

## Research Article

# Salient Region Detection via Feature Combination and Discriminative Classifier

Deming Kong,<sup>1</sup> Liangliang Duan,<sup>2</sup> Peiliang Wu,<sup>2</sup> and Wenji Yang<sup>2</sup>

<sup>1</sup>School of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China

<sup>2</sup>School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China

Correspondence should be addressed to Liangliang Duan; [hengxingdll9@163.com](mailto:hengxingdll9@163.com)

Received 18 July 2015; Revised 12 November 2015; Accepted 24 November 2015

Academic Editor: Jian Guo Zhou

Copyright © 2015 Deming Kong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We introduce a novel approach to detect salient regions of an image via feature combination and discriminative classifier. Our method, which is based on hierarchical image abstraction, uses the logistic regression approach to map the regional feature vector to a saliency score. Four saliency cues are used in our approach, including color contrast in a global context, center-boundary priors, spatially compact color distribution, and objectness, which is as an atomic feature of segmented region in the image. By mapping a four-dimensional regional feature to fifteen-dimensional feature vector, we can linearly separate the salient regions from the clustered background by finding an optimal linear combination of feature coefficients in the fifteen-dimensional feature space and finally fuse the saliency maps across multiple levels. Furthermore, we introduce the weighted salient image center into our saliency analysis task. Extensive experiments on two large benchmark datasets show that the proposed approach achieves the best performance over several state-of-the-art approaches.

## 1. Introduction

Humans have the ability to locate the most interesting region in a cluttered visual scene by selective visual attention. The task of computer vision is to simulate the human intelligence, and the related research has been carried out for many years. The study of human visual systems suggests that the saliency is related to rarity, uniqueness, and surprise of a scene. It has recently gained much attention [1–23], as it has been brought into the various applications, including image classification [24], object recognition [25], and content-aware image editing [26].

Existing saliency region detection methods can be roughly classified into two categories: bottom-up, data driven and top-down, task driven approaches. Bottom-up methods which utilize low-level image features, such as color, intensity, and texture, determine the contrast of image regions to their surroundings, while top-down methods make use of high-level knowledge about “interesting” object. The majority of most bottom-up models can be roughly divided into local and global schemes.

Inspired by the early work by Treisman and Gelade [27] and Koch and Ullman [28], Itti et al. [1] proposed highly influential biologically plausible saliency analysis method and they define image saliency using local center-surrounding operators across multiscale image features, including intensity, color, and orientation. Harel et al. [4] proposed a method to generate saliency map by nonlinearly combining local uniqueness maps from different feature channels. Ma and Zhang [29] propose a novel approach which directly computed center-surround color difference in a fixed neighborhood for each pixel and then utilize a fuzzy growth model to extract image salient region. They classify the saliency into three levels: attended view, attended areas, and attended points. Liu et al. [19] propose a set of novel features, including center-surround histogram, multiscale contrast, and color spatial distribution, which are unified in a CRF learning framework, to detect salient region in images.

Later on, many saliency models are proposed which exploit various types of image features in a global scope for saliency detection. Hou and Zhang [3] propose a spectral residual method that relies on frequency domain processing.

Zhai and Shah [2] define pixel-level saliency based on pixels contrast to all other pixels. To improve computational efficiency, they introduce the color histogram to analyze image saliency. Achanta et al. [5] propose a frequency tuned method which achieves globally consistent results by defining the saliency as the distance between the pixel and the overall mean image color. Cheng et al. [6] also utilize the color histogram and segmented region to analyze image saliency, which enable the assignment of comparable saliency values across similar image regions.

High-level priors have been used to analyze image saliency in recent years. Judd et al. [30] train SVM model using a combination of low-, middle-, and high-level image features, making their approaches potentially suitable for specific high-level computer vision tasks. The concept of center prior was considered in their approach. Shen and Wu [8] unify three higher level priors, including location prior, semantic prior, and color prior, to a low rank matrix recovery framework. Shape prior is proposed in Jiang et al. [7]; concavity context is utilized by [31]. Wei et al. [32] turn to background priors to analyze image saliency, and they assume that the image boundary is mostly background. Subsequently, many recent approaches use boundary prior to guide saliency detection, such as GMR [11], SO [17], PDE [33], AMC [15], and DSR [34], and using those methods can obtain state-of-the-art performance on several public available datasets.

Recent studies indicate that single saliency cue is far from being comprehensive. Some methods such as LC [2], FT [5], and HC [6] only use the contrast cue and the generated saliency maps are disappointing; the contrast cue sometimes produces high saliency values for background regions, especially for regions with complex structures. To alleviate the above problems, some approaches such as SF [18], PD [9] GC [12], PISA [21], PR [22], UFO [14], and HI [10] use multiple cues. Perazzi et al. [18] formulate saliency estimation using high-dimensional Gaussian filters by which region color and region position are, respectively, exploited to measure region uniqueness and distribution. Cheng et al. [12] and Tong et al. [22] also consider color contrast cue and color distribution cue when computing the saliency map. Margolin et al. [9] combine pattern distinctness, color uniqueness, and organization priors to generate saliency result. Shi et al. [21] present a generic framework for saliency detection via employing three terms, including color-based contrast term, structure-based contrast term, and spatial priors. Jiang et al. [14] propose a novel algorithm by integrating three saliency cues, namely, uniqueness, focusness, and objectness. Yan et al. [10] propose a multilayer approach to analyze image saliency. To determine the single-layer saliency cue, they exploit two useful saliency cues, including local contrast and location heuristic, and then a hierarchical inference framework is used to generate the final saliency map. The above-mentioned algorithms compute saliency maps from various cues and heuristically combine them to get the final results.

These methods can generate ideal saliency map when dealing with simple images. When computing the image with complex background, some methods such as [9, 12, 18] can

only highlight part of salient object. Though methods such as [10, 14, 22] can highlight the entire object uniformly, the background may be highlighted too. Thus, to differentiate real salient regions from high-contrast parts, more saliency cues including low-level feature and high-level priors need to be integrated. To the best of our knowledge, there are few works that model the interaction between different saliency cues. Inspired by the work [10, 23], we propose a feature combination strategy which can capture the interaction between different cues. Our main contributions lie in three aspects. Firstly, we introduce feature combination to model the interaction between different cues, which is different from most existing methods that generate saliency maps heuristically from various cues. Secondly, we formulate salient estimation as a classification problem and learn a logistic classifier that can directly map a fifteen-feature vector to a saliency value. Thirdly, the use of smoothing and weighted salient image center can further improve the detection performance. The experimental results show that our method can generate reasonable saliency map, even though the image contains complex background and the salient object has similar color to background.

The framework of the approach is presented in Figure 1. First, our approach includes four main parts. The first one is hierarchical image abstraction, which segments the image to homogenous regions across several layers by using the efficient graph-based image segmentation [35]. Second, four saliency cues, including color contrast in a global context, center-boundary priors, spatially compact color distribution, and objectness, are used as an atomic regional feature. Then we map a four-dimensional regional feature to fifteen-dimensional feature vector which can capture the interaction between different features. Third, a logistic regression classifier is trained for mapping a fifteen-dimensional feature vector to a saliency value. Finally, we combine the saliency map at different layers to obtain our saliency map. Figure 2 shows samples of saliency maps generated by state-of-the-art methods and by ours.

The remainder of this paper is organized as follows. The proposed model is introduced in Section 2. Section 3 presents experiments and results. This paper is summarized in Section 4.

## 2. The Proposed Approach

Our method can be divided into four main stages: hierarchical image abstraction, regional feature generation, training a logistic regression classifier, and multilayer saliency map integration and reinforcement. In the following, we describe the details of the proposed approach.

