ORIGINAL ARTICLE

Saliency bagging: a novel framework for robust salient object detection

Vivek Kumar Singh[1] · Nitin Kumar[1]

Abstract

Salient object detection is a challenging task, and several methods have been proposed for the same in the literature. The problem lies in that most of the methods perform good on a particular set of images but fail when exposed to a variety of different set of images. Here, we address this problem by proposing a novel framework called saliency bagging for detecting salient object(s) in digital images across a variety of images in a robust manner. The proposed framework generates the saliency map of an image in three phases: (i) Selection of existing saliency detection models and generation of initial saliency maps (ii) Generation of integrated binary map from the initial saliency maps by applying adaptive thresholding and majority voting (iii) Computation of final saliency map using integrated binary map and initial saliency maps by applying proposed integration logic. Extensive experiments on six publicly available datasets viz. MSRA10K, DUT-OMRON, ECSSD, PASCAL-S, SED2, and THUR15K have been performed to determine the effectiveness of the proposed method. The performance of the proposed method is measured in terms of Precision, Recall, F-Measure, Mean Absolute Error (MAE) and Receiver Operating Characteristic (ROC) curve and compared with 25 state-of-the-art methods including 17 classic best-performing methods of the last decade, five existing selected, and three aggregation saliency methods. The proposed method outperforms all the compared classic and existing selected methods in terms of Precision, F-Measure, and MAE, while it is comparable to the best-performing methods in terms of Recall and ROC curve across all the six datasets. The proposed framework is computationally very fast than all compared aggregation methods, while performance is almost same on all datasets that support its superiority.

Keywords Ensemble learning · Bagging · Majority voting · Aggregation · Saliency map

1 Introduction

Human visual system (HSV) has the ability to orientate rapidly toward the most salient objects with selective visual attention, which allows humans to semantically understand a visual input quickly and efficiently [18]. This selective visual attention has been intensively analyzed in psychology, neuroscience, and computer vision. In computer vision, there are several methods, for simulating the selective visual attention, which highlight salient object(s) in an image. Nowadays, saliency detection is widely applied to a variety of computer vision applications including image segmentation [10], objection recognition [45], image quality assessment [41], image editing and manipulating [9], information discovery [34], visual tracking [48], content-based image retrieval [6], etc.

Saliency detection methods can be roughly categorized as either bottom-up or top-down approaches. The bottom-up methods [1,8,14,19,20,22,24,29,33,37,42,46,52,55,57,60] are data-driven which usually use low-level visual features like intensity, orientation, or pattern from pixels or regions. These methods do not require training task so that they efficiently perform, while they are sensitive to noise regions due to data-driven characteristics. The contrast prior is one of the most applied principles in bottom-up methods. Based on the surroundings of contrast prior, these methods often categorized into local [19,24] and global [8] contrast prior methods. The local contrast effectively highlights boundaries (e.g., edges and texture) of the object while neglects to highlight the entire

✉ Vivek Kumar Singh
    singh.vivekrajput@gamil.com

    Nitin Kumar
    nitin@nituk.ac.in

[1] National Institute of Technology Uttarakhand, Uttarakhand, India
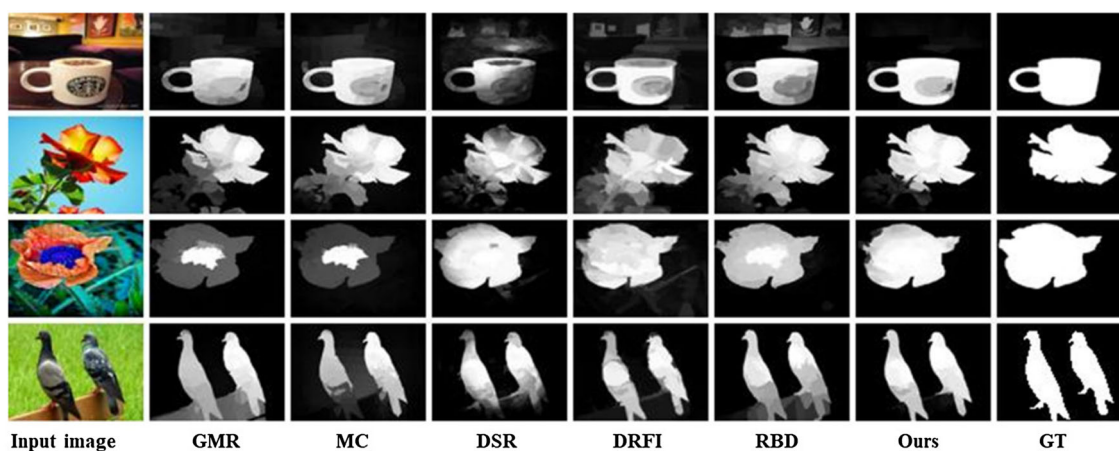
Fig. 1 Examples of saliency aggregation where saliency maps from the state-of-the-art methods do not show similar detection characteristics on variety of input images. From left to right, there are input images, the state-of-the-art methods, e.g., GMR [55], MC [20], DSR [30], DRFI [21], and RBD [21], our proposed saliency bagging framework, and ground truth (GT). Examples are selected from different datasets, e.g., THUR15K [7], MSRA10K [8], ECSSD [54], and SED2 [2] datasets

object. However, global contrast identifies the whole object but fail to keep the inner information of the object. To improve saliency detection performance, some methods jointly consider local and global contrast priors [36]. Furthermore, some bottom-up methods explore visual features, such as boundary prior [46], and central bias [60]. Recently, label propagation-based methods [20,55] have drawn more attention. Initially, seeds are picked based on some prior information and then, labels are propagated from seeds to other nodes in a graph. If the object appears on the image boundary, then these methods inaccurately identify the object. In contrast, top-down methods [13,16,17,21,23,39,47,54,56,59] are task-driven which require high-level information and learning mechanisms. These methods require training task which needs data collection, and hence it is complex and not efficient to implement. These methods are difficult to generalize due to task-driven characteristics. Lately, deep learning methods [15,26,27,31,51,61] have been widely and effectively applied for salient object detection.

Despite the development of several effective methods, none could show superiority than others for a variety of input images. Therefore, the saliency map computed using these state-of-the-art saliency detection methods is not reliable due to their performance dependence on individual images. As depicted in Fig. 1 that individual state-of-the-art method performs better at some images or some part of the image instead of performing well on all images. To address this problem, saliency aggregation has drawn rich attention recently. The saliency aggregation methods [4,25,38,43,53] incorporate several existing saliency maps according to diverse mechanisms. Accordingly, the saliency aggregation result can be superior to each selected existing methods that investigated in [4].

Consequently, the saliency aggregation is proposed to acquire the superiority of several existing saliency methods and alleviate the misleads of individual methods to improve the quality and robustness of the saliency map across a variety of images. Here, we propose a novel framework called saliency bagging in which multiple existing saliency detection methods are exploited to enhance the quality of the saliency map. The main contribution of the proposed research work is threefold:

1. A novel, simple, and efficient framework called saliency bagging is proposed which effectively incorporates advantages of selected saliency methods to improve the quality of saliency maps.
2. A novel integration logic is proposed to combine individual saliency maps into a final saliency map which chooses saliency value among individual saliency values for an image pixel based on the category of the pixel.
3. Extensive experimental evaluation is conducted with 25 different methods including 17 classic best-performing methods of the last decade, five existing selected, and three aggregation methods to verify the superiority of the proposed framework. The experimental results demonstrate that the proposed framework is better performs and robust across a variety of images. In addition, it is also computationally more effective than compared aggregation methods.

The rest of the paper is organized as follows. Section 2 presents related work and state-of-the-art methods in saliency detection. In Sect. 3, the proposed model is discussed in detail along with the step-wise procedure. Experimental results are demonstrated in Sect. 4, while Conclusion and Future Work are given in Sect. 5.

2 Related work

An extremely ample set of saliency detection methods have been investigated and obtained admirable performance for detecting salient objects in images over the last few years. Here, we broadly divide these methods into four classes according to their approaches used:

2.1 Bottom-up methods

Bottom-up methods are usually based on low-level visual attributes (e.g., color, orientation, and intensity) and prior knowledge (e.g., contrast, uniqueness, compactness, and boundary). Biologically motivated first computational saliency method was introduced by Itti et al. [19]. The method extracted center-surround contrast at multiscale with different features to measure the saliency. In [14], graph hypothesis is used to measure activation maps from the raw features. After that, Achanta et al. [1] proposed a simple model based on center-surround contrast that computes saliency of each pixel as the absolute difference between the mean pixel value of the image and blurred image pixel value. Perazzi et al. [42] decomposed an image into suitable elements that reduce unwanted details and measures rarity and distribution of each element in the form of contrast. They assign saliency value to each pixel from element contrast. Klein and Frintrop [24] formulated saliency computation in an information-theoretic way using Kullback–Leibler-Divergence (KLD). Cheng et al. [8] measured saliency via histogram-based contrast and region-based contrast using global knowledge. Uniqueness-, focusness-, and objectness (UFO)-based saliency computation is presented in [22]. Seo and Milanfar [46] computed local regression kernels, and then saliency is determined using self-resemblance measure (SeR). Zhang et al. [60] proposed a method in which rarity was measured using self-information of local visual cues derived from a Bayesian framework (SUN). Background priors used with geodesic distance to compute saliency in [52]. Yang et al. [55] suggested a graph-based manifold ranking (GMR) with background and foreground seeds to compute saliency values. Alternatively, saliency can be measured in terms of absorbed time, which is estimated in the absorbing Markov chain (MC) [20]. Lu et al. [37] proposed multi-graph-based rankings to measure the coarser saliency map and further refine the coarser saliency map. Li et al. [29] determined saliency by aggregating various low-level visual features. Yuan et al. [57] proposed saliency detection model that exploits a robust boundary prior to estimating the saliency. Further, regularized random walk ranking (RRWR) is applied on prior saliency estimation to generate saliency maps. Foreground-center-background (FCB)-based saliency model is proposed by Guang-Hai et al. [33] in which three saliency measures viz. foreground, center, and background are computed and then combined to generate final saliency map.

2.2 Top-down methods

Top-down methods use high-level knowledge and supervised learning approach for certain objects. Kanan et al. [23] proposed appearance-based saliency detection method that computed saliency in a Bayesian framework. Goferman et al. [13] proposed the context-aware (CA) saliency method which identifies salient object along with surrounding context via four principles namely local low-level features, global considerations, visual organization rules, and high-level information. Margolin et al. [39] proposed method to determine distinctive pattern for saliency computation (PCA). Hou and Zhang [16] converted an image into frequency domain and extract the spectral residual (SR) to compute saliency. Jiang et al. [21] suggested discriminatively features integration method (DRFI) to measure saliency values. Yan et al. [54] suggested a saliency model which addresses the scaling problem by a hierarchical framework (HSD) that combines significant information from three image layers at different scales, while this work extended on CSSD [47]. Zhang et al. [59] introduced saliency computation method based on the minimum barrier distance (MBD) transform. Yang et al. [56] measured saliency which jointly learns a conditional random field (CRF) and a visual dictionary. Yuang et al. [17] proposed saliency method which jointly considering object proposals and multiple instances learning (MIL) for saliency measure (MILPS).

