

Adaptive saliency fusion based on quality assessment

Xiaofei Zhou¹ · Zhi Liu¹  · Guangling Sun¹ ·
Xiangyang Wang¹

Received: 23 May 2016 / Revised: 29 September 2016 / Accepted: 24 October 2016 /

Published online: 10 November 2016

© Springer Science+Business Media New York 2016

Abstract A variety of saliency models based on different schemes and methods have been proposed in the recent years, and the performance of these models often vary with images and complement each other. Therefore it is a natural idea whether we can elevate saliency detection performance by fusing different saliency models. This paper proposes a novel and general framework to adaptively fuse saliency maps generated using various saliency models based on quality assessment of these saliency maps. Given an input image and its multiple saliency maps, the quality features based on the input image and saliency maps are extracted. Then, a quality assessment model, which is learned using the boosting algorithm with multiple kernels, indicates the quality score of each saliency map. Next, a linear summation method with power-law transformation is exploited to fuse these saliency maps adaptively according to their quality scores. Finally, a graph cut based refinement method is exploited to enhance the spatial coherence of saliency and generate the high-quality final saliency map. Experimental results on three public benchmark datasets with state-of-the-art saliency models demonstrate that our saliency fusion framework consistently outperforms all saliency models and other fusion methods, and effectively elevates saliency detection performance.

Keywords Saliency model · Adaptive fusion · Quality assessment · Boosting

✉ Zhi Liu
liuzhisjtu@163.com

Xiaofei Zhou
zxforchid@outlook.com

Guangling Sun
sunguangling@shu.edu.cn

Xiangyang Wang
wangxiangyang@shu.edu.cn

¹ School of Communication and Information Engineering, Shanghai University, Shanghai, China

1 Introduction

Visual saliency, inspired by the mechanism of visual attention in humans, aims to make certain regions in the scene stand out from their surroundings, and has received more and more attention in the recent years. Numerous applications such as content-based image/video compression [14, 46], salient object detection and segmentation [30, 32, 48], content-aware image/video retargeting [10–12, 44], classification [45], retrieval [16], to name a few, benefit from saliency detection as a preprocessing step to focus on the area of great importance.

There are numerous literatures on saliency detection, and two benchmarks have been reported in [3, 4], which show the comprehensive comparison among a variety of saliency models. The early research on saliency model is motivated by stimulating the visual attention mechanism of human visual system (HVS), by which only the significant portion of the scene projected onto retina can be processed by human brain for semantic understanding. As a pioneering work on saliency detection, Itti et al. proposed a well-known bottom-up saliency model [19], in which the center-surround differences across multi-scale image features are calculated and then the operation of normalization and summation is used for generating saliency map. A graph-based saliency model is proposed in [15], which utilizes the Markovian approach on an active map. In [17], the spectral analysis in frequency domain is used to detect salient region. In [30], a set of saliency features including multi-scale contrast, center-surround histogram, and color spatial distribution are fused to generate the saliency map under the framework of conditional random field (CRF). A saliency model which exploits the statistics of natural images and introduces the Bayesian framework for saliency computation is proposed in [55]. In [1], an efficient and simple saliency model is proposed based on the center-surround scheme by comparing the color features of each pixel with the average color of the whole image. Multiple cues including local low-level features, visual organization rules and high-level features are simultaneously modeled to improve saliency detection performance with the context of salient object in [13]. Another successful saliency model based on kernel density estimation (KDE) is proposed in [29], where a set of KDE models are constructed based on the region segmentation result. A global contrast based saliency model is proposed in [8], which considers the global region contrast with respect to the entire image and spatial relationships across regions to compute saliency map. In [31], the Gaussian model is adopted to represent each region, and both color and spatial saliency measures of Gaussian models are evaluated and integrated to measure the pixel-level saliency. Background prior is studied in [50] to formulate a geodesic distance based saliency model. Under the framework of low-rank matrix recovery, a region segmentation based object prior is exploited for saliency detection in [58]. Distinctiveness and compactness of regional histograms in [33] as well as global contrast and spatial sparsity of superpixels in [34] are proposed for saliency measurement.

In order to improve saliency detection performance on the basis of existing saliency models, some meaningful works [3, 25, 38, 51] explore the fusion of different saliency models. In [3], a simple fusion method (sum) using three normalization schemes (identity, exponential and logarithmic) is used for combining saliency models, and their results show that combining several best saliency models enhance the saliency detection performance. Bayesian integration is proposed to fuse low-level and mid-level cues for saliency map generation in [51]. A data-driven saliency aggregation approach (denoted as SA) under the CRF framework is proposed in [38], which focuses on modeling the contribution of individual saliency map and the interaction between neighboring pixels. In [25], supervised and unsupervised learning methods are tested to aggregate different saliency models for fixation prediction, and the simple average of saliency maps generated using the two best models is already a good candidate for saliency fusion.

Actually, there are some other methods [20, 26, 53, 57] proposed for the optimization on individual saliency model. In [53], the saliency maps computed on the hierarchical image segmentations are integrated using a tree-structure graphical model. In order to meet several hypotheses on saliency including visual rarity, center bias and mutual correlation, a quadratic programming function which optimizes the saliency values of all superpixels in an image to simultaneously meet all the hypotheses is proposed in [26]. A similar optimization-based framework is proposed in [57] to integrate multiple foreground/background cues. To fuse the high-level object information with pixel-level appearances effectively, a Markov Random Field (MRF) is adopted to enforce the consistence between salient regions in [20]. Although the above mentioned methods [20, 26, 53, 57] cannot be directly used as the fusion method for different saliency models, the research of saliency fusion may refer to these methods and make some adaptations to design an effective fusion method.

With the rapid and continuous development of saliency detection, more and more advanced models are proposed recently. According to the rankings of recently published benchmark [4], there are six state-of-the-art saliency models [2, 21, 22, 27, 35, 57] with the highest performance. In [57], a new boundary prior called “boundary connectivity” and a principled optimization framework are proposed to improve saliency detection performance. In [22], the random forest regression is used to map the regional discriminative feature vector to the saliency score of each region, and saliency scores across multiple levels are fused to obtain the final regional saliency score. In [27], a saliency measure via dense and sparse representation errors of each region is proposed and the final saliency map is generated by integrating multiscale reconstruction errors. A bottom-up saliency detection model is proposed in [21], which considers the appearance divergence and spatial distribution of salient objects and background using the time property in an absorbing Markov chain. The saliency tree model in [35] enables the hierarchical representation of saliency and improves saliency detection performance. The link between quantum mechanics and graph cuts is exploited to generate the saliency map in [2]. In [42], the cellular automaton is used to detect salient object intuitively. In [23], a global saliency model via high dimensional color space transformation and a local saliency model via random forest regression are combined to generate a new saliency model, which estimates object regions from a trimap. In [18], a novel compactness hypothesis including color and texture is proposed as a remedy to address the weakness of contrast hypothesis from the perspective of both color layout and texture layout.

Although the above state-of-the-art saliency models achieve the higher performance statistically on the public benchmarks, there still exists a large margin for performance improvement. In addition, the performance of an individual model varies with images. As illustrated in Fig. 1, the discriminative regional feature integration (DRFI) model [22] performs effectively in the 1st row, which achieves a similar quality with our result by comparing to the ground truth. The saliency tree (ST) model [35] performs better than other models in the 2nd row, the robust background detection (RBD) model [57] achieves the better performance in the 3rd row, and the background-based map optimized via single-layer cellular automata (BSCA) model [42] outperforms other models in the 4th row. The quantum cuts (QCUT) model [2] and the dense and sparse reconstruction (DSR) model [27] work well in the 5th and 6th row, respectively. The high-dimensional color transform (HDCT) model [23] and Markov chain model (MC) [21] generate saliency maps with the highest quality in the 7th and 8th row, respectively. It can be clearly seen from Fig. 1 that all these saliency models cannot perform well in all the eight rows. However, it can be also seen from Fig. 1 that these models often

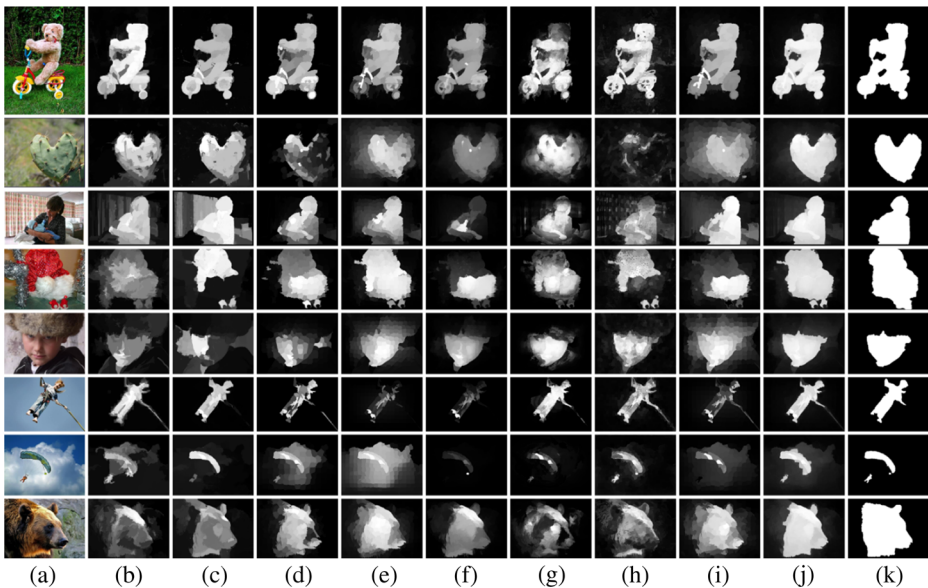


Fig. 1 The performance of individual saliency model varies with images. **a** Images; saliency maps generated using **b** DRFI [22], **c** ST [35], **d** RBD [57], **e** BSCA [42], **f** QCUT [2], **g** DSR [27], **h** HDCT [23] and **i** MC [21], respectively; **j** our fusion results; **k** ground truths

complement each other, for example, in the 2nd row, though RBD suppresses the central part of cactus, DSR could complement it by highlighting the central part.