*2.1. Hierarchical Image Abstraction.* Given an image  $I$ , hierarchical image abstraction can be described as  $H = \{H_1, H_2, \dots, H_M\}$ , where  $M$  is the number of image pyramids and  $H_m$  is the abstraction result of the  $m$ th layer of image pyramid, including  $n_m$  regions. We use the image-level hierarchical image abstraction which is different from regional-level image abstraction [10, 23], as is shown in Figure 1(b). The result of segmentation in first layer of image is described

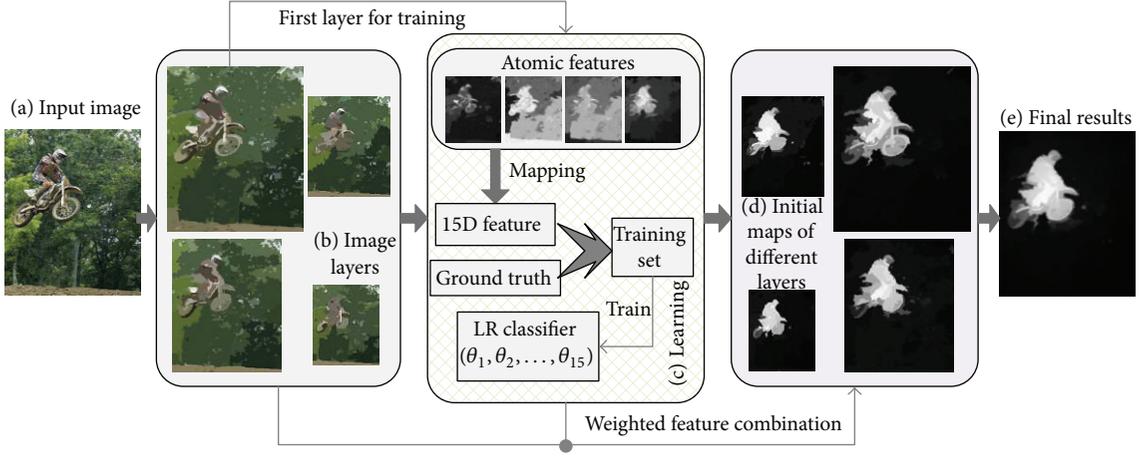


FIGURE 1: An overview of our weighted feature combination framework. We extract four image layers from input and then train a logistic regression classifier by using the four atomic features. Initial saliency maps of the four layers can be obtained by weighted feature combination. Finally, we fuse saliency maps of different layers to obtain the final saliency map.

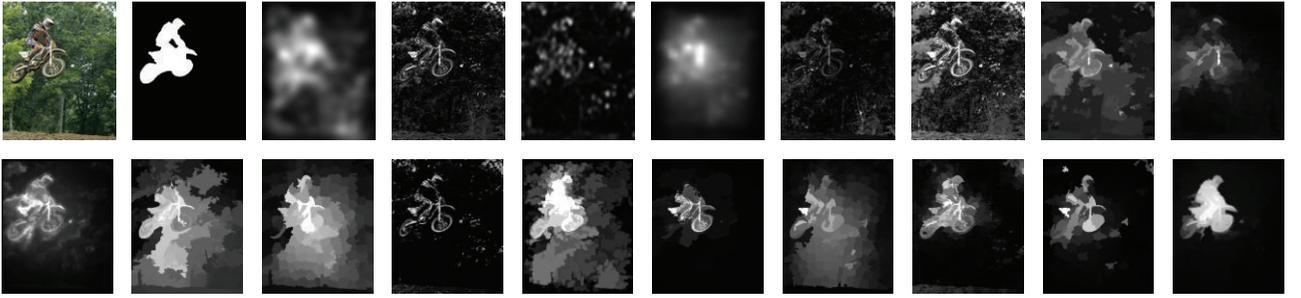


FIGURE 2: Saliency maps (from left to right, top to down): input, G-Truth, IT [1], LC [2], SR [3], GB [4], FT [5], HC [6], RC [6], LR [8], PD [9], HI [10], GMR [11], GC [12], BMS [13], UFO [14], AMC [15], HDCT [16], SO [17], and ours.

as  $H_1 = \{R_1^1, R_2^1, \dots, R_{n_1}^1\}$ , and the segmentation result of other layers can be described in a similar way. Each superpixel  $R_i^m$  is represented by mean color  $f_i^m$  (in CIELab) and a spatial position  $c_i^m$  ( $x$ -coordinate and  $y$ -coordinate) which are defined as  $f_i^m = \sum_{p \in R_i^m} V_{p(c)} / U_i^m$ ,  $c_i^m = \sum_{p \in R_i^m} V_{p(l)} / U_i^m$ , where  $p$  stands for a pixel in the region  $R_i^m$ , while  $V_{p(c)}$  represents the color vector of pixel  $p$  and  $V_{p(l)}$  represents the coordinate vector of pixel  $p$ .  $U_i^m$  is the number of pixels in the segmented region  $R_i^m$ .

## 2.2. Regional Feature Generation

(1) *Color Contrast Cue*. Given the result of segmentation in  $m$  layer of image pyramid which is described as  $H_m = \{R_1^m, R_2^m, \dots, R_{n_m}^m\}$ , the color contrast of a region  $S_{i(1)}^m$  can be formulated as follows:

$$S_{i(1)}^m = \sum_{k=1, k \neq i}^{n_m} \psi(R_i^m, R_k^m) \cdot \tau(R_i^m, R_k^m) \cdot U_k^m, \quad (1)$$

where  $\psi(R_i^m, R_k^m)$  is the smooth term which considers the distance between two regions and  $\tau(R_i^m, R_k^m)$  is the color distance between  $R_i^m$  and  $R_k^m$ .

(2) *Color Distribution Cue*. Inspired by Liu et al. [19], we use the nonoverlapped region as computing unit to compute region color distribution. First, all region colors are represented by Gaussian Mixture Models (GMMs)  $\{\pi_q, \mu_q, \Sigma_q\}_{q=1}^5$ , where  $\{\pi_q, \mu_q, \Sigma_q\}$  is weight, the mean color, and the covariance matrix of the  $q$ th Gaussian component. The probability of a region belonging to the  $q$ th component is given by

$$P(q | R_i^m) = \frac{\pi_q N(R_i^m | \mu_q, \Sigma_q)}{\sum_q \pi_q N(R_i^m | \mu_q, \Sigma_q)}. \quad (2)$$

The number of Gaussian is set to 5 in the subsequent experiment. We exploit  $k$ -means algorithm to initialize the parameters of GMMs and EM algorithm to train

the GMMs. Referring to [19], the horizontal spatial variance of the  $q$ th clustered component of GMMs is defined as

$$D_x(q) = \frac{1}{\Lambda} \sum_{i=1}^{n_m} P(q | R_i^m) \cdot \left\| c_{i(x)}^m - \gamma_{(x)}(q) \right\|^2, \quad (3)$$

$$\gamma_{(x)}(q) = \frac{1}{\Lambda} \sum_{i=1}^{n_m} P(q | R_i^m) \cdot c_{i(x)}^m,$$

where  $\Lambda_q = \sum_{i=1}^{n_m} P(q | R_i^m)$  and  $c_{i(x)}^m$  is the  $x$ -coordinate of  $R_i^m$ . The vertical spatial variance  $D_y(q)$  can be defined in the same way. Different from Liu et al. [19] that use the two variances to compute saliency cue, we only use the horizontal spatial variance. The color distribution of region  $R_i^m$  can be defined as

$$S_{i(2)}^m = \frac{1}{\exp\left(\sum_{q=1}^5 P(q | R_i^m) \cdot D_x(q)\right)}. \quad (4)$$