2.3 Deep learning methods

Recently, many deep learning methods have been developed to saliency analysis and obtained good state-of-the-art performance. However, these methods required an extremely rich set of training data and a huge amount of computation time for deriving suitable network parameters (i.e., a set of learnable parameters). Hence, such methods may be not applicable to the real-time application and degrade performance on unseen visual data. Li and Yu [27] proposed a method which measures saliency using multiscale features computed by convolutional neural networks (CNNs). He et al. [15] suggested superpixel-based CNN that learns the hierarchical contrast features and keeps contextual information to generate a saliency map. A multi-context deep CNN architecture jointly considered the global and local context to compute saliency values [61]. Li et al. [31] introduced a fully convolutional neural network (FCCN) which encode priors knowledge and constitute multi-task learning to explore the inner correlation between saliency detection and semantic image segmentation. Salient instance segmentation approach which generates a saliency map according to a distinct object

instance in an input image [26]. Wang et al. [51] proposed saliency computation based on recurrent fully convolutional networks (RFCNs). The network incorporates visual priors in a more appropriate manner. The recurrent framework has the ability to automatically learn enriching saliency values by iteratively adjusting its previous errors.

2.4 Saliency aggregation methods

Existing saliency detection method varies on a variety of images (e.g., complex structure, boundary touching, small object, larger object, multiple objects, etc). In order to capture advantage of the superiority of each existing saliency method, many researchers have developed the saliency aggregation mechanism that suitably combined these methods in such a way to enhance performance universally on any kind of images. Borji et al. [4] suggested aggregation approaches based on statistical information and linear summation with three combination functions. Mai et al. [38] introduced three data-driven saliency aggregation methods to combine outcomes of individual existing saliency methods which enhance performance compared to individual existing saliency methods. Le et al. [25] suggested various unsupervised and supervised saliency aggregation methods and they observed that the simple average of the saliency maps generated by two best methods. Qin et al. [43] suggested an aggregation method named cuboid cellular automata (CCA) to incorporate various saliency maps generated by exiting saliency methods. Recently, Xu et al. [53] proposed an arbitrator model (AM) to integrate various existing saliency methods. The method derived a reference map on jointly considering the majority voting on multiple existing saliency methods and external knowledge. Afterward, the method learns the expertise of existing saliency methods. Lastly, a novel integration framework is employed based on Bayesian inference to integrate the existing saliency methods of varying expertise and the reference map.

3 Saliency bagging: proposed framework

Bagging [5] is a kind of ensemble learning approach that improves machine learning outcome by aggregating several models. In this approach, various learner models are required to apply on data set to generate various outcomes. These outcomes are integrated together to generate a better outcome than any of the single learner outcome. The integration task is defined in two different ways: the first way is average outcomes generated by various learners if the outcomes of various learners contain numerical values. The second way is majority voting on outcomes generated by various learners if the outcomes of various learners contain class labels. The learner models are independently produced error.

Hence, the integration task reduces the individual error of each learner model and evolves a better learner. Here, we propose a novel framework for salient object detection called saliency bagging which exploits the concept of bagging. In this framework, various existing saliency detection models are exploited to generate a saliency map which is robust than several other existing models. The proposed framework entails three main steps:

Step 1 Selection of existing salient object detection models and generation of *initial saliency maps*
Step 2 Generation of *integrated binary map* using initial saliency maps as obtained in step 1
Step 3 Generation of *final saliency map* using *integration logic*

A schematic representation of the proposed framework is shown in Fig. 2. In the proposed framework, first initial saliency maps are obtained using existing saliency detection models. Next, binary maps corresponding to these initial saliency maps are derived by applying adaptive thresholding as discussed later. Afterward, these binary maps are combined using majority voting to get the *integrated binary map*. The integrated binary map labels the image pixels in a better way in comparison with each individual binary map due to bagging. Based on the pixel labels of the integrated binary map, final saliency map is generated by applying integration logic in which the foreground pixels and background pixels are assigned the highest and the lowest saliency values, respectively, as available from the set of initial saliency maps.

3.1 Selection of existing saliency detection models and generation of initial saliency maps

The performance of the proposed saliency bagging framework depends on the selection of existing saliency detection models. Hence, the selection mechanism to select existing saliency detection models plays an important role here. Thus, the selection mechanism must choose those models which perform effectively on various categories of images. Therefore, we require a common environment to determine the performance of the existing saliency detection models on various datasets. One such study has been carried out by Borji et al. [3] in which several salient object detection models have been compared under the same environment. The performance of these methods has been measured in terms of Precision, Recall, F-measure, Area under the curve (AUC), and receiver operating characteristic (ROC), etc. Based on the benchmark given by Borji et al. [3], we have selected some existing saliency detection models whose performance is best individually in each of the above-mentioned five performance measures. The objective of the proposed framework
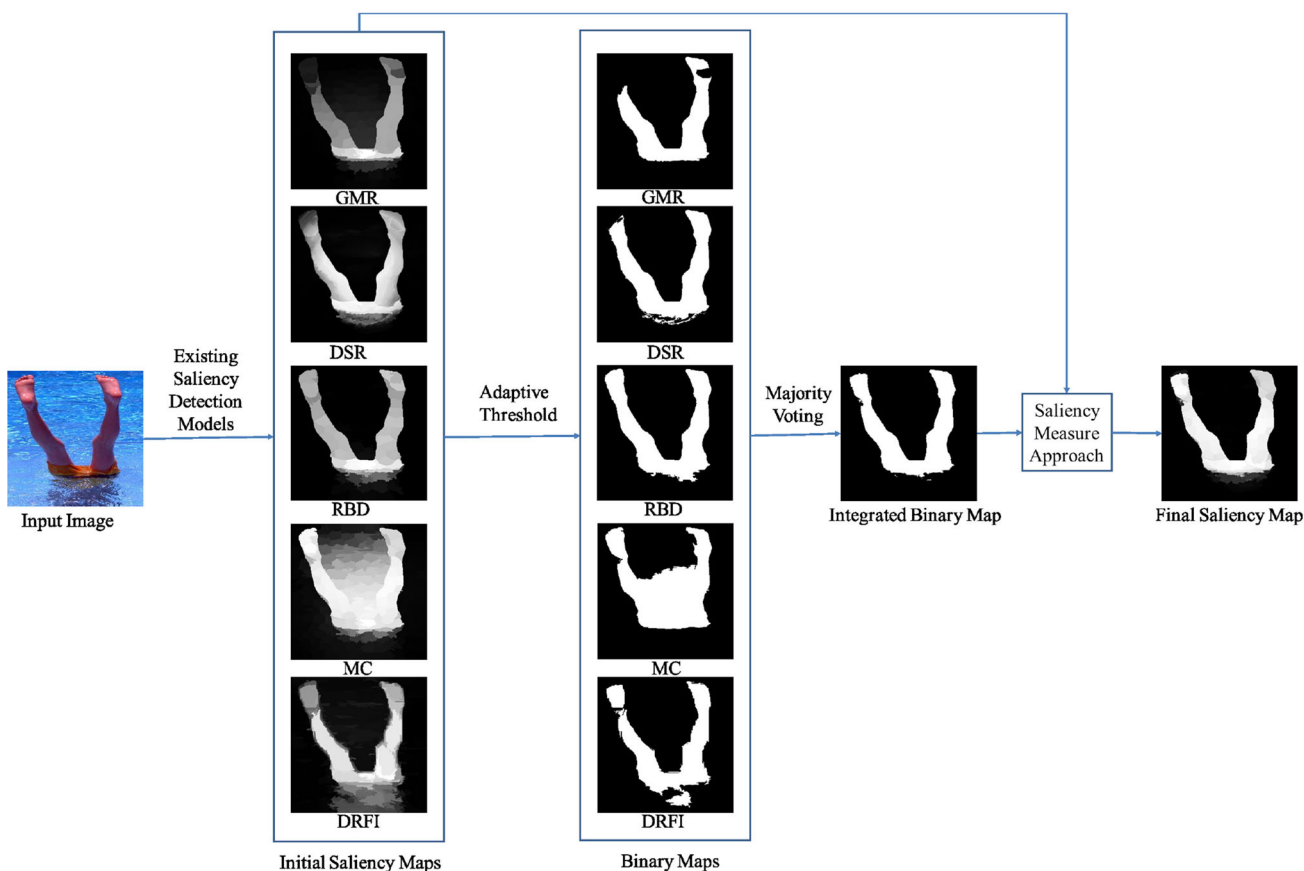
Fig. 2 Schematic representation of proposed framework

is to retain the advantages and suppress the disadvantages of existing saliency detection models. Hence, the proposed model incorporates various effective existing saliency detection models with ensemble learning mechanism, i.e., bagging to improve the quality of the saliency map. Now, we briefly introduce the selected existing saliency detection models which are used to generate the initial saliency maps as follows.

3.1.1 Graph-based manifold ranking (GMR) model

Yang et al. [55] proposed a saliency model that estimates the saliency of each image element (superpixel) as its ranking score. The model represents the input image as a graph $G = (V, E)$, where V is the set of nodes which illustrate elements of the image and E represents the set of edges between any pair of nodes i and j. The weight w_{ij} between nodes i and j is computed as follows:

$$w_{ij} = e^{-\frac{\|c_i - c_j\|}{\sigma^2}} \quad i, j \in V \tag{1}$$

where c_i and c_j is the mean color of corresponding nodes and σ is a constant to regulate the strength of the weight. The objective of the manifold ranking is to estimate rank f_i of each node in the graph with respect to y_i and optimal ranking is computed as follows [55]:

$$\mathbf{f}^* = \arg\min_f \frac{1}{2} \left(\sum_{i,j=1}^n w_{ij} \left\| \frac{f_i}{\sqrt{d_{ii}}} - \frac{f_j}{\sqrt{d_{jj}}} \right\|^2 + \mu \sum_{i=1}^n \|f_i - y_i\|^2 \right) \tag{2}$$

where $d_{ii} = \sum_j w_{ij}$, μ control the smoothing and $\mathbf{f}^*$ is optimized ranking vector. Equation (2) is simplified to obtain the following equation:

$$\mathbf{f}^* = (\mathbf{D} - \mathbf{W})^{-1} \mathbf{y} \tag{3}$$

where $\mathbf{D}$ is the degree matrix and $\mathbf{W}$ is the affinity matrix. Using Eq. (3), saliency map is obtained by selecting a various set of queries.