Therefore, it is a natural idea to study whether the fusion of different saliency models could make improvements on saliency detection performance or not. Specifically, formulating an effective fusion framework to combine different saliency models may generate the better saliency map. Besides, quality estimation recently attracts more and more attention and has been applied in the recommender system [41], which proposes a novel graph-based regularized algorithm that learns the ranking function in the semi-supervised learning framework. Furthermore, for multi-focus image fusion, a dense scale invariant feature transform [36] is proposed to evaluate the clarity of source images, and for integration of multi-view information, a multi-view intact space learning algorithm [52] is proposed to integrate the encoded complementary information in multiple views to discover a latent intact representation of the data.

Motivated by the above analysis, in this paper, we propose a novel and general framework to fuse different saliency models adaptively based on quality assessment of their saliency maps. First, given an input image and its saliency maps generated using different saliency models, we extract effective quality features on the input image and its saliency maps. Second, a quality assessment model is constructed based on quality features using the boosting algorithm with multiple kernels to estimate the appropriate quality scores for these saliency maps. Third, based on the obtained quality scores, a linear summation method with power-law transformation is exploited to fuse these saliency maps adaptively and then generate the fused saliency map. Finally, a graph cut based refinement method is exploited to improve the spatial coherence of the fused saliency map and generate the high-quality final saliency map. The flowchart of the proposed saliency fusion framework is illustrated in Fig. 2.

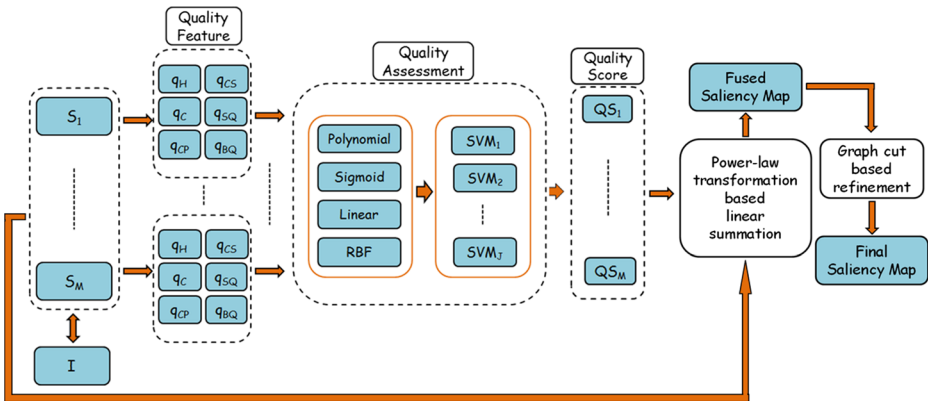


Fig. 2 Flowchart of the proposed saliency fusion framework

The most related work (denoted as IS) to this paper is proposed in [56], which presents a quality based adaptive fusion method for improving saliency detection performance of an existing saliency model. The two works are considerably different. First, their scopes of application are different. The method in [56] can only be used for fusing two saliency maps, since it is used in [56] to fuse initial saliency map and complementary saliency map, which is generated based on initial saliency map, and to achieve an improvement for individual saliency model. In contrast, our method can be used to fuse any number of saliency maps with the aim to elevate saliency detection performance. Second, the quality assessment models are different on the following two aspects: a) The selection strategies of positive and negative samples are different. The AUC (area under the ROC curve) score is exploited in [56], while the coverage score is adopted by our method. b) The use of quality features and the learning algorithms are different, the method in [56] simply concatenates all quality features to form a feature vector and adopts a simple SVM classifier to generate the quality assessment model, while our method adopts the boosting algorithm with multiple kernels [54] to take full use of quality features and to formulate a quality assessment model. Third, the fusion schemes based on quality scores are different. The method in [56] uses the simple linear weighted summation, while our method introduces the power-law transformation and the graph cut based refinement for saliency fusion.

Another related work (denoted as CS) to this paper is proposed in [37], which mainly aims to rank saliency maps according to their quality features without ground truth. Based on quality features, the best salient object detection result from saliency maps are selected to improve saliency detection performance. Overall, our work uses the same quality features as [37], but there are considerable difference between [37] and our work. First, the work in [37] mainly aims to rank saliency maps based on quality features, and simply selects the best one from saliency maps based on the ranking result to improve saliency detection performance; while our work proposes a method to fuse various saliency maps based on quality features. Second, regarding the training strategy, a pairwise-based learning-to-rank methodology [7] is adopted in [37], while our method trains a general quality assessment model for different saliency maps and fuses them based on quality scores. Third, the use of quality features and selection of training samples are different. In [37], each saliency map is represented using a concatenation of all kinds of quality features, and meanwhile AUC values are exploited to obtain the label of each saliency map in the training set. In contrast, our method adopts multiple kernel boosting [54] to take full use of quality features for each saliency map, and

utilizes coverage score to select high-quality/low-quality saliency maps as positive/negative training samples.

Overall, the main contribution of this paper lies in the following three aspects:

- 1) We propose a general quality assessment model, which assigns an appropriate quality score for each saliency map. Specially, in order to take full advantage of different kinds of quality features, we exploit a boosting algorithm with multiple kernels as the core of our quality assessment model.
- 2) Different from previous fusion methods, we propose a linear summation method with power-law transformation to effectively utilize quality scores for adaptive fusion of various saliency maps, and thereafter a graph cut based refinement method to enhance the spatial coherence of saliency map.
- 3) We performed extensive experiments with state-of-the-art saliency models on three public benchmark datasets and detailed comparisons with previous fusion methods. The results demonstrate the effectiveness of our method and show a possible way to further push forward saliency detection performance.

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

2 Proposed method

As shown in Fig. 2, the proposed saliency fusion framework starts from running a total of M saliency models on a given image I with a size of $W * H$, and generates M saliency maps, $\{S_i\}_{i=1}^M$. Each saliency map S_i is normalized to $[0, 1]$. The following subsections are arranged as follows: Sect. 2.1 briefly introduces the quality features; Sect. 2.2 describes the quality assessment model to obtain quality scores for saliency maps; Sect. 2.3 presents the fusion method based on quality scores for generating the fused saliency map; Sect. 2.4 details the graph cut based saliency refinement for generating the final saliency map.

2.1 Quality feature

A number of quality features has been proposed in [37], and here we classify these quality features into two classes: quality features based on only saliency map, and quality features based on interaction between saliency map and input image. The quality features are summarized in Table 1 and briefly described in the following.

Saliency histogram this feature represents the distribution of saliency values. Given a saliency map S_i , its saliency histogram q_H^i is defined as follows:

$$q_H^i(b_j) = \frac{n_j}{W * H}, j = 1, 2, \dots, N_H \quad (1)$$

where n_j denotes the number of pixels falling into the j^{th} bin b_j , and N_H is the total number of bins in the histogram.

Table 1 Quality features of saliency map S_i

Class	Quality Feature	Description	Dim
Quality features based on only saliency map	Saliency histogram q_H^i	The histogram shows peaks at two ends with high-quality saliency map, while the uniform distribution with low-quality saliency map.	20
	Saliency coverage q_C^i	The saliency coverage is abnormally large or small for low-quality saliency map.	10
	Saliency compactness q_{CP}^i	A high-quality saliency map usually concentrates its salient pixels in a compact region.	3
Quality features based on interaction between saliency map and input image	Color separation q_{CS}^i	A high-quality saliency map tends to have a smaller color separation value.	1
	Segmentation quality q_{SQ}^i	A high-quality saliency map usually generate a good segmentation result.	3
	Boundary quality q_{BQ}^i	The accurate and well-defined object boundary is an important property of high-quality saliency maps.	4

Saliency coverage This feature indicates the estimated size of salient object based on saliency map. The feature value of a low-quality saliency map is usually abnormally large or small. Given a saliency map S_i , its saliency coverage q_C^i is defined as follows:

$$q_C^i(t) = \frac{1}{W * H} \sum_{p \in S_i} \Psi[S_i(p) - t] \tag{2}$$

where $\Psi[r]$ is the sign function with a value of 1 when $r > 0$ and 0 otherwise. The threshold $t \in [0, 1]$ is set to 10 different values in the range of $[0, 1]$.

Saliency compactness This feature evaluates the density of salient pixels distributed in the most salient area. Given a saliency map S_i , its saliency compactness q_{CP}^i is defined as follows:

$$\begin{aligned} \{R_j\}_{j=1}^{N_T} \text{ s.t. } & \left(\sum_{p \in R_j} S_i(p) / \sum_p S_i(p) \right) = T \\ j^* = \arg \min_j & (W(R_j) * H(R_j)), j = 1, \dots, N_T \\ q_{CP}^i(T) = & \sum_{p \in R_{j^*}} S_i(p) / |R_{j^*}| \end{aligned} \tag{3}$$

where R_j denotes the j^{th} rectangular window covering a proportion of the total saliency in S_i , and there are totally N_T windows for each threshold T (here we set T to 0.25, 0.5 and 0.75, respectively). For each window R_j , we first compute its area $W(R_j) * H(R_j)$, then find the window R_{j^*} with the smallest area, and compute its mean saliency value as the feature q_{CP}^i where $|R_{j^*}|$ is the number of total pixels in R_{j^*} .