(3) *Center-Boundary Prior Cue*. Location is an important factor in saliency detection. Center and boundary are two priors which are widely used in previous saliency detection methods. After considering above two priors, our center-boundary heuristic is thus defined as

$$S_{i(3)}^m = S_{i(cp)}^m \cdot S_{i(bp)}^m, \quad (5)$$

where  $S_{i(cp)}^m$  is the center prior term which measures the distance between the region  $R_i^m$  and the image center and it is defined as  $S_{i(cp)}^m = 1/(k + \sqrt{\|c_i^m - \mu\|^2/2})$ , where  $\mu$  is the center of image and it is set to (0.5, 0.5); the parameter  $k$  controls the sensitivity of the center prior and it is set to 1 in the experiment.  $S_{i(bp)}^m$  is the boundary prior term which measures the color distance between the region  $R_i^m$  and the image boundary. Inspired by the approach proposed by Yang et al. [11], we define background feature of region  $R_i^m$  as

$$S_{i(bp)}^m = \log\left(\frac{B_{i(\text{top})}^m}{N_t} \cdot \frac{B_{i(\text{bottom})}^m}{N_b} \cdot \frac{B_{i(\text{left})}^m}{N_l} \cdot \frac{B_{i(\text{right})}^m}{N_r}\right), \quad (6)$$

where  $B_{i(\text{top})}^m$  is the sum of distances from region  $R_i^m$  to the top boundary of image which is different from Yang et al. [11] and  $N_t$  is the number of regions that intersect with the top image boundary. We use a simple approach to compute  $B_{i(\text{top})}^m$  which is given by  $B_{i(\text{top})}^m = \sum_{k=1}^{N_t} \|f_i^m - f_k^m\|$ . And  $B_{i(\text{bottom})}^m$ ,  $B_{i(\text{left})}^m$ , and  $B_{i(\text{right})}^m$  can be computed in a similar way.

(4) *Objectness Cue*. Recently, a generic objectness measure is proposed to quantify how likely it is for an image window to contain an object of any class. The measure is based on low-level image cues. As our goal is to obtain a saliency map for the whole image, we should transfer the objectness value

from the bounding boxes to the pixel level first and then we can obtain region level objectness measure. For more details, please refer to UFO [14]. For each region, we can get its region-level objectness  $S_{i(4)}^m = \sum_{p \in R_i^m} V_{p(o)} / U_i^m$ , where  $V_{p(o)}$  is the objectness value for pixel  $p$ .

(5) *Cues Smoothing*. Thus, we get four saliency cues and they are normalized to range [0, 1] using minimum–maximum normalization. Although we can efficiently compute four saliency cues, there exist at least two problems. Firstly, some regions with similar property will have very different saliency value and, secondly, some adjacent regions will be assigned to very different saliency value. To reduce noisy saliency results caused by above-mentioned issues, we use two smoothing procedures to refine the saliency value for each region.

*K-Means Clustering Based Smoothing*. Given the results of segmentation in  $m$  layer of image pyramid which is described as  $H_m = \{R_1^m, R_2^m, \dots, R_{n_m}^m\}$ , we first exploit  $k$ -means clustering algorithm to divide the segmented regions into different clusters in each layer. Referring to [36], we can then define an object function, sometimes called a distortion measure, given by

$$J = \sum_{i=1}^{n_m} \sum_{k=1}^K r_{ik} \|f_i^m - \mu_k\|, \quad (7)$$

which we can easily solve for  $\mu_k$  to give

$$\mu_k = \frac{\sum_{i=1}^{n_m} r_{ik} \cdot f_i^m}{\sum_{i=1}^{n_m} r_{ik}}. \quad (8)$$

$r_{ik}$  is the binary indicator variables; if a region is assigned to cluster  $k$ , then  $r_{ik} = 1$ , and  $r_{ij} = 0$  for  $j \neq k$ . This is known as the 1-of- $K$  coding scheme. The two phases of reassigning data points to clusters and recomputing the cluster means are repeated in turn until there is no further change in the assignments. Then we get the number for each cluster  $cl(k) = \sum_{i=1}^{n_m} r_{ik}$ . We replace the saliency values of each region by the weighted average of the saliency values of the same cluster (measured by  $L^*a^*b^*$  distance). The saliency value of each region  $R_i^m$  can be refined by

$$S_{i(\text{tmp})}^m = \alpha \cdot S_{i(\text{tmp})}^m + (1 - \alpha) \sum_{j=1, j \neq i}^{cl(k)} \frac{d(R_i^m, R_j^m)}{T} S_{j(\text{tmp})}^m, \quad (9)$$

where tmp can be replaced by 1, 2, 3, and 4. The parameter  $\alpha$  controls the importance of color space smoothing term. In our experiment, we set the parameter  $\alpha = 0.5$ . Consider  $T = \sum_{j=1, j \neq i}^{cl(k)} d(R_i^m, R_j^m)$  is the sum of distances between region  $R_i^m$  and the other region in cluster  $k$ . Consider  $d(R_i^m, R_j^m) = \exp(-1 \times \|f_i^m - f_j^m\|/2\sigma^2)$  is the color distance between region  $R_i^m$  and region  $R_j^m$ ; we set  $\sigma^2 = 10$  in the experiments.

*Spatial Based Smoothing*. We assume that two adjacent regions are likely to have similar saliency value. Therefore, we

propose a spatial based approach to refine saliency between adjacent regions and the procedure is very similar to color space smoothing. We replace the saliency values of each region by the weighted average of the saliency values of its neighbors.

(6) *Regional Feature*. After completing the above steps, we can get four atomic features ( $S_{i(1)}^m, S_{i(2)}^m, S_{i(3)}^m$ , and  $S_{i(4)}^m$ ) for each segmented region, including color contrast, center-boundary prior, color distribution, and objectness. In order to capture the interaction between the four different features, a novel feature is generated by mapping a four-dimensional regional feature to fifteen-dimensional feature vector. There are four kinds of combinations: single term, double term, triple term, and quadruple term. For single term, we use the four atomic features ( $x_1-x_4$  in the vector). For double term, there are six elements which are combination of any two atomic features ( $x_5-x_{10}$ ) ( $S_{i(1)}^m \cdot S_{i(2)}^m, S_{i(1)}^m \cdot S_{i(3)}^m, S_{i(1)}^m \cdot S_{i(4)}^m, S_{i(2)}^m \cdot S_{i(3)}^m, S_{i(2)}^m \cdot S_{i(4)}^m$ , and  $S_{i(3)}^m \cdot S_{i(4)}^m$ ). For triple term, the new features are the combination of three different atomic features which are formed to  $x_{11}-x_{14}$  in the new vector ( $S_{i(1)}^m \cdot S_{i(2)}^m \cdot S_{i(3)}^m, S_{i(1)}^m \cdot S_{i(2)}^m \cdot S_{i(4)}^m, S_{i(1)}^m \cdot S_{i(3)}^m \cdot S_{i(4)}^m$ , and  $S_{i(2)}^m \cdot S_{i(3)}^m \cdot S_{i(4)}^m$ ). For the quadruple term, there is only one element which is the product of the four atomic features. The last feature  $x_{15}$  is  $S_{i(1)}^m \cdot S_{i(2)}^m \cdot S_{i(3)}^m \cdot S_{i(4)}^m$ . Finally, we can get a novel fifteen-dimensional feature vector.

2.3. *Learning Framework for Saliency Estimation*. The logistic function is useful because it can take an input with any value from negative to positive infinity, whereas the output always takes values between zero and one [37]. We take full advantage of this property; thus our saliency estimation can be formulated as a probability framework. Let us assume that

$$\begin{aligned} P(y = 1 | x; \theta) &= h_\theta(x), \\ P(y = 0 | x; \theta) &= 1 - h_\theta(x), \end{aligned} \quad (10)$$

where  $h_\theta(x) = g(\theta^T x) = 1/(1 + e^{-\theta^T x})$  is our hypotheses. Consider  $g(z) = 1/(1 + e^{-z})$  is called the logistic function or the sigmoid function. Notice that  $g(z)$  tends towards 1 as  $z \rightarrow +\infty$ , and  $g(z)$  tends towards 0 as  $z \rightarrow -\infty$ . Hence, our hypotheses is always bounded between 0 and 1 and higher value indicates that the region is likely to belong to a salient object.