3.1.2 Absorbing Markov chain (MC) model

Similar to the model given in [55], Jiang et al. [20] also represent the input image as a graph but with different constraints on edges. The weight of the edge is computed in a similar manner to [55]. Here, saliency detection is formulated as an absorbing Markov chain model over image graph. The superpixels at the boundary are duplicated as virtual boundary nodes and are called absorbing nodes in the given model. Rest of the nodes are termed as transient nodes. Afterward, transient matrix $\mathbf{P}$ is obtained as follows [20]:

$$\mathbf{P} = \mathbf{D}^{-1} \times \mathbf{A} \tag{4}$$

where $\mathbf{A}$ is the affinity matrix and $\mathbf{D}$ is the degree matrix define as $\mathbf{D} = \text{diag}(\sum_j a_{ij})$. The fundamental matrix $\mathbf{N}$ is derived from Eq. (4), and absorbing time vector $\mathbf{y}$ is computed as follows:

$$\mathbf{y} = \mathbf{N} \times \mathbf{c} \tag{5}$$

Next, the saliency map is generated with the help of the normalized absorbing time vector $\bar{\mathbf{y}}$ as follows:

$$\mathbf{S}(i) = \bar{\mathbf{y}}(i) \quad i = 1, 2, ..., t \tag{6}$$

Lastly, the saliency map obtained using Eq. (6) is further refined to get the final saliency maps.

3.1.3 Dense and sparse reconstruction (DSR) model

Li et al. [30] designed dense and sparse appearance model for finding salient regions in an images. The input image is segmented into N segments and represented as a matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N] \in \mathbb{R}^{D \times N}$, where $\mathbf{x}_i$ is a feature vector of ith segments and D is feature dimension. A background templates set is constructed with the help of background segments $\mathbf{b}_M$ feature vectors and represented as $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_M]$ where M is number of background segments. Next, dense and sparse reconstruction error is calculated with the help of background template. The dense reconstruction error of segment i is calculated as:

$$\varepsilon_i^d = \|\mathbf{x}_i - (\mathbf{U}_\mathbf{B}\beta_i + \bar{\mathbf{x}})\|_2^2 \tag{7}$$

where $\mathbf{U}_\mathbf{B}$ eigenvectors from the normalized covariance matrix of B, β_i is the reconstruction coefficient of segment i, and $\bar{\mathbf{x}}$ is the mean feature of $\mathbf{X}$. The sparse reconstruction error of segment i is calculated as:

$$\varepsilon_i^s = \|\mathbf{x}_i - \mathbf{B}\alpha_i\|_2^2 \tag{8}$$

where α_i is an encoded representation of segment i. The reconstruction errors are smoothed based on context obtained by K-means clustering. After that, saliency value is assigned to each pixel by combining multiscale reconstruction errors which are further refined with an object biased Gaussian model. Finally, Bayes' formula is applied to integrate both reconstruction error measures for generating final saliency maps.

3.1.4 Discriminative regional feature integration approach (DRFI) model

Supervised learning-based saliency model is proposed by Jiang et al. [21] that extracts several features such as regional contrast, regional property, and regional backgroundness and construct feature vector $\mathbf{x}$ of each region of the input image. In learning phase, the training set contains a set of confident regions $\mathcal{R} = \{R_1, R_2, ..., R_Q\}$ and corresponding saliency scores $\mathcal{A} = \{a_1, a_2, ..., a_Q\}$ from multi-level segmentation of a set of images. The region is considered to be confident if its 80% pixels belong to either foreground or background, and corresponding saliency score is assigned as either 1 or 0. Then, learn a random forest regressor f from the training data $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_Q\}$ and the saliency scores $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_Q\}$. Thereafter, a set of multi-level saliency maps $\{A_1, A_2, ..., A_M\}$ (where M is number of levels) are generated and integrated to get final saliency map.

3.1.5 Robust background detection (RBD) model

A robust background measure-based saliency model is proposed by Zhu et al. [62] that characterizes the spatial geography of image regions with respect to image boundaries and magnitude of robustness. The model represents the input image as graph $G = (V, E)$, where V nodes of the graph represent superpixels of the image and E edges on pair p and q that represent adjacent node. The weight $d_{\text{app}}(p, q)$ is calculated as the Euclidean distance between two nodes p and q on mean color in CIE-Lab color space. After that, the boundary connectivity is calculated with the help of the length along the boundary and spanning area of each superpixel. Next, the background weighted contrast is computed as:

$$wCtr(p) = \sum_{i=1}^{N} d_{\text{app}}(p, p_i) \, w_{\text{spa}}(p, p_i) \, w_i^{\text{bg}} \tag{9}$$

where $w_{\text{spa}}(p, p_i) = exp\left(-\frac{d_{\text{spa}}^2(p, p_i)}{2\sigma_{\text{spa}}^2}\right)$, $d_{\text{spa}}(p, p_i)$ is the distance between the centers of superpixels p and p_i, set $\sigma_{\text{spa}} = 0.25$ and w_i^{bg} is computed from the boundary connectivity value of superpixel p_i represents background probability. Further, an optimization model is designed for optimizing saliency value of superpixel $\{s_i\}_1^N$ in which the cost function is defined as [62]:
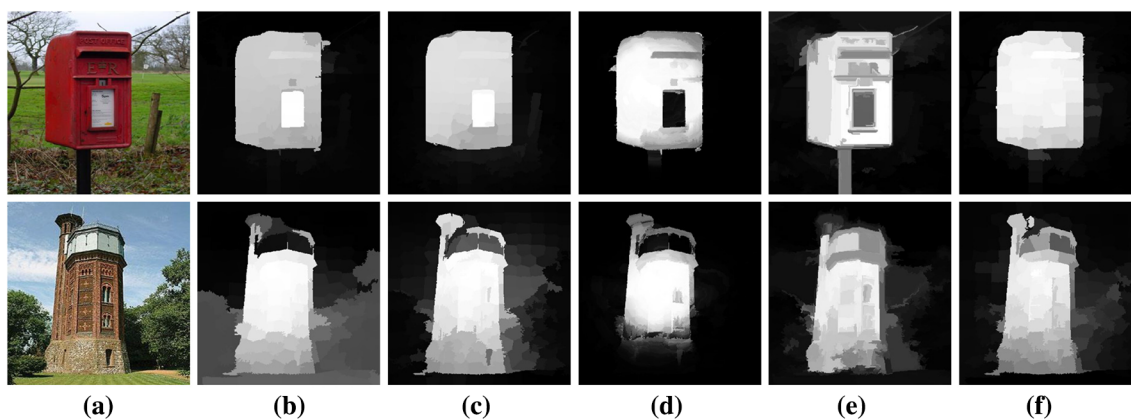
Fig. 3 Initial saliency maps. **a** Images **b** GMR [55] saliency maps **c** MC [20] saliency maps **d** DSR [30] saliency maps **e** DRFI [21] saliency maps **f** RBD [62] saliency maps

$$\sum_{i=1}^{N} w_i^{\text{bg}} s_i^2 + \sum_{i=1}^{N} w_i^{\text{fg}} (s_i - 1)^2 + \sum_{i,j} w_{ij} (s_i - s_j)^2 \quad (10)$$

where w_i^{bg} is foreground probability and w_{ij} controls smoothness. Finally, the saliency map is generated by minimizing the cost function.

3.1.6 Generation of initial saliency maps

In this work, we exploit the bagging approach to measure the saliency of the images. We first compute initial saliency maps $\mathbf{S}^j = \{\mathbf{S}_1^j, \mathbf{S}_2^j, ..., \mathbf{S}_N^j\}$ of j-th image and N represents the number of selected existing saliency detection models. The initial saliency maps are computed with the help of various selected existing saliency detection models. Few sample images along with the initial saliency maps generated using selected existing saliency detection models are shown in Fig. 3.

3.2 Generation of integrated binary map

The initial saliency maps obtained in the previous step are transformed into corresponding binary maps where each pixel has value either 0 or 1. Suppose $\mathbf{B}_i^j$ is the binary map corresponding to initial saliency map $\mathbf{S}_i^j$ generated by ith model on jth image. Thus, the complete set of binary maps can be represented as $\mathbf{B}^j = \{\mathbf{B}_1^j, \mathbf{B}_2^j, ..., \mathbf{B}_N^j\}$. The binary map is obtained from the saliency map using popular adaptive thresholding as given in Achanta et al. [1] as follows:

$$T_i^j = \frac{2}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} S_i^j \quad i = 1, 2.., N \text{ and } j = 1, 2, .., n \quad (11)$$

where W and H are width and height of the input image, n is total number of images in a particular dataset. Hence, binary maps $\mathbf{B}_i^j$ corresponding to initial saliency map $\mathbf{S}_i^j$ are computed as follows:

$$\mathbf{B}_i^j (x, y) = \begin{cases} 1 & \text{if } \mathbf{S}_i^j (x, y) \geqslant T_i^j \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

where (x, y) represents the location of the pixel under consideration and $1 \leq x \leq W$ and $1 \leq y \leq H$. The binary maps thus contain only two values, i.e., 1 and 0 where 1 denotes foreground region and 0 denotes background region in a given image. Therefore, the binary maps are labeled maps which classified the original image pixels into two classes. Further, we use binary maps for integration logic. Since the binary maps represent class labels, we apply the majority voting scheme to obtain an integrated binary map which classifies each pixel as background or foreground in a better manner. To find the integrated binary map $\mathbf{B}^j$ corresponding to j-th image, the following equation is used:

$$\mathbf{B}^j(x, y) = \begin{cases} 1 & \text{if } \sum_{i=1}^{N} \mathbf{B}_i^j(x, y) > N/2 \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

For proper working of the above equation, N must be an odd number. In this research work, we have chosen $N = 5$.

3.3 Generation of final saliency map

The integrated binary map obtained in the previous step helps in the generation of final saliency map $\mathbf{S}_f^j$. For this purpose, we propose a novel integration logic where at the foreground locations in the integrated binary map, the saliency value is chosen to be maximum amid the initial saliency maps. Similarly, at the background locations, minimum saliency value amid the initial saliency maps is chosen. The mathematical

formulation of the proposed integration logic along with initial saliency maps $\mathbf{S}^j$ corresponding to j-th image is given below:

$$\mathbf{S}_f^j(x, y) = \begin{cases} \max\{\mathbf{S}_1^j(x, y), \mathbf{S}_2^j(x, y), ..., \mathbf{S}_N^j(x, y)\} & \text{if } \mathbf{B}^j(x, y) = 1 \\ \min\{\mathbf{S}_1^j(x, y), \mathbf{S}_2^j(x, y), ..., \mathbf{S}_N^j(x, y)\} & \text{otherwise} \end{cases}$$
(14)

Thus, the final saliency map contains maximum saliency value among all the five initial saliency maps at the pixels which are classified as foreground, i.e., the salient region. Therefore, the salient region in the final saliency map is highlighted in a better manner as compared to all the five initial saliency maps. Similarly, the final saliency map contains minimum saliency value among all the five initial saliency maps at the pixels that are classified as background, i.e., the non-salient region. Therefore, the non-salient region in the final saliency map is suppressed in comparison with all the five saliency maps under consideration.

