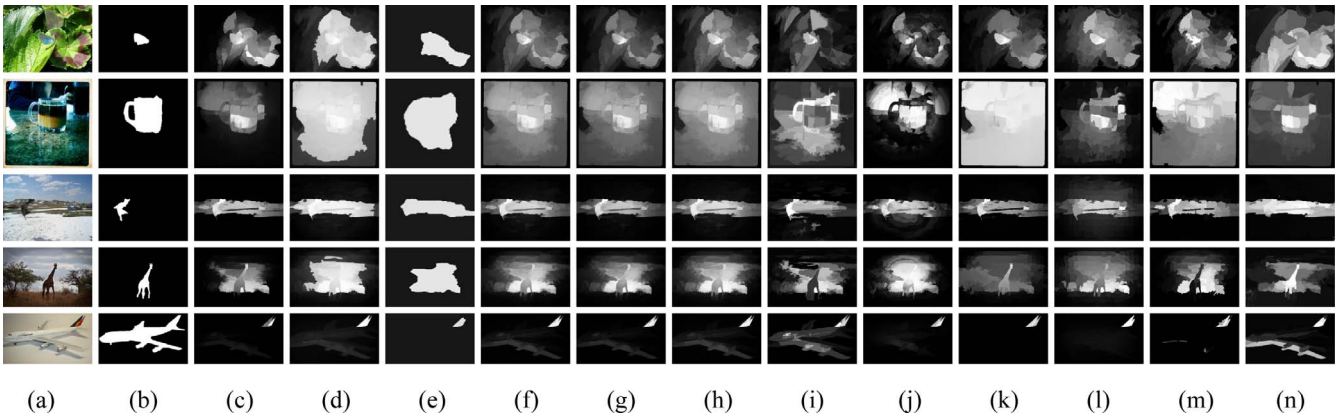


Fig. 7. Failure examples of the proposed approach on the THUR15K dataset. (a) Images; (b) ground truths; saliency maps generated by using (c) our saliency integration approach, five saliency integration approaches including (d) ASF, (e) SA, (f) AVE, (g) EXP and (h) LOG, and six saliency models in *group1* including (i) DRFI, (j) DSR, (k) EQCUT, (l) MC, (m) RBD and (n) ST.

Table 3
Comparison of average F-measure, average F_β^ω-measure and average MAE with the six saliency models in *group2*, five integration approaches on two datasets.

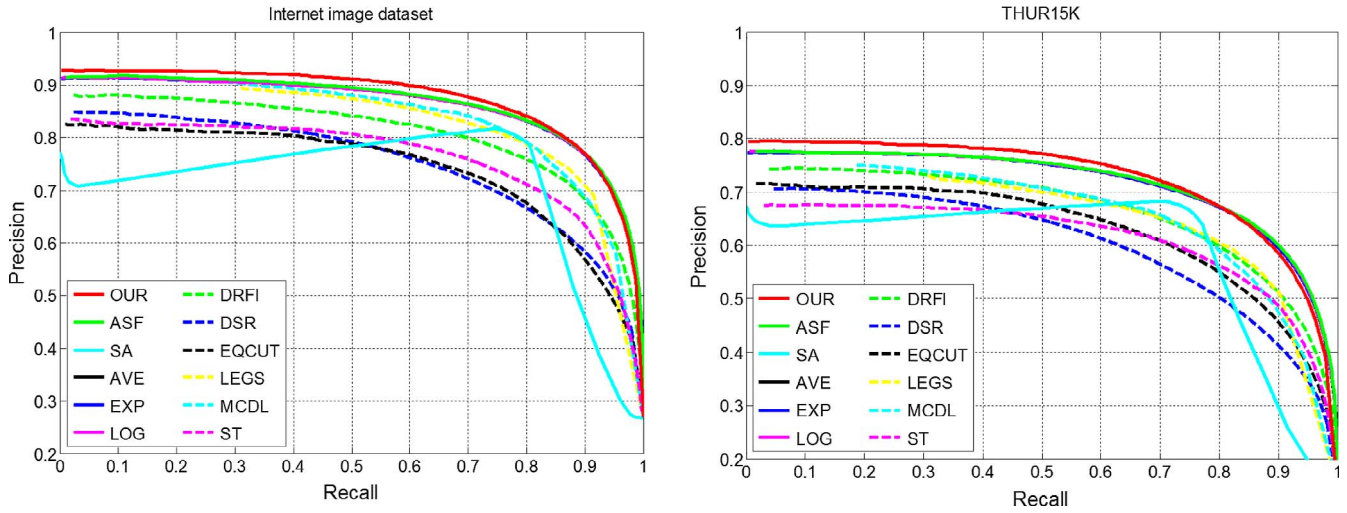
Dataset	Metric	OUR	ASF	SA	AVE	EXP	LOG	DRFI	DSR	EQCUT	LEGS	MCDL	ST
Internet image dataset	F-measure	.815	.804	.777	.804	.808	.802	.755	.670	.684	.773	.766	.710
	F_β^ω −measure	.706	.636	.578	.606	.618	.600	.562	.555	.519	.698	.700	.532
	MAE	.138	.148	.186	.163	.164	.163	.180	.197	.196	.139	.113	.198
THUR15K	F-measure	.677	.649	.650	.654	.663	.648	.612	.562	.582	.629	.577	.565
	F_β^ω −measure	.561	.449	.323	.443	.461	.436	.411	.425	.433	.538	.539	.386
	MAE	.106	.143	.178	.141	.133	.145	.147	.142	.137	.125	.107	.179

