Journal of Ambient Intelligence and Humanized Computing (2023) 14:7173–7188
https://doi.org/10.1007/s12652-021-03570-6

ORIGINAL RESEARCH

Object detection and segmentation by composition of fast fuzzy C-mean clustering based maps

Mehmood Nawaz[1,3] · Rizwan Qureshi[2] · Mansoor Ali Teevno[3,4] · Ali Raza Shahid[3,5]

Received: 5 March 2021 / Accepted: 17 October 2021 / Published online: 24 January 2022

Abstract
The extraction of salient objects from a cluttered background without any prior knowledge is a challenging task in salient object detection and segmentation. A salient object can be detected from the uniqueness, rarity, or unproductivity of the salient regions in an image. However, an object with a similar color appearance may have a marginal visual divergence that is even difficult for the human eyes to recognize. In this paper, we propose a technique which compose and fuse the fast fuzzy c-mean (FFCM) clustering saliency maps to separate the salient object from the background in the image. To be specific, we first generate the maps using FFCM clustering, that contain specific parts of the salient region, which are composed later by using the Porter–Duff composition method. Outliers in the extracted salient regions are removed using a morphological technique in the post-processing step. To extract the final map from the initially constructed blended maps, we use a fused mask, which is the composite form of color prior, location prior, and frequency prior. Experiment results on six public data sets (MSRA, THUR-15000, MSRA-10K, HKU-IS, DUT-OMRON, and SED) clearly show the efficiency of the proposed method for images with a noisy background.

Keywords Salient object · Clustering · Mask construction · Saliency maps · Object segmentation

1 Introduction

The object that attracts the human attention and has a unique spatial position in an image, is a salient object. Recognizing a salient object in a cluttered image is a very complex task. Due to the higher resolution in the retina center, the human eye orients the center of its gaze typically to the spatial area of the visual scene. The salient object is characterized by the contrast of an image in the super-pixel plane (i.e. color, orientation, or intensity), that is the most attractive factor for human vision system. Example of salient object segmentation is shown in Fig. 1.

The saliency detection and segmentation is helpful in different multimedia tasks such as image retrieval (Gao et al. 2015), adaptive image display (Chen et al. 2003), image segmentation (Liu et al. 2010; Azaza et al. 2018; Badoual et al. 2019), content-aware image editing (Ding and Tong 2010), video surveillance (Ding and Tong 2010; Nawaz et al. 2019; Sokhandan and Monadjemi 2018), image compression (Liu et al. 2014), facial expression recognition (Shahid et al. 2020a; Nawaz and Yan 2020) and image classification (Murabito et al. 2018).

Usually, local and global contrast models have been discussed in the literature. With a local contrast model, we compute the saliency map by comparing the characteristics of each region with the adjacent region. The global contrast refers to the difference of a region with global regions as well as local regions. The saliency models are of two types; (1) saliency detection and (2) human fixation. The saliency detection models are frequently used to identify the salient object that utilizes saliency information for assessment for computer vision applications such as content-aware image resizing and image meditation. The human fixation saliency models are expected to be human

✉ Mehmood Nawaz
mnawaz4-c@my.cityu.edu.hk

Rizwan Qureshi
engr.rizwanqureshi786@gmail.com

[1] Department of Biomedical Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong

[2] FAST School of Computing, National University of Computer and Emerging Sciences, Karachi Campus, Islamabad, Pakistan

[3] Department of Electrical Engineering, City University of Hong Kong, Kowloon, Hong Kong

[4] Mehran University of Engineering and Technology, Jamshoro, Pakistan

[5] Department of Electrical and Computer Engineering, COMSATS University Islamabad, Pakistan

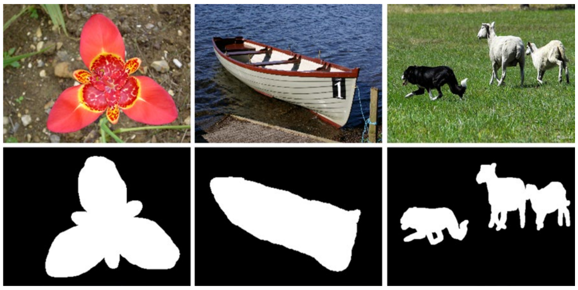


Fig. 1 Examples of salient object detection. The first row shows the original images and second row shows the extracted salient objects

obsession patterns to identify the salient region accurately. The generated maps in these models show the valuable difference in features due to their different purpose in saliency detection. For example, salient detection models produce smooth allied areas, while the fixation models generally produce blob-like salient regions. The visual saliency detection from the complex background poses challenges in the real applications of computer vision, according to the neurobiological and psychological definition (Goferman et al. 2011a) (see Fig. 2).