Color separation this feature denotes the separation on color distribution between salient regions and background regions. It is a weighted color histogram incorporating saliency values. Given a saliency map S_i , its color separation q_{CS}^i is defined as follows:

$$\begin{aligned}
 h_s^i(j) &= \frac{\sum_{p \in I} S_i(p) \delta\{I(p) \in b_j\}}{\sum_{p \in I} S_i(p)} \\
 h_b^i(j) &= \frac{\sum_{p \in I} (1 - S_i(p)) \delta\{I(p) \in b_j\}}{\sum_{p \in I} (1 - S_i(p))} \\
 q_{CS}^i &= \frac{1}{N_{CS}} \sum_{j=1}^{N_S} \min(h_s^i(j), h_b^i(j))
 \end{aligned}
 \tag{4}$$

where b_j denotes the color range of the j^{th} bin (here we set 16 bins per channel in the RGB color space), and $\delta\{\cdot\}$ is the indicator function with a value of 1 if its argument is true and 0 otherwise. h_s^i and h_b^i are the color histogram for salient regions and background, respectively. The feature q_{CS}^i indicates the intersection between the two color histograms, and N_{CS} is the number of histogram bins.

Segmentation quality This feature represents the quality of saliency map by assessing the segmentation result induced by the saliency map. Given a saliency map S_i , its segmentation quality q_{SQ}^i can be computed by using the normalized cut based energy function [47] as follows:

$$q_{SQ}^i(t) = \frac{\sum_{p \in S_{i,S}^t, q \in S_{i,B}^t, q \in N(p)} w_{pq}}{\sum_{p \in S_{i,S}^t, q \in N(p)} w_{pq}} + \frac{\sum_{p \in S_{i,B}^t, q \in S_{i,S}^t, q \in N(p)} w_{pq}}{\sum_{p \in S_{i,B}^t, q \in N(p)} w_{pq}}
 \tag{5}$$

where $S_{i,S}^t$ and $S_{i,B}^t$ denote salient regions and background regions generated from the saliency map S_i with the threshold t , and here we use three thresholds: 0.5, 0.75 and 0.95. $N(p)$ denotes the neighborhood of the pixel p , and w_{pq} represents the color similarity between the neighboring pixels, p and q , with the same definition as [30].

Boundary quality this feature evaluates the correlation between the boundary map $B_{S_i}^M$ generated using the saliency map S_i and the strong edge map E_I generated by performing the structured-forests edge detection [9] on the input image I . Concretely, the boundary quality feature q_{BQ}^i for S_i is defined as follows:

$$\begin{aligned}
 w_p &= S_i(p) \max(|S_i(p) - S_i(p_1)|, |S_i(p) - S_i(p_2)|) \\
 B_{S_i}^M(p) &= \frac{w_p |S_i(p_1) - S_i(p_2)|}{\sum_{p \in M} w_p} \\
 q_{BQ}^{i,n}(I) &= \sum_{p \in I} B_{S_i}^M(p) E_I(p)
 \end{aligned}
 \tag{6}$$

where p_1 and p_2 are two neighboring pixel of p , and they locate orthogonally with respect to the edge direction at p . w_p is the saliency-weighted edge magnitude. The superscript n denotes the n^{th} scale of input image I , and here we use four scales with the scaling ratio of 0.25, 0.5, 0.75 and 1, respectively. The boundary quality feature q_{BQ}^i is the concatenation of feature values $q_{BQ}^{i,n}$ calculated at the four scales.

Therefore, for a given saliency map S_i with the corresponding input image I , there are totally six different kinds of quality features: saliency histogram q_H^i , saliency coverage q_C^i , saliency compactness q_{CP}^i , color separation q_{CS}^i , segmentation quality q_{SQ}^i and boundary quality q_{BQ}^i , and the quality feature of the saliency map S_i is finally defined

$$q^i = \{q_H^i, q_C^i, q_{CP}^i, q_{CS}^i, q_{SQ}^i, q_{BQ}^i\} = \{q_k^i\}_{k=1}^{n_F} \text{ with } n_F = 6.$$

2.2 Quality assessment

Our aim is to learn the quality assessment model from a set of training examples and then assigns an appropriate quality score to each saliency map. As aforementioned, there are totally six different kinds of quality features for a given training sample. In order to take full advantage of these quality features, one of the common schemes is to adopt the kernel transformation of these features with support vector machine (SVM). However, it is not appropriate for our situation, because it is difficult to select an appropriate kernel for the diverse samples with different properties of quality features. To cope with this problem effectively, we exploit multiple kernel boosting (MKB) method [54] to include multiple kernels of different quality features. In the framework of MKB, SVMs with different kernels are regarded as weak classifiers and then a strong classifier can be learned by using the AdaBoost method. Here four types of kernels $\{K_m\}_{m=1}^{n_K}$ (linear, radial basis function (RBF), sigmoid and polynomial, $n_K = 4$) with six kinds of quality features are included in our model.

For learning the quality assessment model, we construct the training set consisting of quality features as follows. First, we select n_D color images and the corresponding ground truths, $\{I_d, G_d\}_{d=1}^n$, from a public image dataset. Second, for a given color image I_d , we obtain the quality features $\{q^i\}_{i=1}^M$ with the saliency maps $\{S_i\}_{i=1}^M$ generated using M saliency models, and each one is defined as $q^i = \{q_k^i\}_{k=1}^{n_F}$. Finally, in order to obtain deterministic quality feature samples, we make a selection among the samples $\{q^i\}_{i=1}^M$ with the ground truth G_d , and assign labels $\{y^i\}_{i=1}^M$ simultaneously:

$$T_i = \frac{\sum_p S_i(p) * G_d(p)}{\sum_p G_d(p)}, \quad y^i = \begin{cases} +1 & T_i \geq T_{high} \\ -1 & T_i \leq T_{low} \end{cases} \quad (7)$$

where T_i denotes the ratio of actual salient pixels indicated by the saliency map S_i compared to the ground truth G_i . T_{low} and T_{high} are the low and high threshold, respectively. According to Eq. (7), we can see that if T_i falls into the range of (T_{low}, T_{high}) , the corresponding quality feature sample will be discarded. The quality feature sample q^i is labeled as a positive sample and y^i is set to +1 if $T_i \geq T_{high}$, or labeled as a negative sample and y^i is set to -1 if $T_i \leq T_{low}$. So

for each color image and its ground truth, $\{I_d, G_d\}$, we obtain a set of quality feature samples $\{q^i, y^i\}_{i=1}^{M^*}$ where $M^* \leq M$. The quality feature samples are collected from all training images and their ground truths, and the training set is finally denoted as $\{q^n, y^n\}_{n=1}^{n_T}$.

Following the framework of MKB, the decision function is defined as follows:

$$F(q) = \sum_{j=1}^J \beta_{(j)} H_{(j)}(q) \tag{8}$$

where $H_{(j)}(q) = \alpha_{(j)}^T \mathbf{K}_{(j)} + b_{(j)}$ denotes a SVM classifier (weak classifier), $J = n_F \times n_K$ is the number of weak classifiers, and $\beta_{(j)}$ is the kernel weight for the j^{th} weak classifier. The kernel is defined as $\mathbf{K}_{(j)} = [K_{(j)}(q, q^1), \dots, K_{(j)}(q, q^{n_T})]^T$, the Lagrange multiplier vector is defined as $\alpha_{(j)} = [\alpha_{(j)}^1 y^1, \dots, \alpha_{(j)}^{n_T} y^{n_T}]^T$, and $b_{(j)}$ is the bias in the SVM classifier. All the above parameters are obtained after the AdaBoost optimization process.

Since the raw decision value $F(q)$ is unbounded, we transform it to the quality score in the range of $[0, 1]$ by using the following function:

$$QS(q) = \frac{1}{1 + e^{-\theta \cdot F(q)}} \tag{9}$$

where θ is the parameter for the decay rate of quality score, and is set to 1 for a moderate decay effect.

By using the MKB algorithm, we obtain the quality assessment model via adaptively integrating the most discriminative features and the corresponding kernels. For each saliency map S^* , using its quality feature as the input, $q = q^*$, the quality assessment model outputs the estimated quality score, $QS(q^*)$, for the saliency map S^* .

2.3 Power-law transformation based linear summation

With the quality scores, which are obtained by the learned quality assessment model for different saliency maps, and for the purpose of utilizing quality scores reasonably and performing an adaptive fusion of saliency maps, a linear summation method with the power-law transformation is proposed to fuse saliency maps adaptively. Specifically, for a test image I_d and the corresponding saliency maps $\{S_i\}_{i=1}^M$ generated using different saliency models, we extract quality features $\{q^i\}_{i=1}^M$ for these saliency maps and estimate the quality scores $\{QS(q^i)\}_{i=1}^M$ via the quality assessment model. The fusion weights for these saliency maps $\{S_i\}_{i=1}^M$ are computed as $w_i = QS(q^i) / \sum_{j=1}^M QS(q^j)$, and the fused saliency map is defined as follows:

$$S_F(p) = P(y_p = 1 | \{S_i(p), w_i\}_{i=1}^M) = \text{Norm} \left[\sum_{i=1}^M (S_i(p))^{w_i} \right] \tag{10}$$

where the operation $\text{Norm}[\cdot]$ normalizes the saliency map into the range of $[0, 1]$. $S_F(p)$ indicates the probability of each pixel p being salient given a total of M saliency maps and their corresponding weights. Eq. (10) performs the linear summation with the power-law transformation, which is illustrated in Fig. 3. According to the quality assessment model, the obtained

quality score of each saliency map ranges from 0 to 1. For the high-quality saliency map S_i with a high quality score, such as $w_i = 0.95$, the saliency value of each pixel after power-law transformation, $S_i(p)^{w_i}$, is similar as the original saliency value, $S_i(p)$, and it indicates that the change of high-quality saliency map after power-law transformation is negligible. With the decrease of quality score, it can be seen from Fig. 3 that the dynamic range of moderate saliency values is progressively narrowed. For example, with $w_i = 0.6$, the range $[0.2, 0.8]$ of original saliency values is mapped to the range $[0.38, 0.87]$ after power-law transformation. Furthermore, for the low-quality saliency map with a rather low quality score, such as $w_i = 0.01$, the original saliency values in the nearly complete range $[0, 1]$ are all transformed to very close to 1.0, and thus a uniform saliency map is generated for fusion. The contribution of such a uniform saliency map to the fused saliency map can be neglected, since it merely adds nearly the same amount of saliency to all pixels and actually cannot change the spatial distribution of saliency values after normalization. Based on the above analysis, using Eq. (10), the quality score reasonably adjusts the contribution of the corresponding input saliency map to the fused saliency map, and such an adaptive fusion method effectively emphasizes high-quality saliency maps and suppresses low-quality saliency maps to generate the fused saliency map.