The parameter  $\theta$  is what we want to learn from the data. We use the first layers of image pyramid for training, given the result of segmentation in 1st layer of image pyramid which is described as  $H_1 = \{R_1^1, R_2^1, \dots, R_{n_1}^1\}$ . A segmented region is considered to be positive if the number of the pixels belonging to the salient object exceeds 90% of the number of the pixels in the region and its saliency value is set to 1. On the contrary, a region is considered to be negative if the number of the pixels belonging to the salient object is under 10% of the number of the pixels in the region and its saliency value is set to 0. As aforementioned, each segmented region is described by a fifteen-dimensional vector  $\mathbf{x}$ . We learn a logistic regression classifier  $f$  from the training data  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$  and the saliency value  $\mathbf{Y} = \{y_1, y_2, \dots, y_N\}$ . Once the parameter  $\theta$

is obtained, we can quickly perform the saliency estimation using (10).

#### 2.4. Multilayer Saliency Map Integration and Reinforcement

2.4.1. *Multilayer Saliency Map Integration*. We combine image pyramid which is the multiscale representation of image to suppress background region. Similar to [1], the saliency map is obtained by adjusting the saliency map to the same scale and point-by-point addition. The fusion strategy is given by  $S_{\text{Fusion}}(I) = \bigoplus_{i=1}^M \text{sal}(I^m)$ , where  $I$  is the input image,  $I^m$  is the  $m$ th layer of image pyramid, and  $\text{sal}(I^m)$  is the saliency detection result of the  $m$ th layer of image pyramid.

2.4.2. *Reinforcement of Salient Region*. Salient object is always distributed in local region of the image, while background has a high degree of dispersion. To use such property, we introduced the weighted salient image center into our saliency estimation, and the newly defined salient center is defined as

$$(SX, SY) = \frac{\sum_{i=1}^N V_{p_i(l)} \cdot S_{\text{Fusion}}(p_i)}{\sum_{i=1}^N S_{\text{Fusion}}(p_i)}, \quad (11)$$

where  $N$  is the number of pixels in image and  $p_i$  is the  $i$ th pixel. Hence, the final pixel-level saliency can be defined as

$$S_{\text{Final}}(p \in I) = S_{\text{Fusion}}(p) \cdot \exp\left(-\frac{D_p^2}{\sigma^2}\right), \quad (12)$$

where  $p$  is a pixel in an image,  $D_p$  is the Euclidean distance between the pixel  $p$  and the weighted salient image center, and the parameter  $\sigma^2$  is the smooth term which controls the strength of spatial weight; we set the parameter  $\sigma^2$  to 0.4.

### 3. Experiments and Results

To validate our proposed approach, we have performed experiments on two publicly available datasets. (1) The first one is the MSRA dataset [19], which contains 5000 images with pixel-level grounds truth. We used the same training set, validating set, and test set as the paper of Jiang et. al [23]. The training set contains 2500 images, the validating set contains 500 images, and the testing set contains 2000 images. (2) The second one is the ECSSD dataset, which contains 1000 images with multiple objects and makes the detection tasks much more challenging and the pixel-level ground truth for image is provided by Yan et al. [10]. On the two datasets, we compare our method with 18 state-of-the-art saliency detection methods, including IT [1], LC [2], SR [3], GB [4], FT [5], HC [6], RC [6], CB [7], LR [8], PD [9], HI [10], GMR [11], GC [12], BMS [13], UFO [14], AMC [15], HDCT [16], and SO [17]. Three parameters of graph-based image segmentation [35] are used in the algorithm, and we set conservative parameters, sigma = 0.5 and min = 10, for all the images. The third parameter  $K$  is set to different

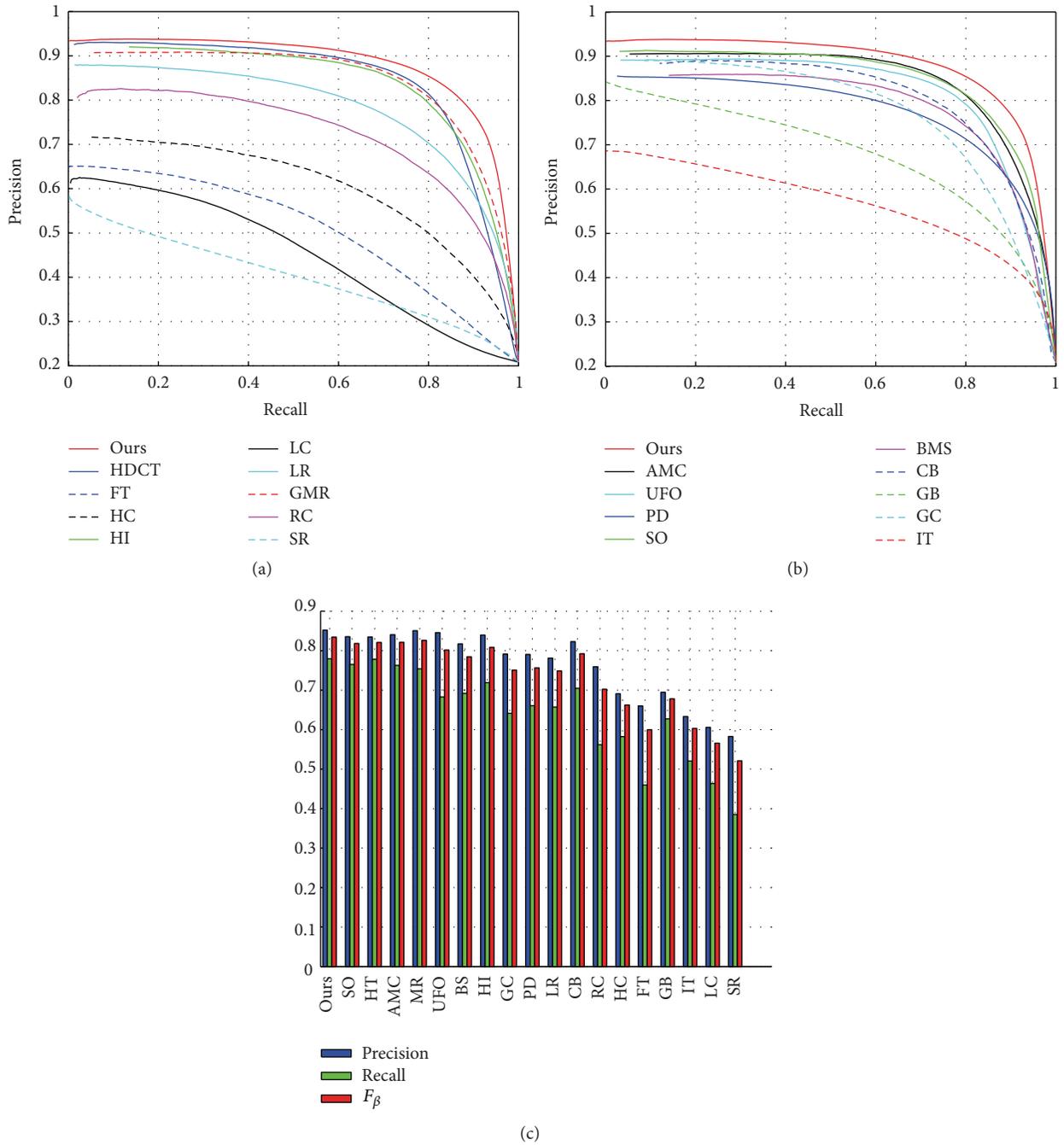


FIGURE 3: Experimental results on the MSRA dataset. (a) and (b) are precision and recall curves of all approaches which are obtained using fixed threshold. The histogram (c) (precision, recall, and  $F_\beta$ ) is obtained using adaptive thresholding.

value according to image layers, We set  $K = 100$  for the first layer,  $K = 75$  for the second layer, and  $K = 50$  for other layers. To evaluate these methods, we either run our own implementations or exploit the results from the original authors.