4 Experimental results

We have performed extensive experiments to determine the efficacy of the proposed method. The experiments have been carried out on six publicly available benchmark datasets: MSRA10K [8], DUT-OMRON [55], ECSSD [54], PASCAL-S [32], SED2 [2], and THUR15K [7]. The qualitative as well as quantitative performance of the proposed method is evaluated and compared with 25 state-of-the-art saliency detection models viz. SR [16], SUN [60], SeR [46], CA [13], SEG [44], SWD [11], FES [49], SIM [40], COV [12], PCA [39], GMR [55], MC [20], DSR [30], DRFI [21], RBD [62], LPS [28], MST [50], HSD [47,54], BMS [58], MILPS [17], RCRR [57], FCB [33], M-EST [25], CCA [43], and AML [53]. Next, we give details of the datasets, evaluation metrics, comparison of the proposed model with above-mentioned models and failure cases. All the experiments are carried out using a desktop PC with following configuration: Intel(R)Core(TM)i7-4770 CPU@3.40GHz.

4.1 Datasets

As discussed above, all the experiments are performed on six publicly available datasets as given in Table 1. MSRA10K dataset [8] is constructed from the MSRA dataset [35] and contains 10,000 images. The dataset also contains pixel-level saliency labels, and it is a large dataset. DUT-OMRON dataset [55] contains 5168 manually selected high-quality images. The images are more complex and contain one or more salient object(s). Hence, the dataset is more difficult and challenging for salient object detection models. ECSSD

Table 1 Summary of datasets used in the experiments

S. no.	Dataset	Total # of images
1	MSRA10K [8]	10, 000
2	DUT-OMRON [55]	5168
3	ECSSD [54]	1000
4	PASCAL-S [32]	850
5	SED2 [2]	100
6	THUR15K [7]	15, 000

dataset [54] contains 1000 natural images, which are semantically meaningful but structurally more complex. PASCAL-S dataset [32] contains 850 natural images. The images have multiple complex object and cluttered background. SED2 dataset [2] contains 100 images. The images have two salient objects. THUR15K dataset [7] is constructed from Flickr with 5 keywords, i.e., Butterfly, Coffee Mug, Dog Jump, Giraffe, and Plane. It contains 15,000 images in which only 6000+ images have pixel-accurate ground truth annotations for salient object regions.

4.2 Evaluation metrics

The performance of the proposed saliency bagging model and 25 state-of-the-art saliency detection models are measured in terms of Precision, Recall, Receiver Operating Characteristics (ROC), F-Measure, and Mean Absolute Error (MAE). Precision, Recall, and ROC are computed based on the overlapping region between ground truth and saliency map. Precision indicates the ratio of correctly computed salient pixels to all the pixel in the computed salient region, while Recall indicates the ratio of correctly computed salient pixels to all the pixels in the ground truth. F-Measure is computed as the weighted harmonic mean of Precision and Recall. The MAE is computed as the mean of the absolute difference between the saliency map and ground truth. Let saliency map whose intensity values are normalized between 0 and 1 is represented as $\mathbf{S}$, while ground truth is represented as $\mathbf{G}$. Then, a binary map ($\mathbf{B}$) is found from the saliency map by using an adaptive threshold as suggested in [1] for computing average Precision, average Recall, and average F-Measure. The threshold is varied between 0 and 1 in equal steps to find ROC performance measures. The notation $|.|$ is used to represent the number of nonzero entries in the binary map and ground truth annotation. The mathematical formulae for various performance measures are given below [3]:

$$\text{Percision} = \frac{|\mathbf{B} \cap \mathbf{G}|}{|\mathbf{B}|}$$
(15)

$$\text{Recall} = \frac{|\mathbf{B} \cap \mathbf{G}|}{|\mathbf{G}|}$$
(16)
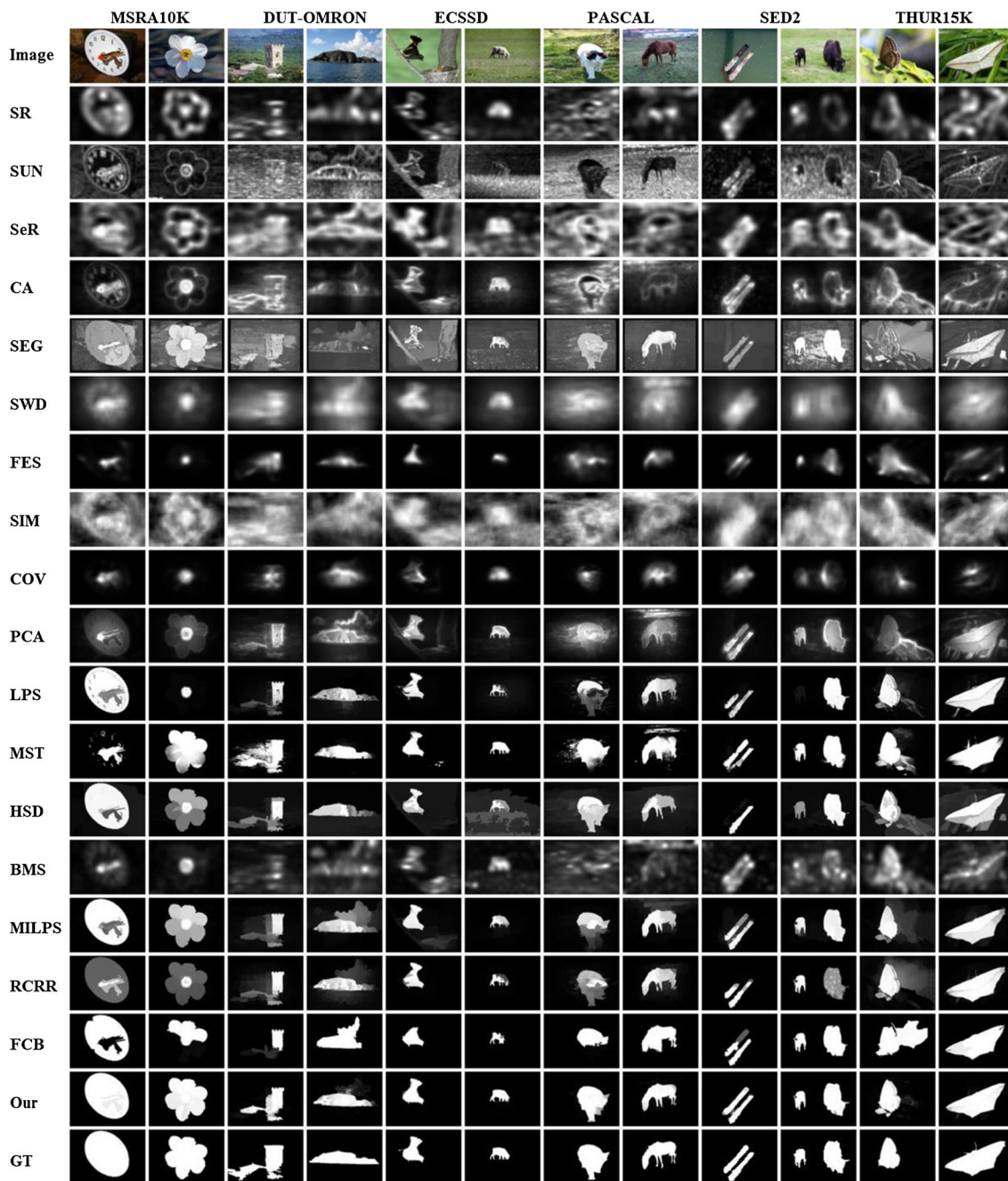
Fig. 4 Qualitative comparison of proposed framework with 17 classic saliency methods on six datasets. The images are sequentially with pair of two from MSRA10K [8], DUT-OMRON [55], ECSSD [54], PASCAL-S [32], SED2 [2], and THUR15K [7] datasets

F-Measure is computed from Precision and Recall as follows [3]:

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

For our experiments, β is set to be 0.3 as suggested in [1] to emphasize Precision more than Recall. Receiver operating characteristics (ROC) curve is the plot in which a false positive rate (FPR) is on the x-axis and true positive rate (TPR) on the y-axis. The TPR and FPR are computed as follows [3]:

$$\text{TPR} = \frac{|\mathbf{B} \cap \mathbf{G}|}{|\mathbf{G}|} \quad (18)$$

$$\text{FPR} = \frac{|\mathbf{B} \cap \bar{\mathbf{G}}|}{\bar{\mathbf{G}}} \quad (19)$$
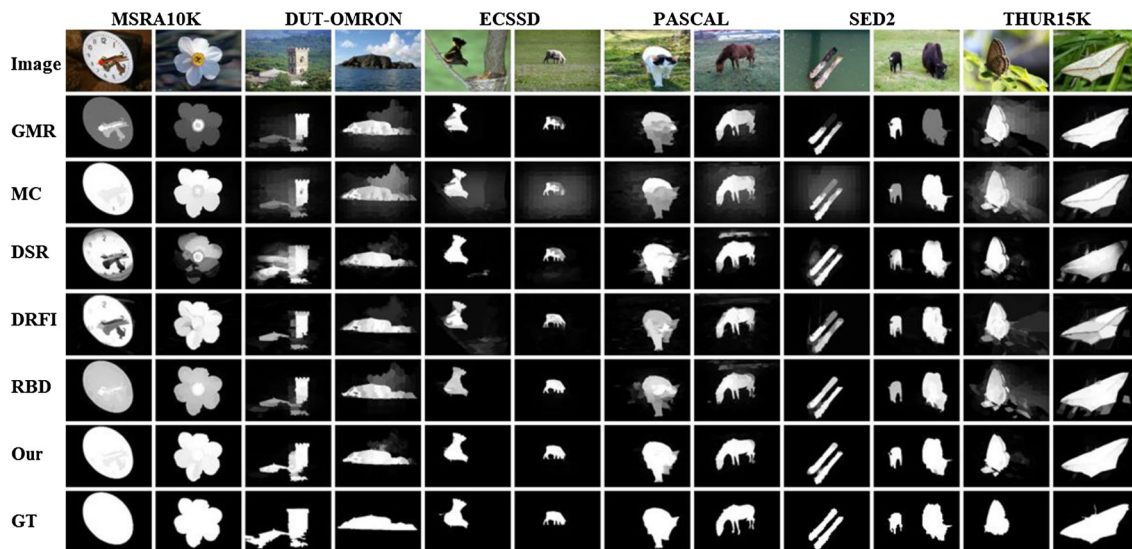
Fig. 5 Qualitative comparison of proposed framework with five existing selected saliency methods. The images are sequentially with pair of two from MSRA10K [8], DUT-OMRON [55], ECSSD [54], PASCAL-S [32], SED2 [2], and THUR15K [7] datasets

Mean Absolute Error (MAE) is found between saliency map and ground truth as follows [3]:

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

4.3 Comparison with state-of-the-art models

4.3.1 Qualitative comparison

Here, we show a qualitative comparison of the proposed framework with 25 state-of-the-art methods. Several visual examples are shown in Figs. 4, 5 and 6 on six datasets. Fig. 4 provides visual comparison of the proposed framework with 17 classic saliency methods on all the six datasets. It can be easily observed that the proposed framework can produce more accurate saliency maps, while most classic saliency methods fail to generate good quality saliency maps on complex background and foreground.