2.4 Graph cut based refinement

In the previous subsection, we have fused different saliency maps based on their quality scores, but the fused saliency map S_F may introduce some noises around object boundaries and may non-uniformly highlight object regions. Therefore, based on the fused saliency map, we exploit a simple yet effective refinement method incorporating saliency and color based on graph cut [24] to improve the spatial coherence of saliency map. Specifically, for the test image I_d , we construct an undirected graph $G = (V, E)$, in which each pixel corresponds to a node in

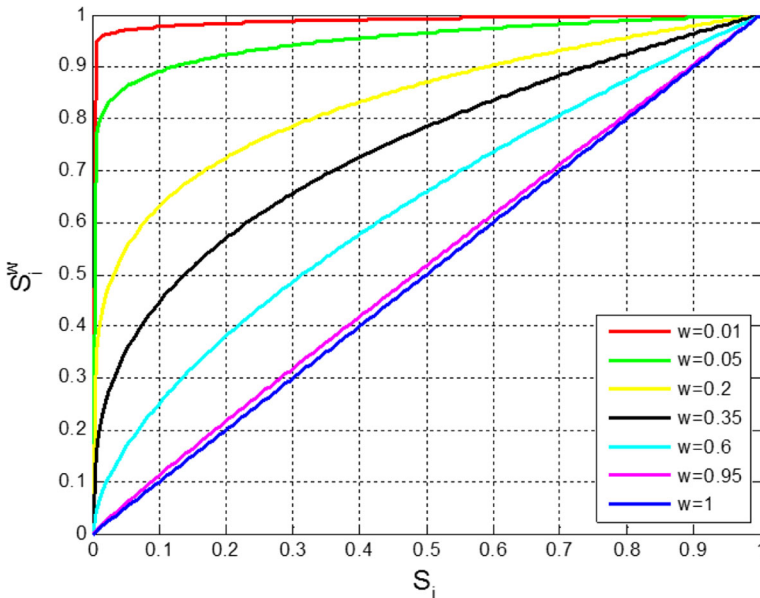


Fig. 3 Illustration of power-law transformation. S_i denotes the original saliency map, and S_i^w denotes the saliency map after power-law transformation

the set V , and E is a set of undirected edges connecting neighboring nodes. The graph cut solves a binary pixel labeling problem, and employs the following energy function:

$$E(L) = \sum_p D(L_p, S_F(p)) + \lambda \sum_{(p,q) \in N} \theta(L_p, L_q, I_d) \quad (11)$$

where $D(\cdot)$ is the data term, $\theta(\cdot)$ is the smoothness term, and the parameter λ is used to balance the two terms. L_p denotes the label of pixel p and $L_p = \{0, 1\}$ with $L_p = 1$ for object label and $L_p = 0$ for background label. The data term is defined as follows:

$$D(L_p, S_F(p)) = \begin{cases} S_F(p), & L_p = 0 \\ 1 - S_F(p), & L_p = 1 \end{cases} \quad (12)$$

The smoothness term models the spatial relationship between two adjacent pixels. Following the contrast function in [5], the smoothness term is defined in a similar way with the incorporation of spatial distance as follows:

$$\theta(L_p, L_q, I_d) = \psi(L_p, L_q) \cdot \frac{\exp(-\mu^{-1} \|I_d(p) - I_d(q)\|^2)}{\text{dist}(p, q)} \quad (13)$$

where $\psi(L_p, L_q) = 1$ if $L_p \neq L_q$, and 0 otherwise, $\text{dist}(p, q)$ denotes the Euclidean distance between p and q , and the decay factor $\mu = \alpha * E(\|I_d(p) - I_d(q)\|^2)$ equals to α times of $E(\cdot)$, the expectation over the whole image. Here α is set to 5 as suggested in [43]. The smoothness term introduces the penalty when adjacent pixels with similar colors are assigned with different labels.

The max-flow algorithm [6] is exploited to perform the graph cut and obtain the binary object mask S_M . The final saliency map is generated by combining the fused saliency map S_F and the binary object mask S_M for the better spatial coherence as follows:

$$S_{\text{final}} = \frac{S_M + S_F}{2} \quad (14)$$

3 Experimental results

3.1 Datasets

We evaluate the performance of our method over three widely used benchmark datasets, which contains a large number of images and have different biases [4].

MSRA10K Dataset Based on the Microsoft Research Asia (MSRA) Salient Object Database [30], MSRA10K dataset [49] was created by randomly selecting 10,000 images with consistent bounding box labeling in MSRA Salient Object Database and annotating salient objects with pixel-level accuracy. There is a large variation among images including natural scenes, animals, indoor, outdoor and human, etc. in this dataset.

ECSSD Dataset The dataset Extended Complex Scene Saliency Dataset (ECSSD) is proposed in [53] for overcoming the weakness of existing datasets, in which background structures are primarily simple. This dataset contains 1000 semantically meaningful but structurally complex natural images, in which objects and background show similar colors or/and objects are non-homogeneous, and is challenging for saliency detection.

PASCAL-S Dataset The PASCAL-S dataset [28] contains 850 images with multiple objects and cluttered background, and the pixel-level ground truth annotations. This dataset provides both fixations and salient object annotations. It is a challenging dataset with abnormally large or small salient objects in many images.

We randomly divided the MSRA10K dataset into three parts: training set, validation set and test set, with 3000, 3000 and 4000 images, respectively. The training set is used for training the quality assessment model, and the validation set is used for setting the parameters (T_{low} , T_{high}) in the quality assessment model and the parameter λ in the graph cut based refinement method. The test set includes the remaining 4000 images in the MSRA10K dataset and all images in both ECSSD dataset and PASCAL-S dataset, which are used for quantitative and qualitative comparison.

3.2 Evaluation metrics

In our experiments, four measures are adopted for evaluation, i.e., mean absolute error (MAE), precision-recall (PR) curve, F-measure, F_{β}^w -measure, Receiver Operating Characteristics (ROC) curves and Area Under ROC Curve (AUC) score [39]. As suggested in [4], different measures should be used for comprehensively evaluating the saliency detection performance.

1) Mean absolute error (MAE)

MAE calculates the average difference at pixel level between the saliency map S and the ground truth G , which are normalized into the range of [0, 1], and is defined as follows:

$$MAE = \frac{1}{W*H} \sum_{p=1}^{W*H} |S(p) - G(p)|. \quad (15)$$

MAE represents how close a saliency map is to the ground truth, and is more meaningful for applications such as salient object segmentation.

2) Precision-recall (PR) curve

For a saliency map S , we can convert it to a binary object mask B using the thresholding operation, and precision and recall are computed by comparing B with the ground truth G at pixel-level as follows:

$$Precision = \frac{\sum_p B(p) \cdot G(p)}{\sum_p B(p)}, \quad Recall = \frac{\sum_p B(p) \cdot G(p)}{\sum_p G(p)} \quad (16)$$

In order to obtain the binary mask B , 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 the x-axis as the recall value and the y-axis as the precision value.

3) F-measure

F-measure is defined as the weighted harmonic mean of precision and recall for a comprehensive evaluation, with the following form:

$$F_{\beta} = \frac{(1 + \beta^2) \text{Precision} \times \text{Recall}}{\beta^2 \text{Precision} + \text{Recall}} \tag{17}$$

where β^2 is set to 1 indicating the equal importance of precision and recall. Similarly as the PR curve, we plot the F-measure curve, in which the average F-measure is plotted against the threshold from 0 to 255. Besides, we also compute the average F-measure for the binary object masks, which are obtained by using the adaptive thresholding method [40], which is simple yet effective.

4) F_{β}^w -measure

F_{β}^w -measure is proposed by [39] for quantitative evaluation of saliency detection performance. It is an intuitive generalization of F-measure and offers a unified solution for evaluation of binary and non-binary maps. Here we compute F_{β}^w -measure for every image and then obtain the average F_{β}^w -measure on a given dataset for performance comparison.

5) Receiver Operating Characteristics (ROC) Curve

ROC curve presents a robust evaluation of saliency detection performance. It plots the true positive rate (TPR) against the false positive rate (FPR) by varying the threshold from 0 to 255. Specifically, TPR and FPR are defined as follows:

$$TPR = \frac{\sum_p B(p) \cdot G(p)}{\sum_p G(p)}, FPR = \frac{\sum_p B(p) \cdot \bar{G}(p)}{\sum_p \bar{G}(p)} \tag{18}$$

where \bar{G} denotes the complement of the ground truth G .