2 Related work

Many salient object detection and segmentation methods have been proposed in the last two decades that are based on superpixel, histogram, contrast, and feature graph (Azaza et al. 2018). Local and global contrast is often used to extract the salient regions from the noisy background. Specifically, local contrast models compare the characteristics of each region with each of the neighboring regions to measure the saliency chart. As a result, the high contrast pixels are called salient and the inner ground pixels belonging to the salient objects are not used in local contrast-based techniques. While in global contrast models, all pixels are considered as salient pixels, which retrieve better internal consistency. However, some background pixels may be considered as high saliency pixel in global contrast models. Some researchers combined both local and global models (Itti et al. 1998; Harel et al. 2007; Perazzi et al. 2012; Xue et al. 2011; Nawaz and Yan 2020; Shahid et al. 2020b; Goferman et al. 2011a; Zhu 2014; Rahtu et al. 2010). The computed contrast of combined models essentially indicates the uniqueness of an area in the image that is sensitive to the detection of salient objects. However, if some of the foreground objects are different, or the background is cluttered, a critical gap can accrue. On the other hand, some partial background areas

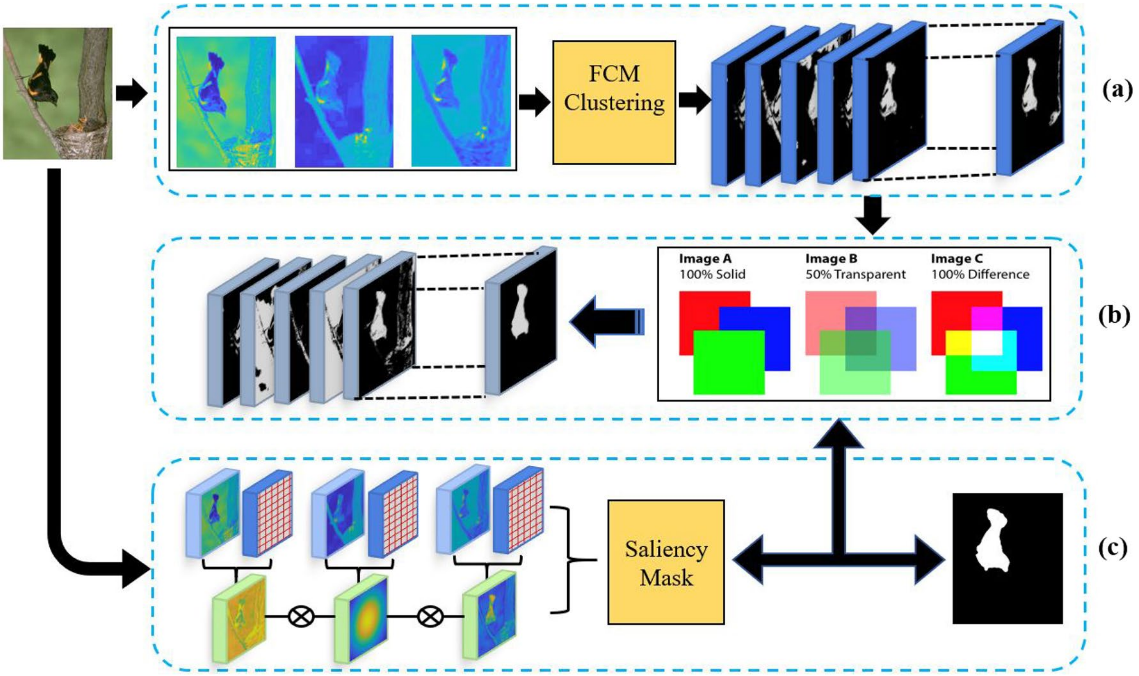


Fig. 2 The flow chart of the proposed object detection and segmentation method. A color image is converted into different maps, which are blended into each other by using the Porter–Duff method. **a** Shows the fast fuzzy c-mean clustering maps, **b** shows the maps blending and morphological filter process to remove outliers in the blended maps, and **c** shows the mask generation using combination of frequency, color, and location prior