respectively. The ablation study is given in Section 3.4, and some failure examples and analysis are presented in Section 3.5.

Some further analyses are presented in Section 3.6 based on the two additional experiments. (1) To evaluate the performance with a different group of saliency models used for integration, we introduced the recently proposed deep learning based saliency models with the higher performance, LEGS [45] and MCDL [46], to replace the previous two saliency models with the relatively lower performance, MC and RBD. Specifically, a new group of saliency models including DRFI, DSR, EQCUT, LEGS, MCDL and ST, denoted *group2*, is used for saliency integration. (2) To evaluate the effectiveness and robustness of our approach more adequately, we performed experiments by using a different dataset, the PASCAL-VOC dataset [47], as the retrieval dataset.

3.2. Quantitative comparison

All the results of average F-measure, average F_β^ω-measure and average MAE for each dataset are listed in Table 1. In each row of Table 1, the best performance, the second best performance and the third one are marked with red, green and blue, respectively. It can be observed that in terms of F-measure, F_β^ω-measure and MAE, our saliency integration approach consistently performs best on both datasets. In terms of PR curve, we can see from Fig. 3 that our approach achieves the highest performance on both datasets. As an overall evaluation based on all the four metrics, our saliency integration approach improves saliency detection performance more effectively than all the other saliency integration approaches.



(a) Internet image dataset (b) THUR15K dataset

Fig. 8. Comparison of precision-recall (PR) curves with six saliency models in *group2*, five integration approaches on two datasets.

Table 4
Comparison of average F-measure, average F_β^ω-measure and average MAE with six saliency models in *group1* and five integration approaches on the Internet image dataset when the retrieval dataset is PASCAL-VOC.

Dataset	Metric	OUR	ASF	SA	AVE	EXP	LOG	DRFI	DSR	EQCUT	MC	RBD	ST
Internet image dataset	F-measure	.761	.753	.757	.756	.750	.756	.755	.670	.684	.659	.667	.710
	F_β^ω – measure	.635	.569	.494	.543	.551	.539	.562	.555	.519	.484	.533	.532
	MAE	.168	.172	.196	.190	.191	.190	.180	.197	.196	.210	.199	.198

Table 5
Comparison of average F-measure, average F_β^ω-measure and average MAE with six saliency models in *group2* and five integration approaches on the Internet image dataset when the retrieval dataset is PASCAL-VOC.