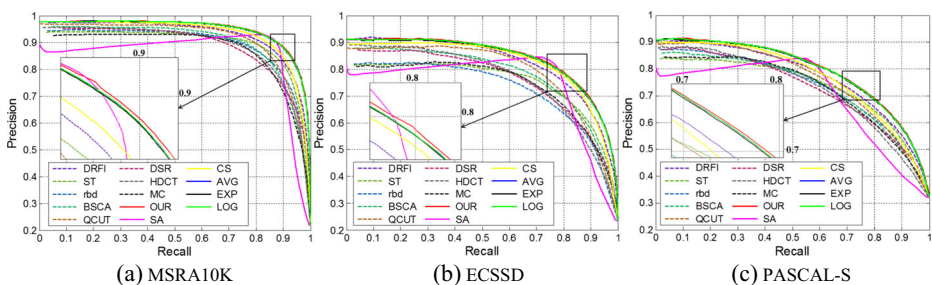


Fig. 4 Comparison of precision-recall curves among TOP8 saliency models, five fusion methods and our method on three public benchmark datasets

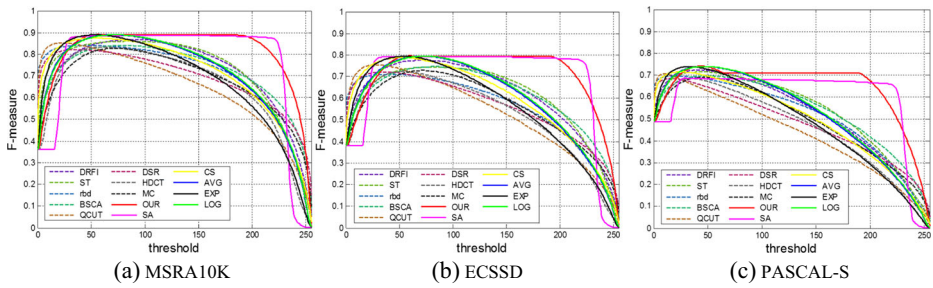


Fig. 5 Comparison of F-measure curves among TOP8 saliency models, five fusion methods and our method on three public benchmark datasets

6) Area Under ROC Curve (AUC) Score

AUC is the area under the ROC curve, which distills the ROC information into a single scalar. Here we compute AUC for each image and then obtain the average AUC on a given dataset for performance comparison.

3.3 Performance comparison

According to the benchmark [4], we selected the six state-of-the-art saliency models with the highest performance, i.e., DRFI [22], QCUT [2], RBD [57], ST [35], DSR [27] and MC [21], and also the recently proposed two saliency models, i.e., BSCA [42] and HDCT [23], to generate saliency maps, which are used for saliency fusion. For the eight saliency models, we used either the codes or the results provided by the authors. Based on the metrics including average F-measure, average F_{β} -measure, average MAE and average AUC, we evaluated the saliency maps generated using all the eight saliency models on the test set, and then obtain the overall ranking performance of all saliency models (in descending order): DRFI > ST > RBD > BSCA > QCUT > DSR > HDCT > MC. According to the ranking result of saliency models, we performed three groups of experiments by using different number of models: “TOP8” includes all the eight saliency models; “TOP4” includes DRFI, ST, RBD and BSCA; “TOP2” includes DRFI and ST.

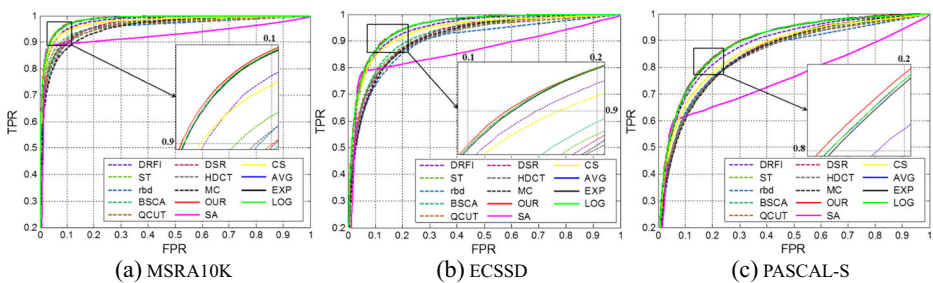


Fig. 6 Comparison of ROC curves among TOP8 saliency models, five fusion methods and our method on three public benchmark datasets

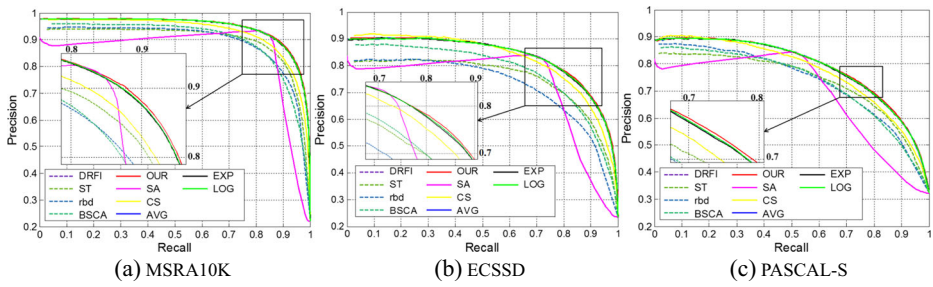


Fig. 7 Comparison of precision-recall curves among TOP4 saliency models, five fusion methods and our method on three public benchmark datasets

Our saliency fusion method is denoted as “OUR”, and is compared to the three fusion methods proposed in [3] with the following definition:

$$P(x|S_1(x), S_2(x), \dots, S_M(x)) \propto \frac{1}{Z} \sum_{i=1}^M \zeta(S_i(x)) \tag{19}$$

where Z is a constant, $\zeta(\cdot)$ denotes one of the three combination functions: 1) $\zeta(y) = y$; 2) $\zeta(y) = \exp(y)$; 3) $\zeta(y) = -1/\log(y)$. The three functions are denoted as “AVG”, “EXP” and “LOG”, respectively. Besides, our method is also compared to the other three fusion methods including SA [38], CS [37] and IS [56]. Note that IS [56] is only applicable to the group “TOP2”.

In the following, both quantitative and qualitative comparisons on saliency detection performance are reported based on the extensive experiments on the test sets from three datasets.

3.3.1 Quantitative comparison

Using the evaluation metrics in Section 3.2, Figs. 4, 5, 6, 7, 8, 9, 10, 11 and 12 show the PR curves, F-measure curves and ROC curves, and Tables 2, 3 and 4 show the average values of F-measure.

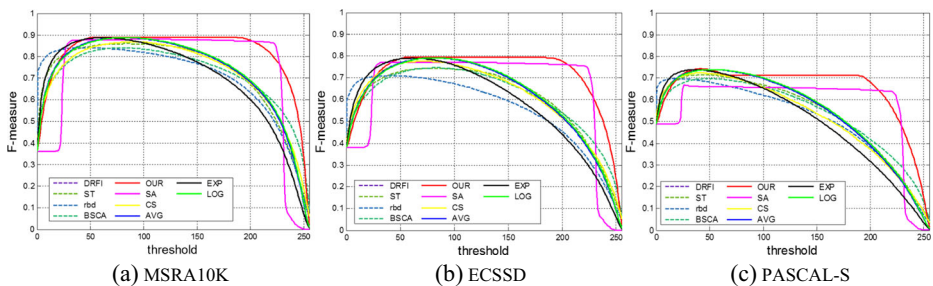


Fig. 8 Comparison of F-measure curves among TOP4 saliency models, five fusion methods and our method on three public benchmark datasets

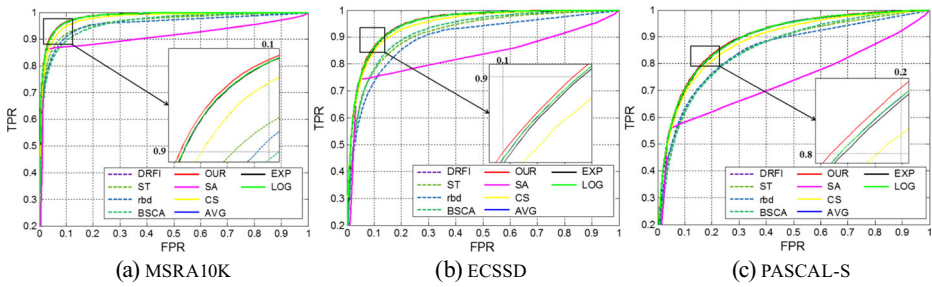


Fig. 9 Comparison of ROC curves among TOP4 saliency models, five fusion methods and our method on three public benchmark datasets

F_{β}^w –measure, MAE and AUC. Specifically, Figs. 4, 5 and 6 with Table 2, Figs. 7, 8 and 9 with Table 3, and Figs. 10, 11 and 12 with Table 4 present the performance comparison of fusion results using “TOP8”, “TOP4” and “TOP2”, respectively.