**3.1. Evaluation Methods.** Following [5, 6, 8], we evaluate the performance of our method measuring its precision and recall rate. Precision measures the percentage of salient pixels

correctly assigned, while recall measures the percentage of salient object detected. In order to study the performance of saliency detection approaches, we use two kinds of objective comparison measures in previous studies.

Firstly, the saliency map is segmented by a fixed threshold. Given a threshold  $T \in [0, 255]$ , the pixels whose saliency values are lower than  $T$  are marked as background; otherwise the pixels are marked foreground. When varying from 0 to 255, it will produce a sequence of

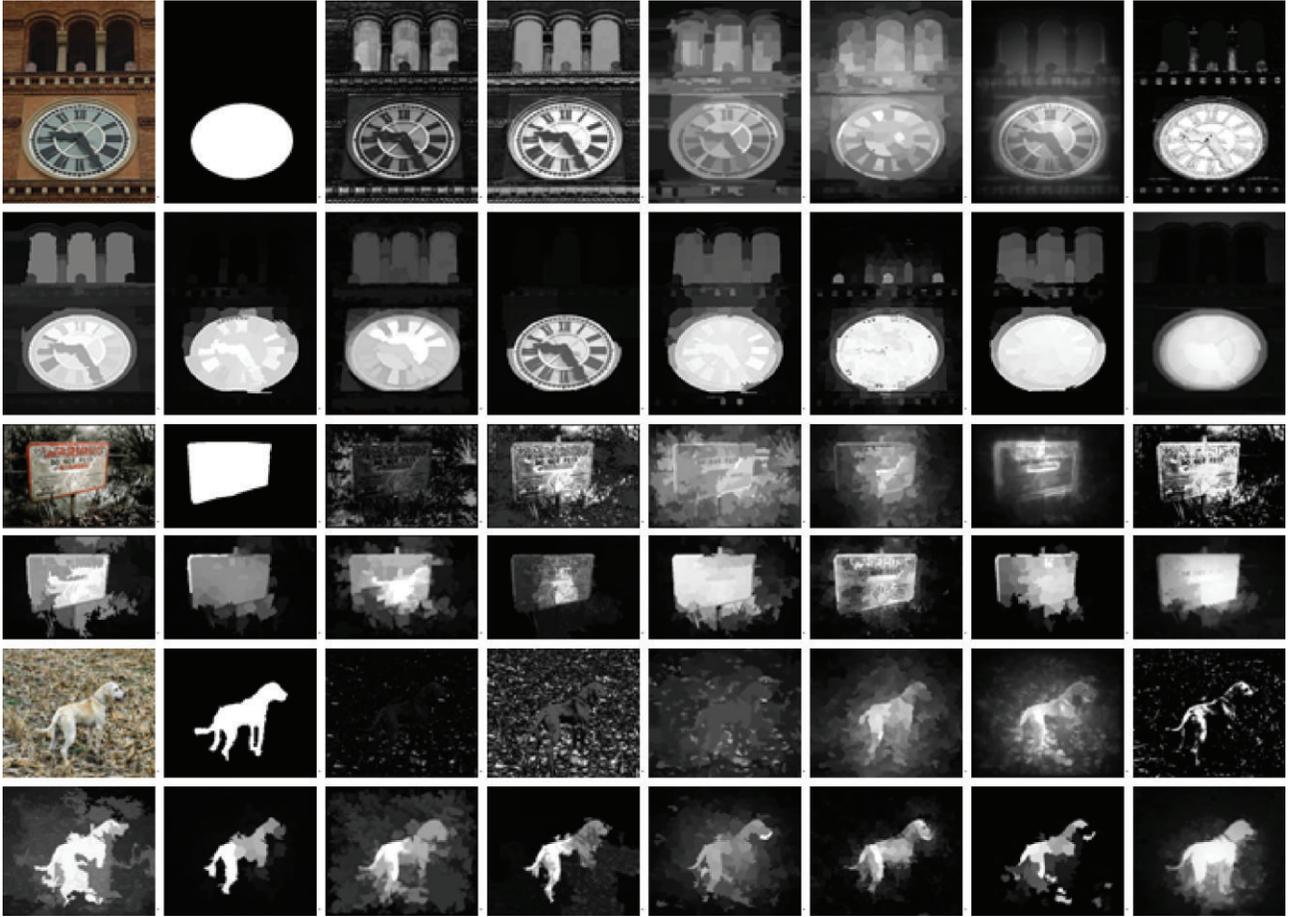


FIGURE 4: Saliency maps (from left to right, top to down): input, G-Truth, FT [5], HC [6], RC [6], LR [8], PD [9], GC [12], HI [10], GMR [11], BMS [13], UFO [14], AMC [15], HDCT [16], SO [17], and ours.

precision-recall pairs, and a precision-recall curve can be obtained.

Secondly, we follow [5, 6, 8] to segment a saliency map by adaptive thresholding. The image is first segmented by mean-shift clustering algorithm. And then we calculate the average saliency value of each nonoverlapped region; an overall mean saliency value over the entire saliency map is calculated as well. The mean-shift segments whose saliency value is larger than twice of the overall mean saliency value will be marked as foreground, and the threshold is defined as

$$T_a = \frac{2}{W \times H} \sum_{x=1}^W \sum_{y=1}^H S_{\text{Final}}(x, y), \quad (13)$$

where  $W$  and  $H$  are the width and height of the saliency map, respectively. In many applications, high precision and better recall rate are both required. In addition to precision and recall, we thus estimate  $F_\beta$ , which is defined as

$$F_\beta = \frac{(1 + \beta^2) \cdot \text{Precision} \cdot \text{Recall}}{\beta^2 \cdot \text{Precision} + \text{Recall}}, \quad (14)$$

where we set  $\beta^2 = 0.3$  as is suggested in [5, 6, 18].

**3.2. Performance on MSRA Dataset.** We report both quantitative and qualitative comparisons of our method with 18 state-of-the-art saliency detection approaches on the MSRA dataset.

*Quantitative Comparison.* Figures 3(a) and 3(b) show the precision-recall curves of all the algorithms on the MSRA-5000 dataset. As observed from Figure 3, the curve of our method is consistently higher than others on this dataset. Besides, we compare the performance of various methods using adaptive thresholding. Each value of our  $P-R-F$  (0.8524, 0.7794, and 0.8343) ranks first among the 18 state-of-the-art methods.

*Qualitative Comparison.* The visual comparison is given in Figure 4. To save the space, we only consider the most recent thirteen models: FT [5], HC [6], RC [6], LR [8], PD [9], GC [12], HI [10], GMR [11], BMS [13], UFO [14], AMC [15], HDCT [16], and SO [17]. Our method produces the best detection results on these images. It is also worth pointing out that our method can handle the challenging cases where the background is very extremely cluttered.