For example, (a) homogeneous background and heterogeneous foreground images: column 1, 2, 6 in Fig. 4, the proposed framework detected salient object as whole; (b) cluttered background: column 3, 4, 8, 11, 12 in Fig. 4, in this scenario the proposed framework clearly removes background noises and highlights salient regions; (c) multiple objects: column 9, 10 in Fig. 4, and the proposed framework effectively identifies all salient object present in the visuals, in contrast others fail to detect them either as whole or partially. (d) object touches the image boundary: column 7, 9 in Fig. 4, and the proposed framework detect salient region that is linked to the boundary of the image. (e) heterogeneous

background: column 5 in Fig. 4, our framework significantly suppressed all different background regions and highlights salient region. Figure 5 demonstrates qualitative effectiveness of the proposed framework against five existing selected saliency methods. It reveals that the proposed framework better performs on a variety of images, while individually selected saliency methods fail to effective on complex scenarios. For example, complex scenarios: column 1, 3, 4, 11, 12 in Fig. 5 the proposed framework more effective against individual selected saliency methods. In addition, the proposed framework also produces good quality salience maps on other variety of images column 2, 5, 6, 7 in Fig. 5.

Figure 6 shows the qualitative comparison with three aggregation saliency methods on six datasets. It can be observed that the proposed framework is consistently superior than M-EST [25] aggregation approach on variety of images. Further, the proposed framework identifies salient region(s) in a visually better manner than AML [53] method on the images containing heterogeneous foreground, e.g., column 3 and 9 in Fig. 6. Here, the proposed framework is superior than AML method in terms of completeness, while the performance of the proposed framework is comparable in other scenarios. The proposed framework efficiently removes foreground noise in a better way than CCA [43] as depicted in columns 1 and 2. It also effectively highlights salient region than CCA [43] as shown in column 3, while the results are similar to CCA [43] in other scenarios. These visual examples of different saliency methods demonstrate superiority of the proposed framework on a variety of images. The proposed framework is effective on a variety of images due to accurately segmenting salient object (majority voting) and effectively reducing background noise (integration logic).
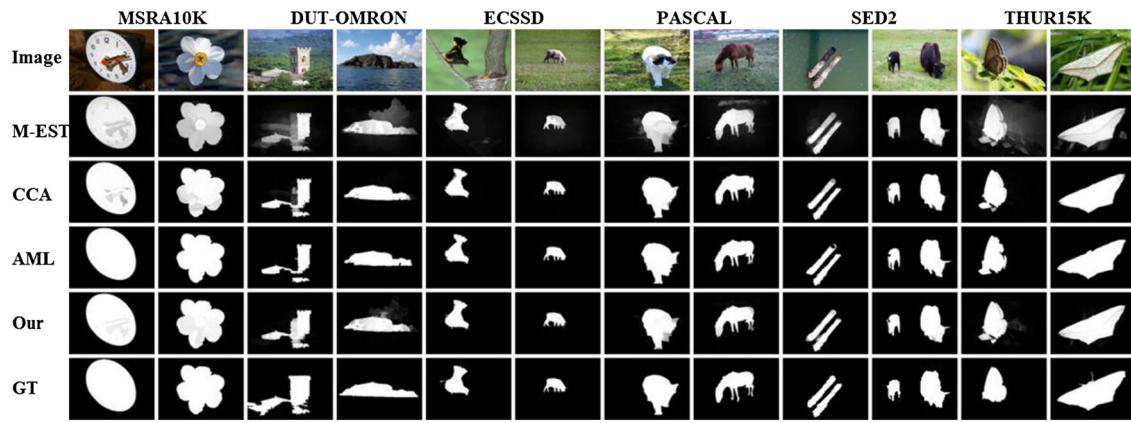
Fig. 6 Qualitative comparison of proposed framework with three aggregation saliency methods. The images are sequentially with pair of two from MSRA10K [8], DUT-OMRON [55], ECSSD [54], PASCAL-S [32], SED2 [2], and THUR15K [7] datasets

Table 2 Qualitative comparisons with classic 17 saliency methods on six datasets (Higher is better for Precision, Recall, and F-Measure while smaller is better for MAE and time

Model	Metric	MSRA- 10K	DUT-OMRON	ECSSD	PASCAL -S	SED2	THUR- 15K	Time (s)
SR [16]	P	0.5463	0.3084	0.4578	0.5575	0.4974	0.3926	
	R	0.4453	0.4135	0.3021	0.3278	0.5438	0.4960	
	F	0.5191	0.3276	0.4092	0.4799	0.5074	0.4124	*0.03*
	M	0.2487	0.2460	0.3106	0.2832	0.2333	0.2243	
SUN [60]	P	0.5237	0.2835	0.4198	0.4922	0.4792	0.3592	
	R	0.3345	0.2437	0.1523	0.1554	0.4113	0.3118	
	F	0.4632	0.2732	0.2987	0.3281	0.4616	0.3471	2.44
	M	0.3095	0.3429	0.3720	0.3434	0.2836	0.3087	
SeR [46]	P	0.5143	0.3305	0.3711	0.4738	0.4253	0.3269	
	R	0.4158	0.3883	0.2096	0.2700	0.4993	0.3591	
	F	0.4876	0.3422	0.3151	0.4035	0.4404	0.3338	0.73
	M	0.3099	0.3524	0.4043	0.3557	0.3023	0.3482	
CA [13]	P	0.6400	0.3749	0.5286	0.6085	0.5646	0.4220	
	R	0.5472	0.5190	0.3769	0.3582	0.5811	0.5387	
	F	0.6159	0.4006	0.4837	0.5240	0.5683	0.4442	34.58
	M	0.2373	0.2545	0.3090	0.2879	0.2269	0.2483	
SEG [44]	P	0.7367	0.5389	0.6446	0.7361	*0.8457*	0.5229	
	R	0.3463	0.3139	0.2457	0.1772	0.3066	0.2689	
	F	0.5846	0.4624	0.4689	0.4260	0.6016	0.4293	5.07
	M	0.3145	0.3369	0.3407	0.3352	0.3121	0.3371	
SWD [11]	P	0.7585	0.4671	0.7037	0.7210	0.5432	0.5217	
	R	0.4732	0.4137	0.3489	0.3063	0.4432	0.4740	
	F	0.6658	0.4536	0.5700	0.5493	0.5163	0.5099	0.10
	M	0.2667	0.3106	0.3183	0.3007	0.2968	0.2893	
FES [49]	P	0.7552	0.4630	0.6714	0.7348	0.6135	0.5089	
	R	0.5882	0.5695	0.5419	0.4577	0.5042	0.5935	
	F	0.7088	0.4839	0.6363	0.6447	0.5843	0.5263	***0.09***
	M	0.1849	0.1558	0.2116	0.2214	0.1955	0.1554	

Table 2 continued

Model	Metric	MSRA- 10K	DUT-OMRON	ECSSD	PASCAL -S	SED2	THUR- 15K	Time (s)
SIM [40]	P	0.4743	0.3278	0.3807	0.4982	0.4257	0.3447	
	R	0.1614	0.1622	0.0790	0.1146	0.3293	0.1607	
	F	0.3277	0.2653	0.2024	0.2811	0.3988	0.2726	0.96
	M	0.3880	0.4294	0.4333	0.3959	0.3842	0.4160	
COV [12]	P	0.7169	0.4512	0.6788	0.7054	0.5062	0.4840	
	R	0.5254	0.5555	0.5243	0.4216	0.4866	0.5727	
	F	0.6613	0.4716	0.6356	0.6105	0.5015	0.5020	13.10
	M	0.1972	0.1562	0.2146	0.2316	0.2103	0.1554	
PCA [39]	P	0.8048	0.4797	0.6611	0.6989	0.7158	0.4983	
	R	0.6750	0.6297	0.5215	0.4614	0.6550	0.6235	
	F	0.7706	0.5076	0.6227	0.6247	0.7008	0.5225	2.77
	M	0.1854	0.2063	0.2466	0.2403	0.2008	0.1985	
LPS [28]	P	*0.8917*	*0.5782*	0.7528	0.7550	0.8254	*0.5681*	
	R	0.7020	0.5894	0.5524	0.4149	0.6184	0.5730	
	F	0.8393	*0.5807*	0.6946	0.6349	0.7662	0.5692	1.07
	M	0.1240	*0.1448*	0.1864	0.2161	0.1414	*0.1503*	
MST [50]	P	0.8555	0.5180	0.7098	0.7449	0.8221	0.5667	
	R	**0.7634**	**0.7279**	**0.7147**	**0.5773**	**0.6811**	**0.7331**	
	F	0.8324	0.5549	0.7109	*0.6981*	**0.7846**	*0.5980*	0.18
	M	*0.0970*	0.1607	*0.1567*	0.1925	*0.1249*	0.1474	
HSD [47,54]	P	0.8728	***0.5571***	0.7408	0.7509	0.7582	0.5453	
	R	0.7122	0.5704	0.5326	0.4067	0.6129	0.5745	
	F	0.8296	0.5601	0.6795	0.6282	0.7189	0.5518	0.33
	M	0.1486	0.2274	0.2274	0.2505	0.2102	0.2198	
BMS [58]	P	0.7171	0.4401	0.5919	0.6538	0.6471	0.4815	
	R	0.5254	0.4910	0.3368	0.3215	0.5970	0.5219	
	F	0.6614	0.4509	0.5038	0.5279	0.6348	0.4903	**0.02**
	M	0.2303	0.2262	0.2856	0.2746	0.2215	0.2169	
MILPS [17]	P	*0.8962*	0.5446	*0.7698*	*0.7647*	***0.8266***	0.5635	
	R	**0.8209**	*0.6916*	*0.6797*	*0.5276*	**0.7501**	*0.7062*	
	F	*0.8776*	*0.5726*	0.7470	*0.6928*	0.8076	*0.5911*	64.48
	M	*0.1101*	0.1675	0.1770	*0.2118*	*0.1280*	0.1663	
RCRR [57]	P	0.8905	0.5464	***0.7693***	***0.7633***	0.8110	*0.5714*	
	R	0.7592	0.6318	0.6526	0.4889	0.6265	0.6478	
	F	*0.8563*	0.5640	*0.7388*	0.6758	0.7594	0.5874	0.33
	M	0.1221	0.1825	0.1838	0.2207	0.1611	0.1779	
FCB [33]	P	0.8620	0.5275	0.7297	0.7046	0.7465	0.5357	
	R	0.6019	0.5310	0.5032	0.3861	0.5273	0.5058	
	F	0.7839	0.5283	0.6610	0.5919	0.6812	0.5285	0.20
	M	0.1206	*0.1488*	*0.1728*	0.2211	0.1603	0.1510	
Ours	P	**0.9157**	**0.5837**	**0.8025**	**0.8007**	**0.8533**	**0.5967**	
	R	*0.7918*	*0.7112*	*0.6776*	*0.5261*	*0.7016*	*0.7108*	
	F	**0.8838**	**0.6089**	**0.7698**	**0.7146**	**0.8127**	**0.6196**	9.80
	M	**0.0805**	**0.1165**	**0.1386**	**0.1799**	**0.1185**	**0.1225**	