In terms of PR curves, we can see from Figs. 4, 7 and 10 that our fusion method consistently outperforms all the eight saliency models and other fusion methods on all the three datasets. In the view of the enlarged PR curves in the zone of high precision and high recall, which is critical for salient object detection and segmentation, it can be seen that our method achieves the higher performance than other fusion methods on all the three datasets. In terms of F-measure curves, we can see from Figs. 5, 8 and 11 that our fusion method consistently outperforms all the eight saliency models and five fusion methods including IS, CS, AVG, EXP and LOG on all the three datasets with large margins; as for the fusion method SA, it achieves the competitive performance compared with our method on MSRA10K dataset, but on the other two datasets, our method consistently outperforms SA. In terms of ROC curves, as shown in Figs. 6, 9 and 12, it can be found that our fusion method also consistently outperforms all the eight saliency models on all the three datasets; meanwhile, in the view of the enlarged ROC curves in the zone of high TPR and low FPR, it can be seen that our method performs better than other fusion methods on all the three datasets. Based on Figs. 4, 5, 6, 7, 8, 9, 10, 11 and 12, PR curves, F-measure curves

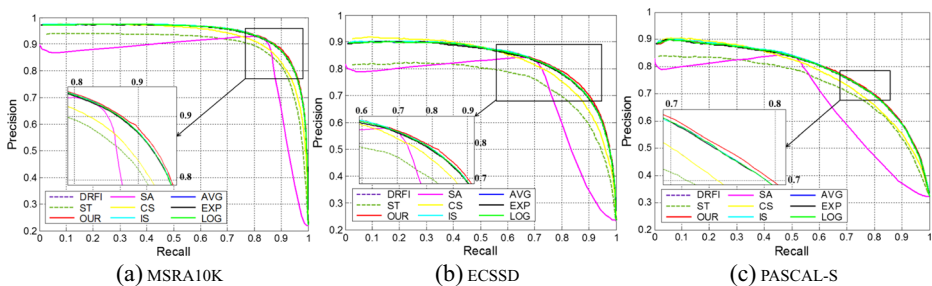


Fig. 10 Comparison of precision-recall curves among TOP2 saliency models, six fusion methods and our method on three public benchmark datasets

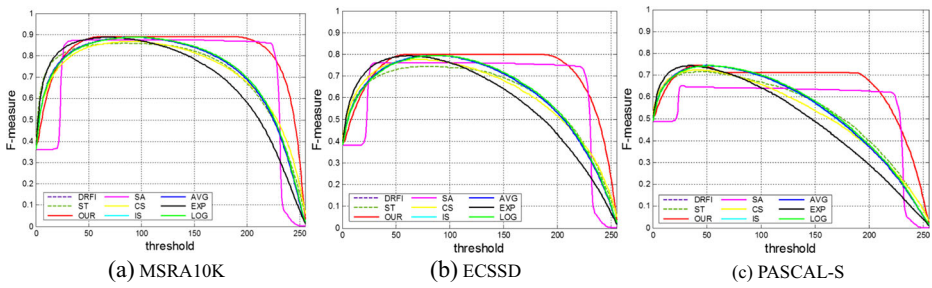


Fig. 11 Comparison of F-measure curves among TOP2 saliency models, six fusion methods and our method on three public benchmark datasets

and ROC curves show that our method achieves the highest performance, which demonstrates the effective improvement on saliency detection performance using our method.

Tables 2–4 show the quantitative performance comparison in terms of F-measure , F_{β}^w -measure , MAE and AUC. In each row of the three tables, the best performance is marked with red color, the second best performance is marked with green color, and the third one is marked with blue color. Table 2 presents the performance of the group “TOP8”. In terms of all the four metrics, it can be seen from Table 2 that our method consistently outperforms all the eight saliency models and other fusion methods on all the three datasets, except for MAE on ECSSD dataset, our method achieves the second performance. For the group “TOP4”, it can be seen from Table 3 that our method consistently achieves the best performance on all the three datasets in terms of all the four metrics. For the group “TOP2”, it can be seen from Table 4 that our method achieves the highest performance on all the three datasets in terms of F_{β}^w -measure, MAE and AUC, except for F-measure on MSRA10K and ECSSD datasets, where our method ranks the third, slightly lower than the best one. Overall, our method is more robust and achieves the better performance than other fusion methods with different combination of saliency models. This objectively shows the overall better quality of final saliency maps fused using our method, and also demonstrates the effectiveness of our method for improving saliency detection performance.

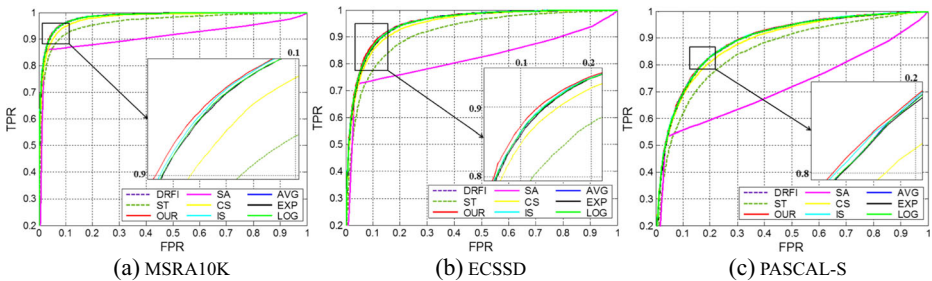


Fig. 12 Comparison of ROC curves among TOP2 saliency models, six fusion methods and our method on three public benchmark datasets

Table 2 Comparison of F-measure (F), F_{β}^w -measure (F_{β}^w), MAE and AUC among TOP8 saliency models, five fusion methods and our method on three public benchmark datasets

D	Metric	OUR	SA	CS	AVG	EXP	LOG	DRFI	ST	RBD	BSCA	QCUT	DSR	HDCT	MC
M	F	.878	.873	.847	.868	.849	.870	.860	.844	.796	.761	.783	.761	.811	.728
	F_{β}^w	.702	.602	.686	.655	.653	.655	.657	.664	.688	.639	.632	.663	.582	.577
	MAE	.097	.131	.105	.114	.116	.115	.116	.120	.106	.123	.113	.118	.142	.143
	AUC	.986	.925	.978	.984	.983	.985	.979	.967	.955	.959	.968	.956	.963	.957
E	F	.760	.759	.740	.754	.735	.755	.752	.706	.664	.660	.681	.634	.674	.643
	F_{β}^w	.572	.527	.551	.518	.514	.519	.542	.509	.512	.513	.490	.514	.452	.454
	MAE	.161	.170	.159	.174	.171	.175	.164	.192	.171	.182	.169	.170	.197	.203
	AUC	.955	.877	.944	.954	.953	.954	.947	.919	.900	.924	.934	.914	.915	.925
P	F	.664	.626	.591	.651	.622	.654	.639	.619	.616	.621	.536	.535	.604	.622
	F_{β}^w	.534	.471	.458	.473	.454	.474	.481	.476	.472	.471	.401	.439	.410	.422
	MAE	.231	.249	.249	.250	.253	.250	.248	.259	.247	.257	.259	.258	.274	.272
	AUC	.897	.783	.877	.896	.895	.896	.880	.856	.852	.856	.871	.851	.854	.864

For Table 2, in the first column, D denotes Dataset, M, E and P denote MSRA10K, ECSSD and PASCAL-S dataset, respectively; in the second column, F and F_{β}^w denote F-measure and F_{β}^w -measure, respectively

For Table 2, in each row, the best performance is marked with red color, the second best performance is marked with green color, and the third one is marked with blue color

3.3.2 Qualitative comparison

For a qualitative comparison, some saliency maps generated using “TOP8” saliency models, five fusion methods including AVG, EXP, LOG, SA and CS, and our method on the three datasets are shown in Figs. 13, 14 and 15 for a subjective comparison. In Figs. 13, 14 and 15, the example images contain heterogeneous objects (row 2 and 4 in Fig. 13, row 4 in Fig. 14, and row 2 and 4 in Fig. 15), low contrast between objects and background (row 1 in Fig. 13, row 1 in Fig. 14, and row 1 and 3 in Fig. 15), clutter background (row 2 in Fig. 13, all rows in Fig. 14, and row 2 and 4 in Fig. 15), large-scale salient object (row 3 in Fig. 14, and row 1 and 3 in Fig. 15) and multiple objects (row 1 and 4 in Fig. 13).

Table 3 Comparison of F-measure (F), F_{β}^w -measure (F_{β}^w), MAE and AUC among TOP4 saliency models, five fusion methods and our method on three public benchmark datasets

D	Metric	OUR	SA	CS	AVG	EXP	LOG	DRFI	ST	RBD	BSCA
M	F	.877	.863	.860	.875	.861	.876	.860	.844	.796	.761
	F_{β}^w	.721	.580	.657	.670	.669	.671	.657	.664	.688	.639
	MAE	.092	.139	.115	.111	.109	.111	.116	.120	.106	.123
	AUC	.985	.916	.976	.984	.984	.983	.979	.967	.955	.959
E	F	.757	.734	.751	.754	.744	.756	.752	.706	.664	.660
	F_{β}^w	.586	.497	.542	.532	.529	.533	.542	.509	.512	.512
	MAE	.156	.181	.164	.172	.168	.173	.164	.192	.171	.182
	AUC	.953	.843	.947	.952	.951	.952	.947	.919	.900	.924
P	F	.665	.600	.639	.662	.637	.664	.639	.619	.616	.621
	F_{β}^w	.550	.445	.481	.493	.474	.495	.481	.476	.472	.471
	MAE	.225	.259	.248	.246	.248	.245	.248	.259	.247	.257
	AUC	.894	.749	.880	.893	.892	.893	.880	.856	.852	.856

For Table 3, in the first column, D denotes Dataset, M, E and P denote MSRA10K, ECSSD and PASCAL-S dataset, respectively; in the second column, F and F_{β}^w denote F-measure and F_{β}^w -measure, respectively

For Table 3, in each row, the best performance is marked with red color, the second best performance is marked with green color, and the third one is marked with blue color

Table 4 Comparison of F-measure (F), F_{β}^w -measure (F_{β}^w), MAE and AUC among TOP2 saliency models, six fusion methods and our method on three public benchmark datasets

D	Metric	OUR	IS	SA	CS	AVG	EXP	LOG	DRFI	ST
M	F	.878	.880	.857	.860	.879	.869	.879	.860	.844
	F_{β}^w	.712	.669	.579	.657	.667	.666	.667	.657	.664
	MAE	.096	.113	.139	.115	.114	.113	.114	.116	.120
	AUC	.984	.983	.914	.979	.983	.983	.983	.979	.967
E	F	.770	.770	.721	.751	.769	.756	.769	.752	.706
	F_{β}^w	.586	.535	.490	.542	.533	.531	.533	.542	.509
	MAE	.156	.174	.182	.164	.175	.169	.175	.164	.192
	AUC	.953	.952	.835	.947	.951	.950	.952	.947	.919
P	F	.664	.669	.579	.639	.669	.639	.671	.639	.619
	F_{β}^w	.548	.492	.432	.481	.490	.467	.492	.481	.476
	MAE	.227	.248	.263	.248	.249	.252	.249	.248	.259
	AUC	.888	.886	.741	.880	.887	.886	.887	.880	.856