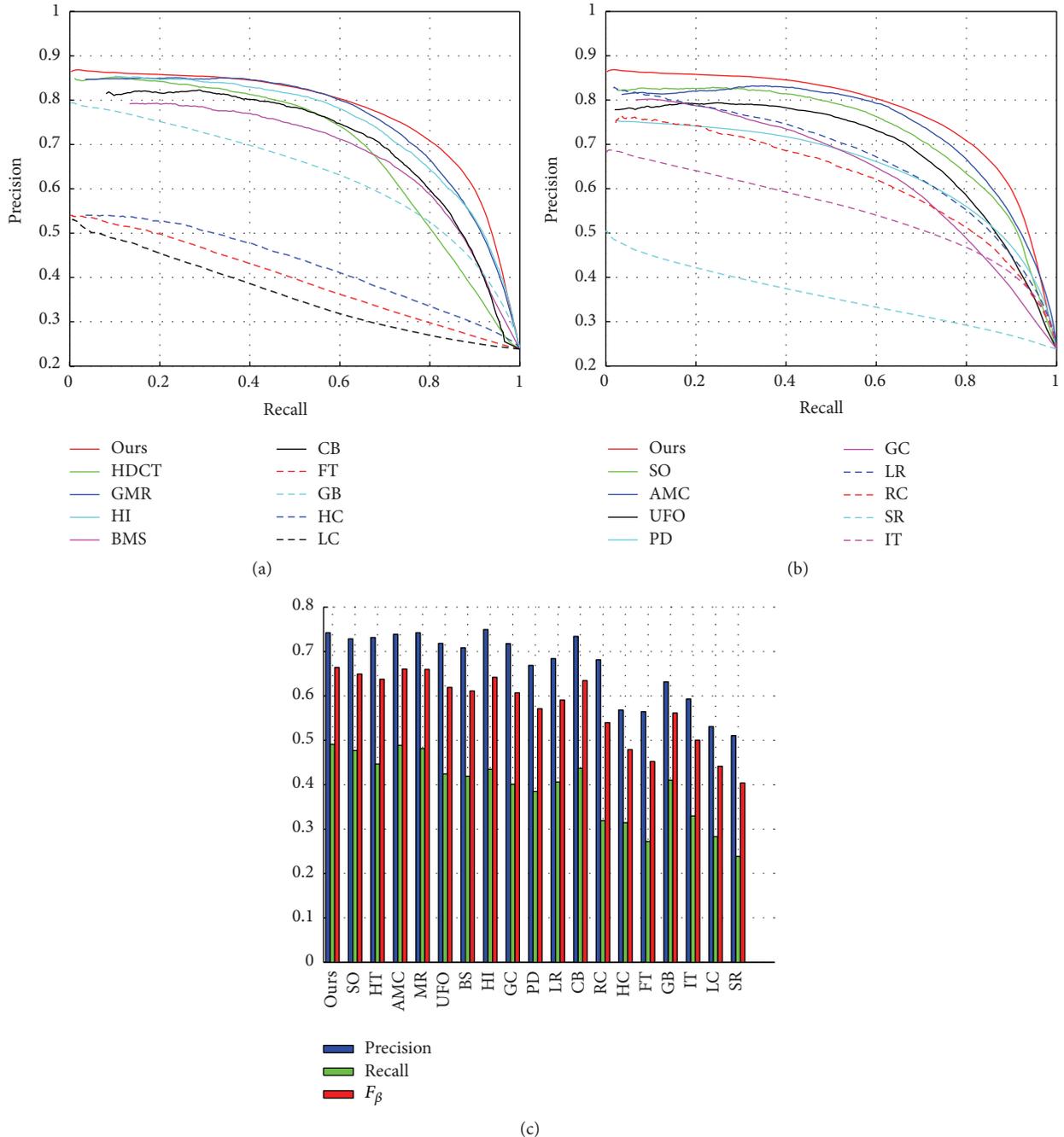


FIGURE 5: Experimental results on the ECSSD dataset. (a) and (b) are precision and recall curves of all approaches which are obtained using fixed threshold. The histogram (c) (precision, recall, and  $F_\beta$ ) is obtained using adaptive thresholding.

**3.3. Performance on ECSSD Dataset.** The ECSSD dataset is a more challenging dataset provided by Yan et al. [10]. As is shown in Figure 5, our approach achieves the best precision-recall curve. We also evaluate average precision, recall, and  $F_\beta$  using adaptive thresholding; our recall and  $F_\beta$  value rank first among all the methods.

We also provide the visual comparison of different approaches in Figure 6, from which we see that our approach produces the best detection results on these images and

can highlight the entire salient object uniformly. We only consider the most recent thirteen models: FT [5], HC [6], RC [6], LR [8], PD [9], GC [12], HI [10], GMR [11], BMS [13], UFO [14], AMC [15], HDCT [16], and SO [17].

**3.4. Evaluation on Different Feature Combination.** To verify the effectiveness of the proposed feature combination method, we plot the corresponding  $P$ - $R$  curves and the histogram of four combination schemes on the ASD dataset.

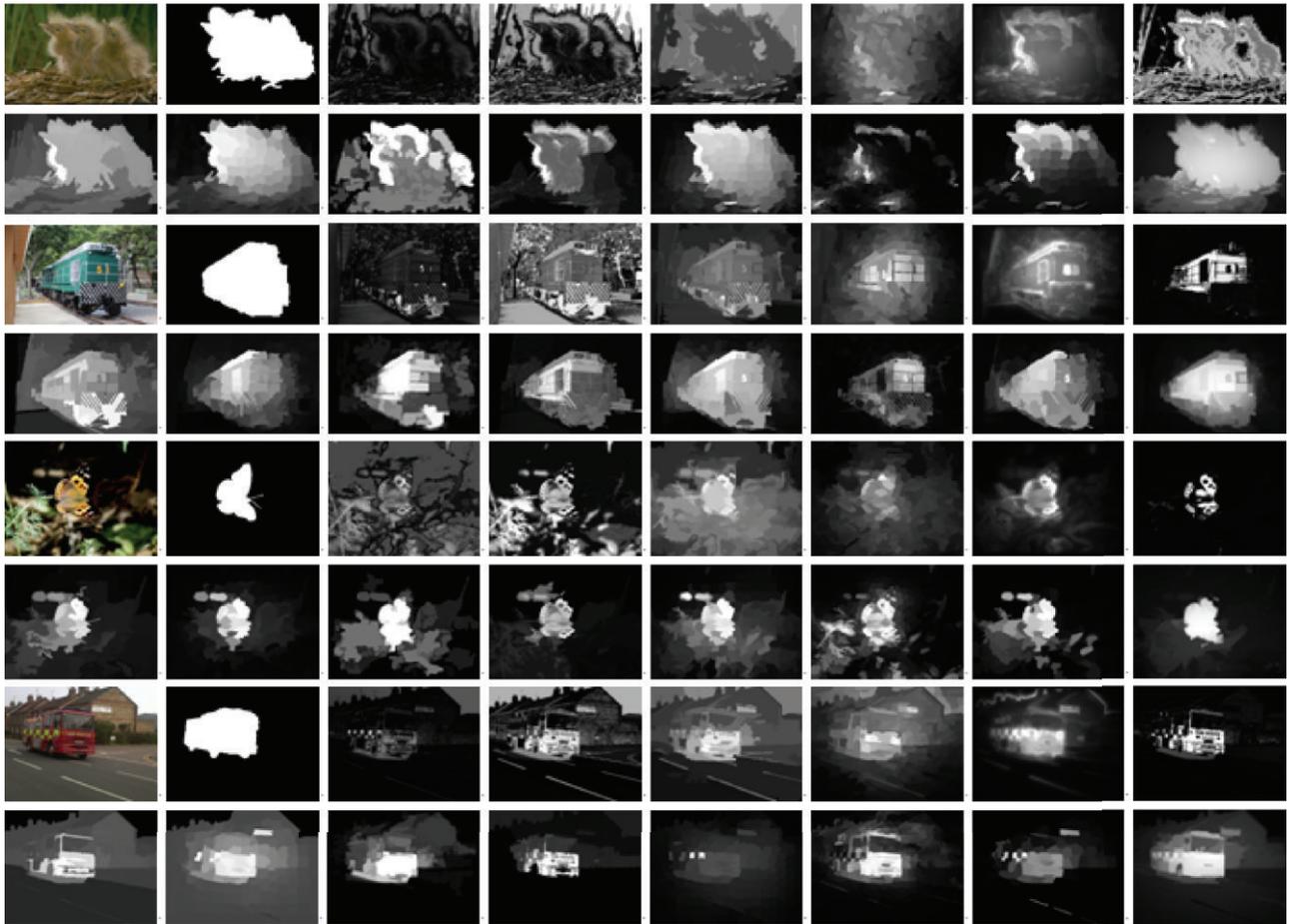


FIGURE 6: Saliency maps (from left to right, top to down): input, G-Truth, FT [5], HC [6], RC [6], LR [8], PD [9], GC [12], HI [10], GMR [11], BMS [13], UFO [14], AMC [15], HDCT [16], SO [17], and ours.