Dataset	Metric	OUR	ASF	SA	AVE	EXP	LOG	DRFI	DSR	EQCUT	LEGS	MCDL	ST
Internet image dataset	F-measure	.812	.804	.777	.804	.808	.802	.755	.670	.684	.773	.766	.710
	F_β^ω – measure	.701	.636	.578	.606	.618	.600	.562	.555	.519	.698	.700	.532
	MAE	.139	.148	.186	.163	.164	.163	.180	.197	.196	.139	.113	.198

3.3. Qualitative comparison

Some saliency maps on both datasets are shown in Figs. 4 and 5 for a qualitative comparison. Overall, the saliency maps integrated by our approach show the best visual quality compared to those generated by the six saliency models and integrated by all the other saliency integration approaches. It can be seen that comparing with all the other saliency integration approaches, our approach can suppress background regions more effectively and highlight salient objects more uniformly with well-defined boundaries. For example, the sky in the 5th row of Fig. 4 and the trunk behind the giraffe in the 7th row of Fig. 5 are highlighted incorrectly in overwhelming individual saliency maps, but our final saliency maps suppress such background regions effectively. When some indistinctive objects mix in the background, such as the leaf behind the butterfly (in the 2nd row of Fig. 5) and the flaming sun (in the last row of Fig. 5), most of individual saliency maps falsely highlight them, while our final saliency maps can suppress them effectively. The comparison clear demonstrates the superiority of our approach to promote saliency detection performance.

3.4. Ablation study

To analyze the contribution of each step in our approach, we evaluate two variants of the proposed saliency integration approach with different settings. As shown in Fig. 6 and Table 2, "Without-fusion" represents the first variant without the adaptive fusion and "Without-propagation" represents the second variant without inter-image propagation. Considering the feasibility and comparability of such experiments, we use the average of multiple saliency maps to replace the saliency fusion map in the first variant in order to provide the potential object seeds for propagation. The results of the two variants are compared with the results of our approach on all the four metrics. In each row of Table 2, the best performance is marked with boldface. The comparisons on both datasets show the reasonability and superiority of our approach.

3.5. Failure examples

Some failure examples are shown in Fig. 7. The main reason for these failure examples is that the low contrasts between salient objects and background induce the fatal errors in saliency detection. For instance, in the first row of Fig. 7, the green butterfly shows similar colors with the leaves in the background. Since the saliency integration heavily depends on the individual saliency maps involved in the fusion, the proposed approach cannot pop out the salient objects well when all the individual saliency maps are incapable of highlighting salient objects effectively. It can be seen from Fig. 7 that all the other results generated by using individual saliency models and all the other saliency integration approaches also fail to effectively highlight the salient objects.

3.6. Further analysis

We further performed experiments with the six saliency models in *group2*, which includes DRFI, DSR, EQCUT, LEGS, MCDL and ST. The experimental results of average F-measure, average F_β^ω-measure and average MAE on the Internet image dataset and the THUR15K dataset are listed in Table 3. In each row of Table 3, the best performance, the second best performance and the third one are marked with red, green and blue, respectively. The PR curves on both datasets are shown in Fig. 8. We can see from Table 3 and Fig. 8 that our approach achieves the overall better performance than the six saliency models in *group2* and five integration approaches on both datasets. These comparison results signify that our approach works well with different groups of saliency models for integration. Comparing with the results for *group1* (Table 1 and Fig. 3), the results for *group2* (Table 3 and Fig. 8) achieve the further improvement due to the introduction of two deep learning based saliency models, LEGS and MCDL, which show the better performance. Therefore, we can conclude that the saliency detection performance of our approach has a positive correlation with the performance of saliency models used for integration.

To further evaluate the effectiveness and robustness of our approach to different retrieval dataset, the PASCAL-VOC dataset is used as the

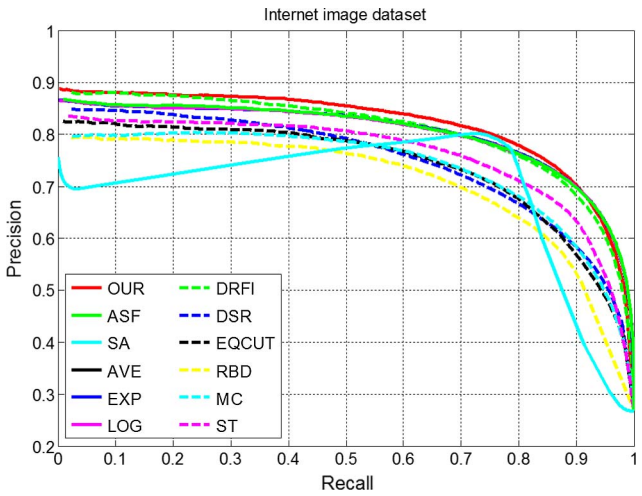


Fig. 9. Comparison of precision-recall (PR) curves with six saliency models in *group1* and five integration approaches on the Internet image dataset when the retrieval dataset is PASCAL-VOC.