For Table 4, in the first column, D denotes Dataset, M, E and P denote MSRA10K, ECSSD and PASCAL-S dataset, respectively; in the second column, F and F_{β}^w denote F-measure and F_{β}^w -measure, respectively

For Table 4, in each row, the best performance is marked with red color, the second best performance is marked with green color, and the third one is marked with blue color

Furthermore, some of these example images are coupled with two or more issues mentioned above, such as row 2 in Fig. 13, row 3 in Fig. 14, and row 4 in Fig. 15, etc. All these examples are challenging images for saliency detection. Compared with all the eight saliency models and other fusion methods, we can see that our adaptive fusion method is able to suppress background regions and highlight salient object regions more completely and uniformly with well-defined boundaries. This demonstrates that our method can further elevate saliency detection performance by fusing saliency maps generated using state-of-the-art saliency models, especially for the complicated images with heterogeneous objects, low contrast, clutter background, large-scale objects and multiple objects.

3.3.3 Computational complexity

Our method is implemented using Matlab on a PC with an Intel Core i7 4.0 GHz CPU and 16 GB RAM. Excluding the time of generating saliency maps using different saliency models and considering all the eight saliency models to be used for fusion, the training time for quality assessment model is around 54 h, and the average testing time for an image with a resolution of

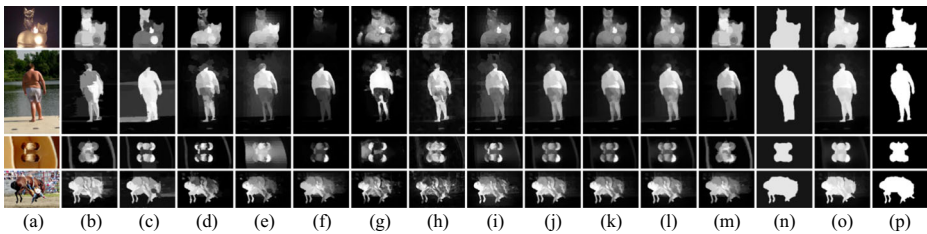


Fig. 13 Comparison of our method with eight state-of-the-art saliency models and five fusion methods on MSRA10K dataset. **a** Images; saliency maps generated using **b** DRFI [22], **c** ST [35], **d** RBD [57], **e** BSCA [42], **f** QCUT [2], **g** DSR [27], **h** HDCT [23], **i** MC [21], **j** AVG, **k** EXP, **l** LOG, **m** CS [37], **n** SA [38] and **o** OUR; **p** ground truths

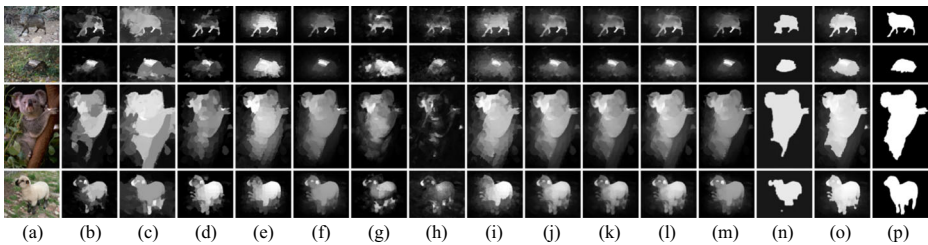


Fig. 14 Comparison of our method with eight state-of-the-art saliency models and five fusion methods on ECSSD dataset. **a** Images; saliency maps generated using **b** DRFI [22], **c** ST [35], **d** RBD [57], **e** BSCA [42], **f** QCUT [2], **g** DSR [27], **h** HDCT [23], **i** MC [21], **j** AVG, **k** EXP, **l** LOG, **m** CS [37], **n** SA [38] and **o** OUR; **p** ground truths

400×300 is 101.727 s including the extraction of quality features (Section 2.1 takes 101.454 s), estimation of quality scores (Section 2.2 takes 0.025 s), summation (Section 2.3 takes 0.037 s) and refinement (Section 2.4 takes 0.211 s). Although the current implementation of our method is time-consuming, we believe that the computational efficiency of our method can be substantially accelerated by using a C++ implementation and even a parallel GPU implementation.

4 Conclusion

In this paper, we have presented a general framework to adaptively fuse saliency maps generated using various saliency models via quality assessment and to generate a high-quality saliency map. First, in order to take full advantage of different kinds of quality features, we exploit multiple kernel boosting to formulate an effective quality assessment model. Second, for the purpose of utilizing quality scores reasonably and performing an adaptive fusion of saliency maps, quality scores are used as the weights with the power-law transformation for linear summation of different saliency maps. Third, a graph cut based refinement method is exploited to improve the spatial coherence of final saliency map. Experimental results show the performance improvement of the proposed saliency fusion method compared to the state-of-the-art saliency models and other fusion methods. The saliency maps obtained by our method can well suppress background regions and uniformly highlight salient objects, especially for complicated images. In our future work, based on the research results of this

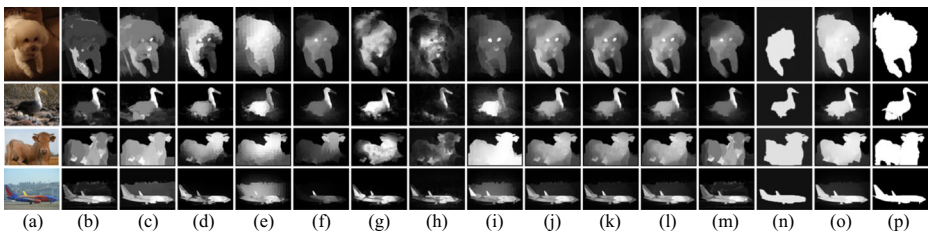


Fig. 15 Comparison of our method with eight state-of-the-art saliency models and five fusion methods on PASCAL-S dataset. **a** Images; saliency maps generated using **b** DRFI [22], **c** ST [35], **d** RBD [57], **e** BSCA [42], **f** QCUT [2], **g** DSR [27], **h** HDCT [23], **i** MC [21], **j** AVG, **k** EXP, **l** LOG, **m** CS [37], **n** SA [38] and **o** OUR; **p** ground truths

work, we will investigate the fusion of spatial saliency map and temporal saliency map for effective saliency detection in videos.

Acknowledgments This work was supported by National Natural Science Foundation of China under Grant No. 61471230 and No. 61171144, and by the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning.

References

1. Achanta R, Hemami SS, Estrada FJ, Süsstrunk S (2009) Frequency tuned salient region detection. Proc. of IEEE Conference on Computer Vision Pattern Recognition, In, pp. 1597–1604
2. Aytekin C, Kiranyaz S, Gabbouj M (2014) Automatic object segmentation by quantum cuts. Proc. of IEEE International Conference on Pattern Recognition, In, pp. 112–117
3. Borji A, Sihite D.N., Itti L (2012) Salient object detection: a benchmark. In: Proc. of European Conference on Computer Vision, pp. 414–429
4. Borji A, Cheng MM, Jiang H, Li J (2015) Salient object detection: a benchmark. IEEE Trans Image Process 24(12):5706–5722
5. Boykov Y, and Jolly MP (2001) Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In: Proc. of IEEE International Conference on Computer Vision, pp. 105–112
6. Boykov Y, Kolmogorov V (2004) An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Trans Pattern Anal Mach Intell 26(9):1124–1137
7. Cao Z, Qin T, Liu T.Y., Tsai M.F., Li H (2007) Learning to rank: from pairwise approach to listwise approach. In: Proc. of International Conference on Machine Learning, pp.129–136
8. Cheng MM, Zhang GX, Mitra NJ, Huang X, Hu SM (2011) Global contrast based salient region detection. Proc. of IEEE Conference on Computer Vision Pattern Recognition, In, pp. 409–416
9. Dollár P, Zitnick C (2013) Structured forests for fast edge detection. In: Proc. of IEEE International Conference on Computer Vision, pp. 1841–1848
10. Du H, Liu Z, Jiang J, Shen L (2013) Stretchability-aware block scaling for image retargeting. J Vis Commun Image Represent 24(4):499–508
11. Fang Y, Chen Z, Lin W, and Lin C.-W. (2012) Saliency detection in the compressed domain for adaptive image retargeting. IEEE Trans Image Process, 21(9): 3888–3901
12. Fang Y, Lin W, Lee B.-S., Lau CT, Chen Z, and Lin C.-W. (2012) Bottom-up saliency detection model based on human visual sensitivity and amplitude spectrum. IEEE Trans Multimed, 12(1): 187–198
13. Goferman S, Zelnik-Manor L, Tal A (2010) Context-aware saliency detection. Proc. of IEEE Conference on Computer Vision Pattern Recognition, In, pp. 2376–2383
14. Guo C, Zhang L (2010) A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. IEEE Trans Image Process 19(1):85–198
15. Harel J, Koch C, and Perona P (2007) Graph-based visual saliency. In Proc. of Advances in Neural Information Processing Systems, pp. 545–552
16. Hiremath P, Pujari J (2008) Content based image retrieval using color boosted salient points and shape features of an image. Int J Image Process 2(1):10–17
17. Hou X, Zhang L (2007) Saliency detection: a spectral residual approach. Proc. of IEEE Conference on Computer Vision Pattern Recognition, In, pp. 1–8
18. Hu P, Wang W, Zhang C, Lu K (2016) Detecting salient objects via color and texture compactness hypotheses. IEEE Trans Image Process 25(10):4653–4664
19. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20(11):1254–1259
20. Jia Y, and Han M (2013) Category-independent object-level saliency detection. In: Proc. of European Conference on Computer Vision, pp. 1761–1768
21. Jiang B, Zhang L, Lu H, Yang C, and Yang MH (2013) saliency detection via absorbing Markov chain. In: Proc. of IEEE International Conference on Computer Vision, pp. 1665–1672
22. Jiang H, Wang J, Yuan Z, Wu Y, Zheng N, and Li S (2013) salient object detection: a discriminative regional feature integration approach. In: Proc. of IEEE Conference on Computer Vision Pattern Recognition, pp. 2083–2090
23. Kim J, Han D, Tai YW, Kim J (2016) Salient region detection via high-dimensional color transform and local spatial support. IEEE Trans Image Process 25(1):9–23