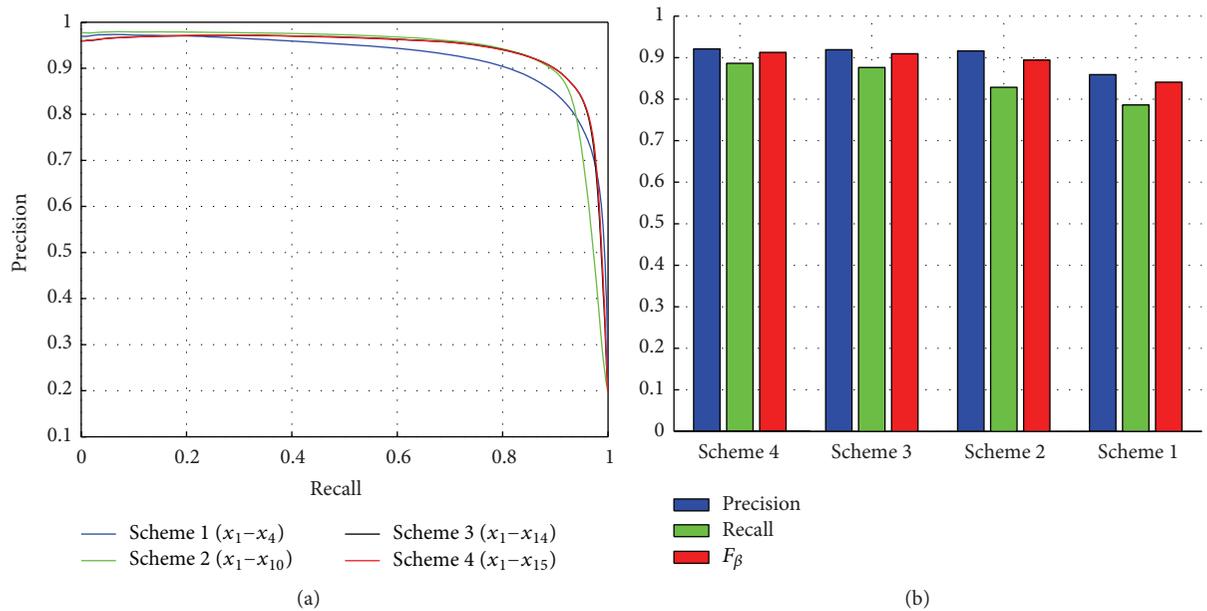


FIGURE 7: Experimental results on the ASD1000 dataset. (a) is PR curves of four schemes. The histogram (b) (precision, recall, and  $F_\beta$ ) is obtained using adaptive thresholding.

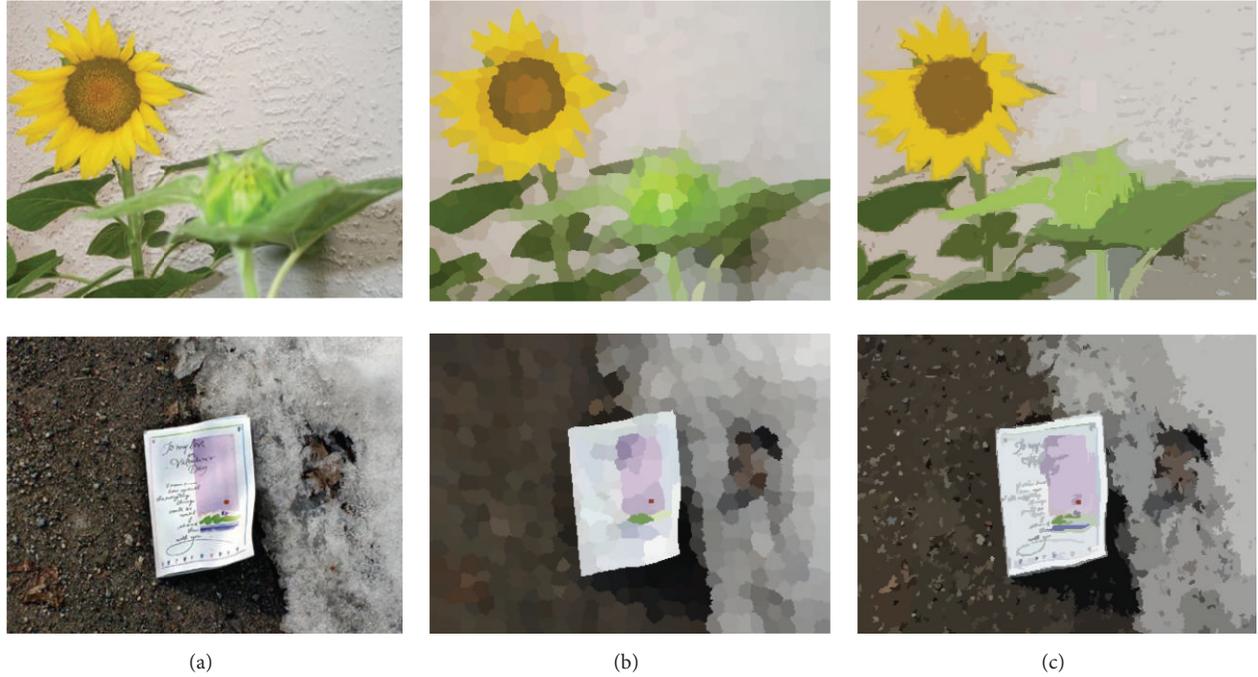


FIGURE 8: Visual comparison of SLIC and superpixel segmentation result, from left to right: input image, SLIC segmentation result, and superpixel segmentation result.

Four logistic classifiers are trained to get the parameters of four different combination schemes. We use the learned parameters to detect the images and also provide the quantitative comparison of different combination schemes in Figure 7. As can be seen, our approach (shown in Figure 7(b)) can get better result when we use more features. Similar conclusions can be obtained from the results of  $P$ - $R$  curves of different schemes (shown in Figure 7(a)): Scheme 1, four atomic features ( $x_1$ - $x_4$ ), Scheme 2, any combination of two atomic features and Scheme 1 ( $x_1$ - $x_{10}$ ), Scheme 3, any combination of three atomic features and Scheme 2 ( $x_1$ - $x_{14}$ ), and Scheme 4, any combination of four atomic features and Scheme 3 ( $x_1$ - $x_{15}$ ). Weighted combination of atomic features without considering any combination gets the lowest  $P$ - $R$  curves in most of the range. The curve of Scheme 2 is very close to Scheme 3 and Scheme 4 when recall is less than 0.9. Scheme 3 and Scheme 4 get almost the same curve.