Top three models are emphasized in bold, italic, and bolditalic, respectively). *P*-Precision, *R*-Recall, *F*-F-Measure, *M*-MAE

Table 3 Qualitative comparisons with five existing selected saliency methods on six datasets (Higher is better for Precision, Recall, and F-Measure while smaller is better for MAE and time

Model	Metric	MSRA- 10K	DUT-OMRON	ECSSD	PASCAL -S	SED2	THUR- 15K	Time (s)
GMR [55]	*P*	***0.8888***	***0.5509***	0.7694	0.7643	0.8077	0.5730	
	R	0.7501	0.6232	0.6430	0.4797	0.6185	0.6373	
	F	0.8524	0.5661	0.7360	0.6722	0.7544	0.5867	***0.40***
	M	0.1257	0.1870	0.1861	0.2235	0.1659	0.1811	
MC [20]	*P*	0.8899	0.5489	***0.7700***	***0.7747***	***0.8185***	***0.5767***	
	R	0.7529	0.6691	0.6490	0.5031	0.6545	0.6797	
	F	***0.8540***	***0.5726***	***0.7382***	0.6889	0.7738	***0.5976***	**0.09**
	M	0.1450	0.1863	0.2023	0.2219	0.1838	0.1853	
DSR [30]	*P*	0.8537	0.5357	0.7515	0.7500	0.7689	0.5531	
	R	0.7367	0.6792	0.6443	0.5108	0.6827	0.6769	
	F	0.8236	0.5632	0.7237	0.6768	0.7471	0.5775	3.49
	M	*0.1208*	*0.1388*	*0.1710*	0.2038	0.1401	*0.1425*	
DRFI [21]	*P*	0.8782	*0.5598*	*0.7896*	*0.7837*	0.7932	*0.5856*	
	R	**0.8151**	**0.7422**	**0.6942**	*0.5313*	**0.7628**	**0.7572**	
	F	*0.8628*	*0.5935*	0.7653	*0.7063*	**0.7860**	*0.6179*	5.64
	M	0.1259	0.1496	*0.1702*	**0.2014**	*0.1363*	*0.1501*	
RBD [62]	*P*	0.8662	0.5253	0.7258	0.7474	*0.8443*	0.5345	
	R	***0.7885***	*0.7138*	***0.6621***	**0.5484**	*0.7119*	**0.7003**	
	F	0.8470	0.5594	0.7101	0.6896	*0.8096*	0.5654	*0.15*
	M	*0.1109*	***0.1462***	0.1713	*0.1972*	*0.1299*	0.1509	
Ours	*P*	**0.9157**	**0.5837**	**0.8025**	**0.8007**	**0.8533**	**0.5967**	
	R	*0.7918*	***0.7112***	*0.6776*	***0.5261***	***0.7016***	*0.7108*	
	F	**0.8838**	**0.6089**	**0.7698**	**0.7146**	**0.8127**	**0.6196**	9.80
	M	**0.0805**	**0.1165**	**0.1386**	**0.1799**	**0.1185**	**0.1225**	

Top three models are emphasized in bold, italic, and bolditalic, respectively). P-Precision, R-Recall, F-F-Measure, M-MAE

4.3.2 Quantitative comparison

To verify the effectiveness of the proposed framework, we also estimate quantitative comparison in three segments including *classic*, *existing selected*, and *aggregation* saliency methods as illustrated in Tables 2, 3, and 4, respectively.

Classic saliency methods The quantitative results of proposed framework and 17 classic saliency methods are presented in Table 2 in terms of various performance metrics Precision, Recall, F-Measures, MAE and computational time in second on MSRA10K [8] dataset. As can be observed, the proposed framework outperforms all the other methods across all the datasets in terms of Precision, F-Measure, and MAE due to the proposed framework take the advantages of existing combined saliency methods to accurately segment the salient and non-salient regions and aggregate saliency values using novel integration approach. It is comparable in terms of Recall with top performer MILPS [17] and MST [50] on MSRA10K [8], SED2 [2] and DUT-OMRON [55], ECSSD [54], PASCAL-S [32], respectively.

As suggested by Liu et al. [36], Recall plays a less important role than Precision in salient object detection performance. In addition, on MSRA10K [8], the proposed framework improves the Precision and F-Measure scores by approx. 2.2% and 0.7% over the best method MILPS [17] while improves MAE score by approx. 17% over the best method MST [50] among the compared classic methods. On DUT-OMRON [55], the proposed framework improves the Precision, F-Measure, and MAE scores by approx. 1%, 4.9%, and 19.5% over the best method LPS [28] among the compared classic methods. Next, on ECSSD [54], the proposed method improves the Precision and F-Measure scores by approx. 4.3% and 3.1% over the best method MILPS [17] while improves MAE score by approx. 11.6% over the best method MST [50] among the compared classic methods. On PASCAL-S [32], the proposed framework improves the Precision score by approx. 4.7% over the best method MILPS [17] while improves and F-Measure, and MAE scores by approx. 2.4% and 6.6% over the best method MST [50] among the compared classic methods. Further on SED2 [2], the proposed method improves the Precision,

Table 4 Qualitative comparisons with three aggregation saliency methods on six datasets (Higher is better for Precision, Recall, and F-Measure while smaller is better for MAE and time

Model	Metric	MSRA- 10K	DUT-OMRON	ECSSD	PASCAL -S	SED2	THUR- 15K	Time (s)
M-EST [25]	P	0.8894	0.5498	0.7787	0.7763	0.8035	0.5738	
	R	**0.8288**	**0.7482**	**0.7254**	**0.5654**	**0.7560**	**0.7580**	
	F	0.8746	0.5857	0.7657	*0.7148*	0.7920	0.6079	12.74
	M	0.1159	0.1682	0.1719	0.2031	0.1437	0.1660	
CCA [43]	P	*0.9191*	*0.5904*	*0.8071*	*0.8013*	*0.8558*	*0.6016*	
	R	*0.7966*	*0.7424*	*0.7071*	*0.5645*	*0.7153*	*0.7386*	
	F	*0.8876*	*0.6197*	*0.7816*	**0.7306**	*0.8187*	*0.6285*	10.03
	M	*0.0732*	***0.1176***	*0.1262*	*0.1726*	*0.1161*	***0.1226***	
AML [53]	P	**0.9423**	**0.6636**	**0.8496**	**0.8423**	**0.9041**	**0.6459**	
	R	0.7499	0.6697	0.6379	0.4831	0.6471	0.6750	
	F	**0.8897**	**0.6650**	**0.7891**	*0.7189*	**0.8282**	**0.6524**	*12.47*
	M	**0.0695**	**0.1004**	**0.1209**	**0.1725**	**0.1151**	**0.1094**	
Ours	P	*0.9157*	*0.5837*	*0.8025*	*0.8007*	*0.8533*	*0.5967*	
	R	*0.7918*	*0.7112*	*0.6776*	*0.5261*	*0.7016*	*0.7108*	
	F	*0.8838*	*0.6089*	*0.7698*	0.7146	*0.8127*	*0.6196*	**9.80**
	M	*0.0805*	*0.1165*	*0.1386*	*0.1799*	*0.1185*	*0.1225*	

Top three models are emphasized in bold, italic, and bolditalic, respectively). P-Precision, R-Recall, F-F-Measure, M-MAE

F-Measures and MAE scores by approx. 0.9 %, 0.6% and 5.1% over the best method SEG [44], MILPS [17], and MST [50] among the compared classic methods, respectively. On THUR15K [7], the proposed framework improves the Precision score by approx. 4.4% over the best method RCRR [57] while improves F-Measure and MAE scores by approx. 3.6% and 16.9% over the best method MST [50] among the compared classic methods. Hence, it is apparent that the proposed model improves Precision, F-Measure, and MAE by approx. 2.9%, 2.55, and 12.78 on average over the best method among the compared classic methods across all six datasets. The above discussion shows that the proposed framework possess effective performance in terms of correctness, completeness, and suppressing non-salient regions on different images scenarios which illustrate its robustness characteristics. In addition, the computational time of the proposed framework includes all combined saliency methods (i.e., 0.40, 0.09, 3.49, 5.64, and 0.15) and aggregation approach (i.e., 0.03) computational time. It is faster than several methods such as CA [13], COV [12], and MILPS [17].

Existing selected saliency methods In order to evaluate the effective performance of the proposed framework against existing selected saliency methods, we also estimate qualitative comparisons with combined saliency methods as illustrated in Table 3 in terms of various performance metrics Precision, Recall, F-Measures, MAE, and computational time in second on MSRA10K [8] dataset.

Table 5 Computational time of aggregation component

Model	M-EST [25]	CCA [43]	AML [53]	Ours
Time (s)	2.97	0.26	1.07	**0.03**

Bold is used to signify the least time required by the proposed framework

The scores presented in Table 3 demonstrate that the proposed framework outperforms all the individual methods in terms of Precision, F-Measure, and MAE because the proposed framework clearly suppressed background noises which are not suppressed by all individual saliency detection methods on same regions. However, these existing selected methods perform better on different regions of the image. We exploit this advantage of each individual method to find a saliency map which significantly improves the quality of salient object detection.