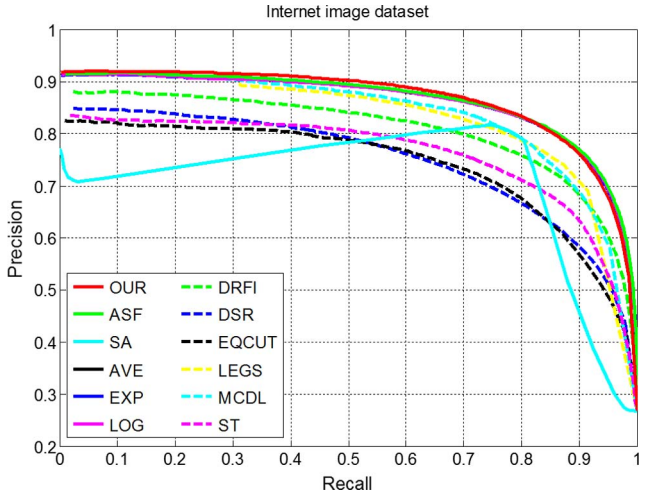


Fig. 10. Comparison of precision-recall (PR) curves with six saliency models in *group2* and five integration approaches on the Internet image dataset when the retrieval dataset is PASCAL-VOC.

retrieval dataset and the Internet image dataset is used as the testing dataset. Specifically, the PASCAL-VOC dataset includes 20 classes and a total of 1037 images, which are manually annotated with pixel-wise binary ground truths. For the Internet image dataset, the three classes of images including airplane (470 images), car (1204 images) and horse (810 images) have the corresponding classes in the PASCAL-VOC dataset, *i.e.*, aeroplane (66 images), car (63 images) and horse (40 images), respectively, as the retrieval dataset. The experiments were implemented on two groups of saliency models including *group1* and *group2*. The results of average F-measure, average F_β^ω-measure and average MAE are listed in Tables 4 and 5, and the PR curves are shown in Figs. 9 and 10. We can see that our approach outperforms all individual saliency models and integration approaches on all the four evaluation metrics, except for the saliency model MCDL only in terms of the average MAE. Therefore, with the different retrieval dataset, we can conclude that our approach is still effective and shows the better robustness.

4. Conclusion

In this paper, we propose a novel saliency integration approach driven by similar images. Specifically, the saliency fusion map is first generated via adaptively fusing multiple saliency maps of the input image. Secondly, in order to further improve saliency detection performance, a complementary saliency map, namely the saliency propagation map, is computed by the inter-image propagation, which transfers the confident saliency values from the similar images to the input image. Finally, the combination of saliency fusion map and saliency propagation map results in the high-quality final saliency map. Experimental results on two public datasets demonstrate the effectiveness of the proposed saliency integration approach to boost saliency detection performance.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grants 61771301 and 61502424, Shanghai Municipal Natural Science Foundation under Grant No. 16ZR1411100, and Zhejiang Provincial Natural Science Foundation of China under Grants LY15F020028 and LY18F020032.