**3.5. Analysis of the Influencing Factors of Segmentation.** Recently, low-level image segmentation methods have been widely used for saliency analysis. SLIC [38] and superpixel [35] approaches are two efficient algorithms and the source codes are publicly available. Because of considering different segmentation criterion, the segmentation results are quite different from each other, as is shown in Figure 8. From the figure, we can see that the result of SLIC method has more local compactness than superpixel method. We also provide the visual comparison of four saliency cues and final saliency map produced by the above two segmentation algorithms. Figure 9 shows that different segmentation algorithms can produce different salient cues

and final saliency map. The superpixel approach can generate high-quality saliency map, while the SLIC segmentation algorithm may highlight some nonsalient region. Finally, we provide the quantitative comparison of SLIC and superpixel segmentation algorithm. To verify the effectiveness of two segmentation algorithms, we plot the corresponding precision-recall curves on the ASD dataset. As observed from Figure 10, the use of superpixel algorithm can obtain better precision-recall curves when compared with SLIC clustering algorithm.

## 4. Conclusion

In this paper, a novel salient region detection approach based on feature combination and discriminative classifier is presented. We use four saliency cues as atomic feature of segmented region in the image. To capture the interaction among different features, a novel feature vector is generated by mapping a four-dimensional regional feature to a fifteen-dimensional feature vector. A logistic regression classifier is trained to map a regional feature to a saliency value. We further introduce the multilayer saliency map integration and salient center for improvement. We evaluate the proposed approach on two publicly available datasets and the experiments results show that our model can generate high-quality saliency map which can uniformly highlight the entire salient object.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.



FIGURE 9: Comparison of salient feature with different segmentation methods. (a) Input image. (b) Ground truth. (c) Saliency map generated by using SLIC method. (d) Saliency map generated by using superpixel method. (e) Color contrast based salient feature by using SLIC method. (f) Color contrast based salient feature by using superpixel method. (g) Color distribution based salient feature by using SLIC method. (h) Color distribution based salient feature by using superpixel method. (i) Objectness based salient feature by using SLIC method. (j) Objectness based salient feature by using superpixel method. (k) High prior based salient feature by using SLIC method. (l) High prior based salient feature by using superpixel method.

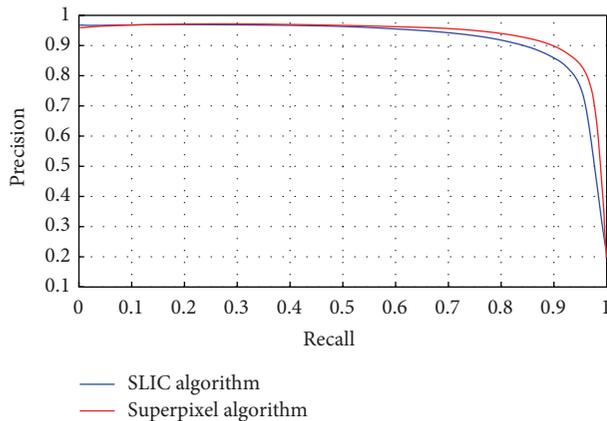


FIGURE 10: Comparison of different segmentation methods on the ASD dataset.

## Acknowledgments

This research is partly supported by the National Natural Science Foundation of China (no. 61305113, no. 61501394, and no. 61462038), Hebei Province Science and Technology Support Program, China (no. 13211801D), and The Specialized Research Foundation for the Doctoral Program of Higher Education by the Ministry of Education of PR China (no. 20131333110015).

## References

- [1] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [2] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proceedings of the 14th Annual ACM International Conference on Multimedia (MM '06)*, pp. 815–824, ACM, October 2006.
- [3] X. Hou and L. Zhang, "Saliency detection: a spectral residual approach," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, Minneapolis, Minn, USA, June 2007.
- [4] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS '06)*, pp. 545–552, Vancouver, Canada, December 2006.
- [5] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09)*, pp. 1597–1604, Miami, Fla, USA, June 2009.
- [6] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 409–416, IEEE, Providence, RI, USA, June 2011.
- [7] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li, "Automatic salient object segmentation based on context and shape prior," in *Proceedings of the British Machine Vision Conference (BMVC '11)*, p. 7, Dundee, UK, August-September 2011.
- [8] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 853–860, IEEE, Providence, RI, USA, June 2012.
- [9] R. Margolin, A. Tal, and L. Zelnik-Manor, "What makes a patch distinct?" in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 1139–1146, IEEE, Portland, Ore, USA, June 2013.
- [10] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 1155–1162, IEEE, Portland, Ore, USA, June 2013.
- [11] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 3166–3173, Portland, Ore, USA, June 2013.
- [12] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and N. Crook, "Efficient salient region detection with soft image abstraction," in *Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV '13)*, pp. 1529–1536, Sydney, Australia, December 2013.
- [13] J. Zhang and S. Sclaroff, "Saliency detection: a boolean map approach," in *Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV '13)*, pp. 153–160, Sydney, Australia, December 2013.
- [14] P. Jiang, H. Ling, J. Yu, and J. Peng, "Salient region detection by UFO: uniqueness, focusness and objectness," in *Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV '13)*, pp. 1976–1983, IEEE, Sydney, Australia, December 2013.
- [15] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang, "Saliency detection via absorbing Markov chain," in *Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV '13)*, pp. 1665–1672, IEEE, Sydney, Australia, December 2013.
- [16] J. Kim, D. Han, Y.-W. Tai, and J. Kim, "Salient region detection via high-dimensional color transform," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 883–890, IEEE, Columbus, Ohio, USA, June 2014.
- [17] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 2814–2821, IEEE, Columbus, Ohio, USA, June 2014.
- [18] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: contrast based filtering for salient region detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 733–740, IEEE, Providence, RI, USA, June 2012.
- [19] T. Liu, Z. Yuan, J. Sun et al., "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353–367, 2011.
- [20] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1915–1926, 2012.
- [21] K. Shi, K. Wang, J. Lu, and L. Lin, "PISA: pixelwise image saliency by aggregating complementary appearance contrast measures with spatial priors," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 2115–2122, Portland, Ore, USA, June 2013.

- [22] N. Tong, H. Lu, L. Zhang, and X. Ruan, "Saliency detection with multi-scale superpixels," *IEEE Signal Processing Letters*, vol. 21, no. 9, pp. 1035–1039, 2014.
- [23] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: a discriminative regional feature integration approach," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 2083–2090, Portland, Ore, USA, June 2013.
- [24] G. Sharma, F. Jurie, and C. Schmid, "Discriminative spatial saliency for image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 3506–3513, IEEE, Providence, RI, USA, June 2012.
- [25] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition?" in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 2, pp. II-37–II-44, Washington, DC, USA, June 2004.
- [26] H. Wu, Y.-S. Wang, K.-C. Feng, T.-T. Wong, T.-Y. Lee, and P.-A. Heng, "Resizing by symmetry-summarization," *ACM Transactions on Graphics*, vol. 29, no. 6, article 159, 2010.
- [27] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, no. 1, pp. 97–136, 1980.
- [28] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, no. 4, pp. 219–227, 1985.
- [29] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proceedings of the 11th Annual ACM International Conference on Multimedia (MM '03)*, pp. 374–381, Berkeley, Calif, USA, November 2003.
- [30] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Proceedings of the 12th International Conference on Computer Vision (ICCV '09)*, pp. 2106–2113, 2009.
- [31] Y. Lu, W. Zhang, H. Lu, and X. Xue, "Salient object detection using concavity context," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 233–240, Barcelona, Spain, November 2011.
- [32] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Proceedings of the 12th European Conference on Computer Vision (ECCV '12)*, pp. 29–42, Florence, Italy, October 2012.
- [33] R. Liu, J. Cao, Z. Lin, and S. Shan, "Adaptive partial differential equation learning for visual saliency detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 3866–3873, Columbus, Ohio, USA, June 2014.
- [34] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV '13)*, pp. 2976–2983, Sydney, Australia, December 2013.
- [35] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.
- [36] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [37] J. D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, John Wiley & Sons, 2004.
- [38] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2281, 2012.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