The performance of the proposed framework is comparable with DRFI [21] and RBD [62] on MSRA10K [8], ECSSD [54], THUR15K [7], PASCAL-S [32], and DUT-OMRON [55], SED2 [2], respectively, in terms of Recall. Furthermore, on MSRA10K [8], the proposed framework improves the Precision, F-Measure, and MAE scores by approx. 3%, 2.4%, and 27.4% over the best methods, i.e., MC [20], DRFI [21], and RBD [62] among the existing selected methods, respectively. On DUT-OMRON [55] dataset, the proposed framework improves Precision and F-Measure scores by approx. 4.3% and 2.6% over the best
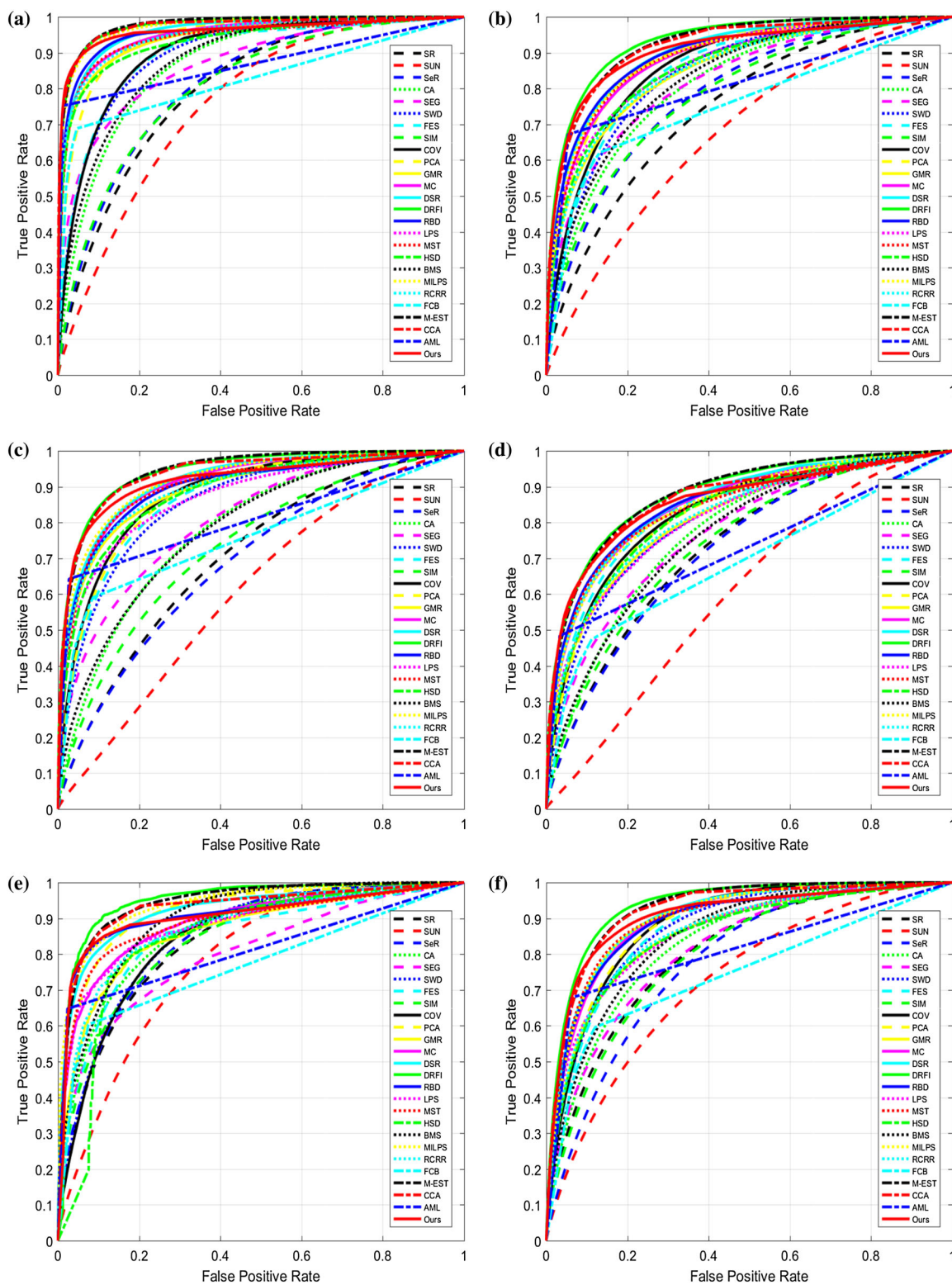
Fig. 7 ROC on the six datasets: **a** MSRA10K [8], **b** DUT-OMRON [55], **c** ECSSD [54], **d** PASCAL-S [32], **e** SED2 [2], **f** THUR15K [7]

method DRFI [21] while improves MAE score by approx. 16.1% over the best method DSR [30] among the existing selected methods.

On ECSSD [54], the proposed framework improves the Precision, F-Measure, and MAE scores by approx. 2.4%, 0.6%, and 18.6% over the best method DRFI [21] among the
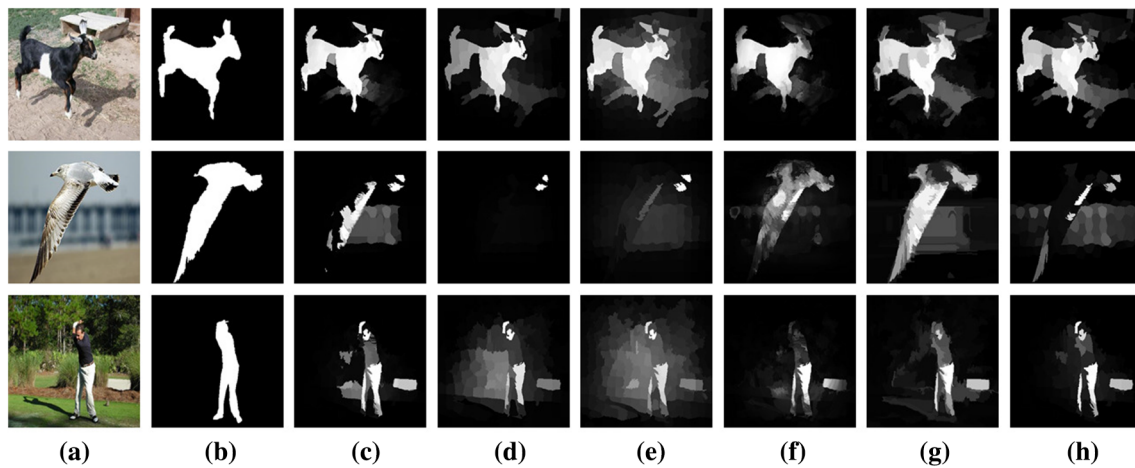
Fig. 8 Failure cases. **a** Images. **b** Ground truth. **c** Our saliency maps **d** GMR [55] saliency maps **e** MC [20] saliency maps **f** DSR [30] saliency maps **g** DRFI [21] saliency maps **h** RBD [62] saliency maps

existing selected methods. On PASCAL-S [32], the proposed framework improves the Precision and F-Measure scores by approx. 2.2% and 1.2% over the best method DRFI [21] while improves MAE score by approx. 8.8% over the best method RBD [62] among the existing selected methods. Similarly, on SED2 [2], the proposed framework improves the Precision, F-Measure, and MAE scores by approx. 0.8%, 0.4%, and 8.8% over the best method RBD [62] among the existing selected methods. On THUR15K [7], the proposed framework improves the Precision and F-Measure scores by approx. 1.9 % and 0.3% over the best method DRFI [21] while improves MAE score by approx. 14% over the best method DSR [30] among the existing selected methods. Thus, the proposed model improves Precision, F-Measure, and MAE scores by approx. 2.4%, 1.7%, and 15.6% on average over the best-performing method among the existing selected saliency methods across all six datasets.

Aggregation saliency methods The quantitative performance of the proposed framework with aggregation saliency methods is shown in Table 5 in terms of different metrics including Precision, Recall, F-Measure MAE and Table 5 also presented computational time on MSRA10K [8] dataset. The statistical data in Table 5 illustrate that the proposed framework outperforms on all six except PASCAL-S [32] datasets in terms of Precision, F-Measure, and MAE while it outperforms in terms of Precision and MAE from M-EST [25] on PASCAL-S [32] dataset. In addition, the proposed framework is more closer to CCA [43] in terms of all metrics, i.e., the performance is almost equal. In contrast, the proposed framework outperforms with AML [53] in term of Recall across all datasets while it is comparable on other metrics. The proposed framework performs in a consistent and robust manner on a variety of images. Furthermore, the computational time of M-EST [25], AML [53], and CCA [43] aggregation methods includes computation time of (i) indi-

vidual saliency methods and (ii) aggregation approach . As demonstrated in Table 4, the proposed framework is very fast in processing an image, and it is computationally more efficient than all compared aggregation methods which verify the superiority of the proposed framework against other compared aggregation methods. Table 5 displays the run time of the aggregation component of different aggregation methods. The proposed framework appreciably minimizes the aggregation computational time approx. 98.99%, 88.46%, and 97.20% from compared aggregation methods M-EST [25], CCA [43] and AML [53], respectively.

Furthermore, we evaluate the effectiveness of the proposed framework on varying threshold in terms of ROC against 25 different saliency methods as shown in Fig. 7 on six datasets. It can be seen that the proposed framework performs better than AML [53] while comparable with M-EST [25] and CCA [43] on all datasets. Hence, the proposed framework accurately separates the salient and non-salient regions on different thresholds that AML [53]. It can be also seen that DFRI [21] is better performer among all the methods on DUT-OMRON [55], SED2 [2], and THUR15K [7], while CCA [43] is better performs on other datasets. Hence, the performance of the proposed framework is comparable with better performing saliency detection methods and better than several classic state-of-the-art methods.

4.4 Failure cases

In this paper, we formulate the saliency detection problem as saliency bagging by using existing saliency detection models. The proposed models are highly effective on most of the saliency detection tasks. However, when the selected existing saliency detection models are simultaneously unable to detect the particular salient region then our model also fails to detect that particular region as shown in Fig. 8 top row.

As shown in Fig. 8, middle row majority of existing saliency detection models detected salient region as a background region that influences the performance of our model. In the bottom row of Fig. 8, the above two cases appear simultaneously.

5 Conclusion and future work

In this paper, we have proposed a novel framework for saliency detection called saliency bagging. The proposed framework gives robust performance across a variety of images and works in three steps. Firstly, existing saliency detection models are selected and a set of initial salience maps are computed. Secondly, an integrated binary map is found from the initial saliency maps by applying adaptive thresholding and majority voting. Lastly, the final saliency map is obtained using an integrated binary map and initial saliency maps using a novel proposed integration logic. The efficacy of the proposed framework is supported by performing extensive experiments on six publicly available datasets, and the performance is compared with 25 state-of-the-art methods which include 17 classic best-performing methods of the last decade, five existing selected, and three aggregation saliency methods. The experimental results indicate that the proposed method significantly enhances the quality of saliency maps. It outperforms all the compared classic and existing selected methods in terms of Precision, F-Measure, and MAE, while it is comparable to the best-performing methods in terms of Recall and ROC curve across all the six datasets. It is computationally efficient than all compared aggregation methods while performance is comparable on all datasets that validate its superiority. In future work, we shall look for improving the performance of salient object detection in terms of Recall and ROC curves using machine learning techniques.

Compliance with ethical standards

Conflict of interest All the authors declare that they have no conflict of interest.