References

[1] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, S. Li, Salient object detection: A discriminative regional feature integration approach, Proc. IEEE CVPR, Portland, Oregon, USA, 2013, pp. 2083–2090.
[2] X. Li, H. Lu, L. Zhang, X. Ruan, M.-H. Yang, Saliency detection via dense and sparse reconstruction, Proc. IEEE ICCV, Sydney, NSW, Australia, 2013, pp. 2976–2983.
[3] C. Aytekin, E. Ozan, S. Kiranyaz, and M. Gabbouj, Visual saliency by extended quantum cuts, in: Proc. IEEE ICIP, Quebec City, QC, Canada, 2015, pp. 1692–1696.
[4] B. Jiang, L. Zhang, H. Lu, C. Yang, M.-H. Yang, Saliency detection via absorbing Markov chain, Proc. IEEE ICCV, Melbourne, Victoria, Australia, 2013, pp. 1665–1672.
[5] W. Zhu, S. Liang, Y. Wei, J. Sun, Saliency optimization from robust background detection, Proc. IEEE CVPR, Columbus, Ohio, USA, 2014, pp. 2814–2821.
[6] Z. Liu, W. Zou, O. Le Meur, Saliency tree: a novel saliency detection framework, IEEE Trans. Image Process. 23 (2014) 1937–1952.
[7] C. Yang, L. Zhang, H. Lu, R. Xiang, M.H. Yang, Saliency detection via graph-based manifold ranking, Proc. IEEE CVPR, Portland, Oregon, USA, 2013, pp. 3166–3173.
[8] C. Lang, J. Feng, G. Liu, J. Tang, S. Yan, J. Luo, Improving bottom-up saliency detection by looking into neighbors, IEEE Trans. Circuits Syst. Video Technol. 23 (2013) 1016–1028.
[9] P. Siva, C. Russell, X. Tao, L. Agapito, Looking beyond the image: unsupervised learning for object saliency and detection, Proc. IEEE CVPR, Portland, Oregon, USA, 2013, pp. 3238–3245.
[10] H. Song, Z. Liu, H. Du, G. Sun, O. Le Meur, T. Ren, Depth-aware salient object detection and segmentation via multiscale discriminative saliency fusion and bootstrap learning, IEEE Trans. Image Process. 26 (2017) 4204–4216.
[11] H. Song, Z. Liu, Y. Xie, L. Wu, M. Huang, RGBD co-saliency detection via bagging-based clustering, IEEE Signal Process. Lett. 23 (2016) 1722–1726.
[12] D. Zhang, J. Han, L. Jiang, S. Ye, X. Chang, Revealing event saliency in unconstrained video collection, IEEE Trans. Image Process. 26 (2017) 1746–1758.
[13] Y. Fang, W. Lin, Z. Chen, C.M. Tsai, C.W. Lin, A video saliency detection model in compressed domain, IEEE Trans. Circuits Syst. Video Technol. 24 (2014) 27–38.
[14] Z. Liu, J. Li, L. Ye, G. Sun, L. Shen, Saliency detection for unconstrained videos using superpixel-level graph and spatiotemporal propagation, IEEE Trans. Circuits Syst. Video Technol. (2016), http://dx.doi.org/10.1109/TCSVT.2016.2595324.
[15] D. Zhang, D. Meng, J. Han, Co-saliency detection via a self-paced multiple-instance learning framework, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 865–878.
[16] D. Zhang, J. Han, C. Li, J. Wang, X. Li, Detection of co-salient objects by looking deep and wide, Int. J. Comput. Vis. 120 (2016) 215–232.
[17] D. Zhang, J. Han, J. Han, L. Shao, Cosaliency detection based on intrasaliency pior tansfer and deep intersaliency mining, IEEE Trans. Neural Netw. Learn. Syst. 27 (2016) 1163–1176.
[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] Y. Yang, M. Song, N. Li, J. Bu, C. Chen, What is the chance of happening: a new way to predict where people look, Proc. ECCV, Springer, Hersonissos, Heraklion, Crete, Greece, 2010, pp. 631–643.
[20] M. Song, C. Chen, S. Wang, Y. Yang, Low-level and high-level prior learning for visual saliency estimation, Inf. Sci. 281 (2014) 573–585.
[21] Z. Liu, R. Shi, L. Shen, Y. Xue, K.N. Ngan, Z. Zhang, Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut, IEEE Trans. Multimedia 14 (2012) 1275–1289.
[22] W. Zou, Z. Liu, K. Kpalma, J. Ronsin, Y. Zhao, N. Komodakis, Unsupervised joint salient region detection and object segmentation, IEEE Trans. Image Process. 24 (2015) 3858–3873.
[23] G. Sharma, F. Jurie, C. Schmid, Discriminative spatial saliency for image classification, Proc. IEEE CVPR, Providence, Rhode Island, USA, 2012, pp. 3506–3513.
[24] M.-M. Cheng, N.J. Mitra, X. Huang, S.-M. Hu, SalientShape: group saliency in image collections, Vis. Comput. 30 (2014) 443–453.
[25] V. Setlur, T. Lechner, M. Nienhaus, B. Gooch, Retargeting images and video for preserving information saliency, IEEE Comput. Graph. Appl. 27 (2007) 80–88.
[26] H. Du, Z. Liu, J. Jiang, L. Shen, Stretchability-aware block scaling for image retargeting, J. Vis. Commun. Image Represent. 24 (2013) 499–508.
[27] Y. Fang, Z. Chen, W. Lin, C. Lin, Saliency detection in the compressed domain for adaptive image retargeting, IEEE Trans. Image Process. 21 (2012) 3888–3901.
[28] C. Guo, L. Zhang, A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression, IEEE Trans. Image Process. 19 (2010) 185–198.
[29] L. Shen, Z. Liu, Z. Zhang, A novel H.264 rate control algorithm with consideration of visual attention, Multimed. Tools Appl. 63 (2013) 709–727.
[30] A. Borji, D.N. Sihite, L. Itti, Salient object detection: a benchmark, Proc. ECCV, Springer, Florence, Tuscany, Italy, 2012, pp. 414–429.
[31] L. Mai, Y. Niu, F. Liu, Saliency aggregation: a data-driven approach, in: Proc. IEEE CVPR, Portland, Oregon, USA, 2013, pp. 1131–1138.
[32] O. Le Meur, Z. Liu, Saliency aggregation: does unity make strength? Proc. ACCV, Springer, Singapore, 2014, pp. 18–32.
[33] X. Zhou, Z. Liu, G. Sun, X. Wang, Adaptive saliency fusion based on quality assessment, Multimed. Tools Appl. (2016), http://dx.doi.org/10.1007/s11042-016-4093-8.
[34] L. Ye, Z. Liu, X. Zhou, L. Shen, J. Zhang, Saliency detection via similar image retrieval, IEEE Signal Process. Lett. 23 (2016) 838–842.
[35] W. Wang, J. Shen, L. Shao, F. Porikli, Correspondence driven saliency transfer, IEEE Trans. Image Process. 25 (2016) 5025–5034.
[36] T.V. Nguyen, M. Kankanhalli, As-similar-as-possible saliency fusion, Multimed.