24. Kolmogorov V, Zabini R (2004) What energy functions can be minimized via graph cuts? *IEEE Trans Pattern Anal Mach Intell* 26(2):147–159
25. Le Meur O, and Liu Z (2014) Saliency aggregation: does unity make strength? In *Proc. of Asia Conference on Computer Vision*, pp. 18–32
26. Li J, Tian Y, Duan L, Huang T (2013) Estimating visual saliency through single image optimization. *IEEE Signal Process Lett* 20(9):845–848
27. Li X, Lu H, Zhang L, Ruan X, and Yang MH (2013) Saliency detection via dense and sparse reconstruction. In: *Proc. of IEEE International Conference on Computer Vision*, pp. 2976–2983
28. Li Y, Hou X, Koch C, Rehg JM, and Yuille AL (2014) The secrets of salient object segmentation. In: *Proc. of IEEE Conference on Computer Vision Pattern Recognition*, pp. 280–287
29. Liu Z, Xue Y, Shen L, Zhang Z (2010) Nonparametric saliency detection using kernel density estimation. *Proc. of IEEE International Conference on Image Processing*, In, pp. 253–256
30. Liu T, Yuan Z, Sun J, Wang J, Zheng N, Tang X, Shum HY (2011) Learning to detect a salient object. *IEEE Trans Pattern Anal Mach Intell* 33(2):53–367
31. Liu Z, Xue Y, Yan H, Zhang Z (2011) Efficient saliency detection based on Gaussian models. *IET Image Process* 5(2):122–131
32. Liu Z, Shi R, Shen L, Xue Y, Ngan KN, Zhang Z (2012) Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut. *IEEE Trans Multimed* 14(4):1275–1289
33. Liu Z, Le Meur O, Luo S, Shen L (2013) Saliency detection using regional histograms. *Opt Lett* 38(5):700–702
34. Liu Z, Le Meur O, Luo S (2013) Superpixel-based saliency detection. *Proc. of IEEE International Workshop on Image and Audio Analysis for Multimedia Interactive Services*, In, pp. 1–4
35. Liu Z, Zou W, Le Meur O (2014) Saliency tree: a novel saliency detection framework. *IEEE Trans Image Process* 23(5):1937–1952
36. Liu Y, Liu S, Wang Z (2015) Multi-focus image fusion with dense SIFT. *Inf Fusion* 23:139–155
37. Mai L, Liu F (2014) Comparing salient object detection results without ground truth. In: *Proc. of European Conference on Computer Vision*, pp. 76–91
38. Mai L, Niu Y, and Liu F (2013) Saliency aggregation: a data-driven approach. In: *Proc. of IEEE Conference on Computer Vision Pattern Recognition*, pp. 1131–1138
39. Margolin R, Zelnik-Manor L, and Tal A (2014) How to evaluate foreground maps? In: *Proc. of IEEE Conference on Computer Vision Pattern Recognition*, pp. 248–255
40. Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* 9(1):62–66
41. Pan Z, You X, Chen H, Tao D, Pang B (2013) Generalization performance of magnitude-preserving semi-supervised ranking with graph-based regularization. *Inf Sci* 221:284–296
42. Qin Y, Lu H, Xu Y, Wang H (2015) Saliency detection via cellular automata. *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, In, pp. 110–119
43. Rother C, Kolmogorov V, Blake A (2004) GrabCut: interactive foreground extraction using iterated graph cuts. *ACM Trans Graph* 23(3):309–314
44. Shamir A, Avidan S (2009) Seam carving for media retargeting. *Commun ACM* 52(1):77–85
45. Sharma G, Jurie F, Schmid C (2012) Discriminative spatial saliency for image classification. *Proc. of IEEE Conference on Computer Vision Pattern Recognition*, In, pp. 3506–3513
46. Shen L, Liu Z, Zhang Z (2013) A novel H.264 rate control algorithm with consideration of visual attention. *Multimed Tools Appl* 63(3):709–727
47. Shi J, Malik J (1997) Normalized cuts and image segmentation. In: *Proc. of IEEE Conference on Computer Vision Pattern Recognition*, pp. 731–737
48. Shi R, Liu Z, Du H, Zhang X, Shen L (2012) Region diversity maximization for salient object detection. *IEEE Signal Process Lett* 19(4):215–218
49. THUR15000 dataset [Online] (2015) Available: <http://mmcheng.net/gsal/>
50. Wei Y, Wen F, Zhu W, Sun J (2012) Geodesic saliency using background priors. *Proc. of European Conference on Computer Vision*, In, pp. 29–42
51. Xie Y, Lu H, Yang MH (2013) Bayesian saliency via low and mid-level cues. *IEEE Trans Image Process* 22(5):1689–1698
52. Xu C, Tao D, Xu C (2015) Multi-view intact space learning. *IEEE Trans Pattern Anal Mach Intell* 37(12): 2531–2544
53. Yan Q, Xu L, Shi J, and Jia J (2013) Hierarchical saliency detection. In: *Proc. of IEEE Conference on Computer Vision Pattern Recognition*, pp. 1155–1162
54. Yang F, Lu H, and Chen YW (2010) Human tracking by multiple kernel boosting with locality affinity constraints. In *Proc. of Asia Conference on Computer Vision*, pp. 39–50.
55. Zhang L, Tong M, Marks T, Shan H, Cottrell G (2008) SUN: a Bayesian framework for saliency using natural statistics. *J Vis* 8(7):1–20

56. Zhou X, Liu Z, Sun G, Ye L, Wang X (2016) Improving saliency detection via multiple kernel boosting and adaptive fusion. *IEEE Signal Process Lett* 23(4):517–521
57. Zhu W, Liang S, Wei Y, and Sun J (2014) Saliency optimization from robust background detection. In: *Proc. of IEEE Conference on Computer Vision Pattern Recognition*, pp. 2814–2821.
58. Zou W, Kpalma K, Liu Z, Ronsin J (2013) Segmentation driven low-rank matrix recovery for saliency detection. *Proc. of British Machine Vision Conference*, In, pp. 1–13



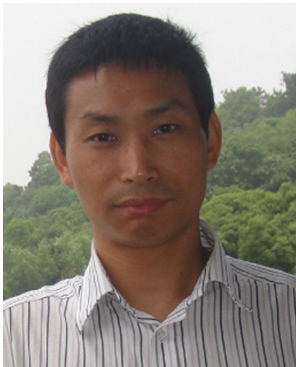
Xiaofei Zhou received the B.E. degree from Anhui Polytechnic University, Wuhu, China, in 2012, and the M.E. degree from Shanghai University, Shanghai, China, in 2015. He is currently pursuing the Ph.D. degree at the School of Communication and Information Engineering, Shanghai University. His research interests include image processing, pattern recognition and computer vision.



Zhi Liu received the B.E. and M.E. degrees from Tianjin University, Tianjin, China, and the Ph.D. degree from Institute of Image Processing and Pattern Recognition, Shanghai Jiaotong University, Shanghai, China, in 1999, 2002, and 2005, respectively. He is currently a Professor with the School of Communication and Information Engineering, Shanghai University, Shanghai, China. From Aug. 2012 to Aug. 2014, he was a Visiting Researcher with the SIROCCO Team, IRISA/INRIA-Rennes, France, with the support by EU FP7 Marie Curie Actions. He has published more than 120 refereed technical papers in international journals and conferences. His research interests include saliency models, image/video segmentation, image/video retargeting, video coding, and multimedia communication. He was a TPC member in ICME 2014, WIAMIS 2013, IWVP 2011, PCM 2010, ISPACS 2010, etc. He co-organized special sessions on visual attention, saliency models, and applications at WIAMIS 2013 and ICME 2014. He is an area editor of *Signal Processing: Image Communication* and served as a guest editor for the special issue on *Recent Advances in Saliency Models, Applications and Evaluations* in this journal. He is a senior member of IEEE.



Guangling Sun received the B.S. degree in electronic engineering from Northeast Forestry University, China, in 1996 and the M.E. and Ph.D. degrees in computer application technology from Harbin Institute of Technology, China, in 1998 and 2003, respectively. Since 2006, she has been with the faculty of the School of Communication and Information Engineering, Shanghai University, where she is currently an Associate Professor. She was with the University of Maryland, College Park as a visiting scholar from December 2013 to December 2014. Her research interests include saliency detection, face recognition, and image/video processing.



Xiangyang Wang received the PhD degree in pattern recognition and intelligent systems from the Institute of Image Processing and Pattern Recognition, Shanghai Jiaotong University, Shanghai, China, in 2006. He is currently an associate professor with the School of Communication and Information Engineering, Shanghai University, Shanghai, China. His research interests include computer vision, machine learning and medical image analysis.