References

1. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009. pp. 1597–1604. IEEE (2009)
2. Alpert, S., Galun, M., Basri, R., Brandt, A.: Image segmentation by probabilistic bottom-up aggregation and cue integration. In: IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR\'07. pp. 1–8. IEEE (2007)
3. Borji, A., Cheng, M.M., Jiang, H., Li, J.: Salient object detection: a benchmark. IEEE Trans. Image. Process. **12**(24), 5706–5722 (2015)
4. Borji, A., Sihite, D.N., Itti, L.: Salient object detection: a benchmark. In: Proceedings of the 12th European Conference on Computer Vision-Volume Part II, pp. 414–429. Springer (2012)
5. Breiman, L.: Bagging predictors. Mach. Learn. **24**(2), 123–140 (1996)
6. Chen, T., Cheng, M.M., Tan, P., Shamir, A., Hu, S.M.: Sketch2photo: internet image montage. ACM Trans. Graph. (TOG) **28**(5), 124 (2009)
7. Cheng, M.M., Mitra, N.J., Huang, X., Hu, S.M.: Salientshape: group saliency in image collections. Vis. Comput. **30**(4), 443–453 (2014)
8. Cheng, M.M., Mitra, N.J., Huang, X., Torr, P.H., Hu, S.M.: Global contrast based salient region detection. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 569–582 (2015)
9. Chia, A.Y.S., Zhuo, S., Gupta, R.K., Tai, Y.W., Cho, S.Y., Tan, P., Lin, S.: Semantic colorization with internet images. ACM Trans. Gr. (TOG) **30**(6), 156 (2011)
10. Donoser, M., Urschler, M., Hirzer, M., Bischof, H.: Saliency driven total variation segmentation. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 817–824. IEEE (2009)
11. Duan, L., Wu, C., Miao, J., Qing, L., Fu, Y.: Visual saliency detection by spatially weighted dissimilarity. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 473–480. IEEE (2011)
12. Erdem, E., Erdem, A.: Visual saliency estimation by nonlinearly integrating features using region covariances. J. Vis. **13**(4), 11–11 (2013)
13. GOFERMAN, S.: Context-aware saliency detection. In: Proceedings of IEEE Conference Computer Vision and Pattern Recognition, 2010, pp. 2376–2383 (2010)
14. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: Advances in Neural Information Processing Systems, pp. 545–552 (2007)
15. He, S., Lau, R.W., Liu, W., Huang, Z., Yang, Q.: Supercnn: a superpixelwise convolutional neural network for salient object detection. Int. J. Comput. Vis. **115**(3), 330–344 (2015)
16. Hou, X., Zhang, L.: Saliency detection: a spectral residual approach. In: IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07. pp. 1–8. IEEE (2007)
17. Huang, F., Qi, J., Lu, H., Zhang, L., Ruan, X.: Salient object detection via multiple instance learning. IEEE Trans. Image Process. **26**(4), 1911–1922 (2017)
18. Itti, L., Koch, C.: Computational modelling of visual attention. Nat. Rev. Neurosci. **2**(3), 194 (2001)
19. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. **20**(11), 1254–1259 (1998)
20. Jiang, B., Zhang, L., Lu, H., Yang, C., Yang, M.H.: Saliency detection via absorbing markov chain. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1665–1672 (2013)
21. Jiang, H., Wang, J., Yuan, Z., Wu, Y., Zheng, N., Li, S.: Salient object detection: a discriminative regional feature integration approach. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2083–2090 (2013)
22. Jiang, P., Ling, H., Yu, J., Peng, J.: Salient region detection by ufo: uniqueness, focusness and objectness. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1976–1983 (2013)
23. Kanan, C., Tong, M.H., Zhang, L., Cottrell, G.W.: Sun: top-down saliency using natural statistics. Vis. Cognit. **17**(6–7), 979–1003 (2009)
24. Klein, D.A., Frintrop, S.: Center-surround divergence of feature statistics for salient object detection. In: 2011 International Conference on Computer Vision, pp. 2214–2219. IEEE (2011)
25. Le Meur, O., Liu, Z.: Saliency aggregation: Does unity make strength? In: Asian Conference on Computer Vision, pp. 18–32. Springer (2014)

26. Li, G., Xie, Y., Lin, L., Yu, Y.: Instance-level salient object segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2386–2395 (2017)

27. Li, G., Yu, Y.: Visual saliency based on multiscale deep features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5455–5463 (2015)

28. Li, H., Lu, H., Lin, Z., Shen, X., Price, B.: Inner and inter label propagation: salient object detection in the wild. IEEE Trans. Image Process. 24(10), 3176–3186 (2015)

29. Li, R., Cai, J., Zhang, H., Wang, T.: Aggregating complementary boundary contrast with smoothing for salient region detection. Vis. Comput. 33(9), 1155–1167 (2017)

30. Li, X., Lu, H., Zhang, L., Ruan, X., Yang, M.H.: Saliency detection via dense and sparse reconstruction. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2976–2983 (2013)

31. Li, X., Zhao, L., Wei, L., Yang, M.H., Wu, F., Zhuang, Y., Ling, H., Wang, J.: Deepsaliency: multi-task deep neural network model for salient object detection. IEEE Trans. Image Process. 25(8), 3919–3930 (2016)

32. Li, Y., Hou, X., Koch, C., Rehg, J.M., Yuille, A.L.: The secrets of salient object segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 280–287 (2014)

33. Liu, G.H., Yang, J.Y.: Exploiting color volume and color difference for salient region detection. IEEE Trans. Image Process. 28(1), 6–16 (2019)

34. Liu, H., Zhang, L., Huang, H.: Web-image driven best views of 3d shapes. Vis. Comput. 28(3), 279–287 (2012)

35. Liu, T., Sun, J., Zheng, N.N., Tang, X., Shum, H.Y.: Learning to detect a salient object. In: IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07. pp. 1–8. IEEE (2007)

36. Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.Y.: Learning to detect a salient object. IEEE Trans. Pattern Anal. Mach. Intell. 33(2), 353–367 (2011)

37. Lu, Y., Zhou, K., Wu, X., Gong, P.: A novel multi-graph framework for salient object detection. Vis. Comput. (2019). https://doi.org/10.1007/s00371-019-01637-2

38. Mai, L., Niu, Y., Liu, F.: Saliency aggregation: A data-driven approach. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1131–1138 (2013)

39. Margolin, R., Tal, A., Zelnik-Manor, L.: What makes a patch distinct? In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1139–1146 (2013)

40. Murray, N., Vanrell, M., Otazu, X., Parraga, C.A.: Saliency estimation using a non-parametric low-level vision model. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 433–440. IEEE (2011)

41. Ninassi, A., Le Meur, O., Le Callet, P., Barba, D.: Does where you gaze on an image affect your perception of quality? Applying visual attention to image quality metric. In: IEEE International Conference on Image Processing, 2007. ICIP 2007. vol. 2, pp. II–169. IEEE (2007)

42. Perazzi, F., Krähenbühl, P., Pritch, Y., Hornung, A.: Saliency filters: contrast based filtering for salient region detection. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 733–740. IEEE (2012)

43. Qin, Y., Feng, M., Lu, H., Cottrell, G.W.: Hierarchical cellular automata for visual saliency. Int. J. Comput. Vis. 126(7), 751–770 (2018)

44. Rahtu, E., Kannala, J., Salo, M., Heikkilä, J.: Segmenting salient objects from images and videos. In: Computer Vision–European Conference on Computer Vision 2010, pp. 366–379. Springer (2010)

45. Rutishauser, U., Walther, D., Koch, C., Perona, P.: Is bottom-up attention useful for object recognition? In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. vol. 2, pp. II–II. IEEE (2004)

46. Seo, H.J., Milanfar, P.: Static and space-time visual saliency detection by self-resemblance. J. Vis. 9(12), 15–15 (2009)

47. Shi, J., Yan, Q., Xu, L., Jia, J.: Hierarchical image saliency detection on extended cssd. IEEE Trans. Pattern Anal. Mach. Intell. 38(4), 717–729 (2016)

48. Stalder, S., Grabner, H., Van Gool, L.: Dynamic objectness for adaptive tracking. In: Asian Conference on Computer Vision, pp. 43–56. Springer (2012)

49. Tavakoli, H.R., Rahtu, E., Heikkilä, J.: Fast and efficient saliency detection using sparse sampling and kernel density estimation. In: Scandinavian Conference on Image Analysis, pp. 666–675. Springer (2011)

50. Tu, W.C., He, S., Yang, Q., Chien, S.Y.: Real-time salient object detection with a minimum spanning tree. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2334–2342 (2016)

51. Wang, L., Wang, L., Lu, H., Zhang, P., Ruan, X.: Salient object detection with recurrent fully convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. 41(7), 1734–1746 (2018)

52. Wei, Y., Wen, F., Zhu, W., Sun, J.: Geodesic saliency using background priors. In: European Conference on Computer Vision, pp. 29–42. Springer (2012)

53. Xu, Y., Hong, X., Porikli, F., Liu, X., Chen, J., Zhao, G.: Saliency integration: an arbitrator model. IEEE Trans. Multimed. 21(1), 98–113 (2018)

54. Yan, Q., Xu, L., Shi, J., Jia, J.: Hierarchical saliency detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1155–1162 (2013)

55. Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M.H.: Saliency detection via graph-based manifold ranking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3166–3173 (2013)

56. Yang, J., Yang, M.H.: Top-down visual saliency via joint crf and dictionary learning. IEEE Trans. Pattern Anal. Mach. Intell. 39(3), 576–588 (2017)

57. Yuan, Y., Li, C., Kim, J., Cai, W., Feng, D.D.: Reversion correction and regularized random walk ranking for saliency detection. IEEE Trans. Image Process. 27(3), 1311–1322 (2018)

58. Zhang, J., Sclaroff, S.: Exploiting surroundedness for saliency detection: a boolean map approach. IEEE Trans. Pattern Anal. Mach. Intell. 38(5), 889–902 (2016)

59. Zhang, J., Sclaroff, S., Lin, Z., Shen, X., Price, B., Mech, R.: Minimum barrier salient object detection at 80 fps. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1404–1412 (2015)

60. Zhang, L., Tong, M.H., Marks, T.K., Shan, H., Cottrell, G.W.: Sun: a bayesian framework for saliency using natural statistics. J. Vis. 8(7), 32–32 (2008)

61. Zhao, R., Ouyang, W., Li, H., Wang, X.: Saliency detection by multi-context deep learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1265–1274 (2015)

62. Zhu, W., Liang, S., Wei, Y., Sun, J.: Saliency optimization from robust background detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2814–2821 (2014)

Vivek Kumar Singh received his M.Tech (Software Engineering) from Motilal Nehru National Institute of Technology Allahabad, Allahabad, India. He is currently pursuing Ph.D. from National Institute of Technology Uttarakhand, India. His current research areas are saliency detection, object detection, computer vision, image processing, and pattern recognition.

Nitin Kumar received his Doctor of Philosophy and Master of Technology from School of Computer and System Sciences, Jawaharlal Nehru University, New Delhi. He is currently working as Assistant Professor at National Institute of Technology, Uttarakhand, India. His current research includes visual attention modeling, pattern recognition, face recognition, and image processing.