Tools Appl. 76 (2017) 10501–10519.
[37] C.L. Lawson, R.J. Hanson, Solving Least Squares Problems, Prentice-Hall, New York, 1974.
[38] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, SLIC superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2012) 2274–2282.
[39] V. Kolmogorov, R. Zabin, What energy functions can be minimized via graph cuts? IEEE Trans. Pattern Anal. Mach. Intell. 26 (2004) 147–159.
[40] M. Rubinstein, A. Joulin, J. Kopf, C. Liu, Unsupervised joint object discovery and segmentation in internet images, in: Proc. IEEE CVPR, Portland, Oregon, USA, 2013, pp. 1939–1946.
[41] R. Margolin, L. Zelnik-Manor, A. Tal, How to evaluate foreground maps? in: Proc. IEEE CVPR, Columbus, Ohio, USA, 2014, pp. 248–255.
[42] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. 9 (1979) 62–66.
[43] A. Borji, M.-M. Cheng, H. Jiang, J. Li, Salient object detection: a benchmark, IEEE Trans. Image Process. 24 (2015) 5706–5722.
[44] THUR15000 dataset [Online], 2015. Available: http://mmcheng.net/gsal/.
[45] L. Wang, H. Lu, M. Yang, Deep networks for saliency detection via local estimation and global search, in: Proc. IEEE CVPR, Boston, Massachusetts, USA, 2015, pp. 3183–3192.
[46] R. Zhao, W. Ouyang, H. Li, X. Wang, Saliency detection by multi-context deep learning, in: Proc. IEEE CVPR, Boston, Massachusetts, USA, 2015, pp. 1265–1274.
[47] M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman, The PASCAL visual object classes (VOC) challenge, Int. J. Comput. Vis. 88 (2010) 303–338.