may be considered as essential areas of the salient object, which causes the degradation of saliency detection during clustering. Different supervised and unsupervised saliency detection methods have been proposed to overcome these problems. More recently, the convolutional neural network-based models are used in computer vision applications. Xuan yang et al. (Xi et al. 2019) proposed a supervised approach of detection and segmentation that was presented on the basis of an efficient network of end-to-end saliency regression. This strategy did not achieve reasonable results to locate a salient region using pre-processing and post-processing practices. It employed VGG-16 as a backbone to train the model. It outperforms on a single object detection scene, however it breaks down to detect multiple objects. Wenguan et al. comprehensively reviewed the salient object detection problem and their solution in Wang et al. (2019). They efficiently differentiate the supervised and unsupervised SOD (salient object detection) models by using different statistical models with extensive experiments.

The unsupervised methods are based on low-level features (e.g., background prior, color, contrast). Zhang et al. (2013) presented a novel technique to evaluate the importance of the region by integrating three fundamental priors, color, frequency, and location. A visual patch attention saliency detection and segmentation method was suggested by Jian et al. (2014), where patches of information and direction are used to detect the salient object as neuronal signals. Barranco et al. (2014) has proposed a visual savings architecture with top-down care modulation for a field programmable gate array.

This model comprises a hardware architecture in real-time that combines FPGA into the robot system. It uses a robust biological operation to detect and segment the salient area in a color image. Li et al. (2013b) used contextual hypergraph modeling for the detection and segmentation of salient objects, which first retrieved the contextual characteristics of the image. A cost-sensitive support vector machine is then used to find the salient object. Wang et al. (2016) proposed an algorithm that uses information from the background to detect the salient object that uses the previous background information to make precise and stronger salient maps. Kim et al. (2017) suggested the Gaussian mixture models for fully automatic object segmentation as a saliency-based initialization. In this approach, the Gaussian mixture model determines the outer color of the object's background and foreground pixels, followed by average, covariance matrices, and coefficients of mixing that are further used in an image to locate the prominent pixels. A regularized random walk system for saliency detection and segmentation was proposed in Yuan et al. (2017). This approach first eliminates the adjacent boundary of the foreground superpixels, leading to a random walking ranking model that calculates the prior salience for each pixel in the image.

There are several disadvantages to the above-discussed approaches. For example, classifying the background clusters and foreground clusters in contrast and graph-based saliency technique is a challenging task. More specifically, one or more background clusters involve a portion (Yang et al. 2013b) of the foreground cluster (salient region) that can cause error and reduce the accuracy of the image segmentation. The salient maps of other methods of segmentation of images are binaries of different threshold techniques in initial saliency map process, but the proposed salient method did not use any threshold technique at this stage. Due to different threshold values (adaptive threshold, global threshold, fixed threshold, etc.) in saliency detection models, the efficiency of automated segmentation is reduced. The superpixel methods promote the boundary-based preprocessing saliency detection algorithm, as discussed above. The assignment of the same saliency value to all pixels in a patch in superpixel-based methods ignores certain important information, which can lead to a decrease in the image's visual quality. Based on this consideration, we have proposed a technique, which uses the Porter–Duff method (Duff 2015) to compose the salient object maps obtained by fast fuzzy c-mean (FFCM) clustering. The FFCM clustering technique is a powerful tool, which accurately clusters the large number of features points based on the histogram and intensity values of the color image as shown in Fig. 3. The histogram in an image usually refers to a pixel intensity in an image processing module. This histogram is a graph that shows how many pixels have similar values in an image. In the color image, different channel histograms (like red, green, and blue) can be retrieved. A histogram with three color channels is used to select a class membership at different threshold value. It calculates the chisquared distance (X) between pixels, which is defined as:

$$X^2(h_i, h_j) = \sum_{1}^{K} \frac{\left(h_i(k) - h_j(k)\right)^2}{h_i(k) + h_j(k)} \tag{1}$$

where i and j are the coordinates of pixel intensity and K is the total number of pixels. Benefiting from the clustering techniques, the obtained maps contain part of salient features (salient regions), which are composed using Porter–Duff composition method. Outliers in the extracted salient maps are removed using morphological techniques in a post-processing step.

This work is an extension of our conference presentation (Nawaz et al. 2019). The main extended contributions of this work are summarized as:

- We proposed a clustering-based segmentation technique, in which fast fuzzy c-mean clustering-based maps are blended by the baseline Porter–Duff composition method (Duff 2015). The blending method decides how the

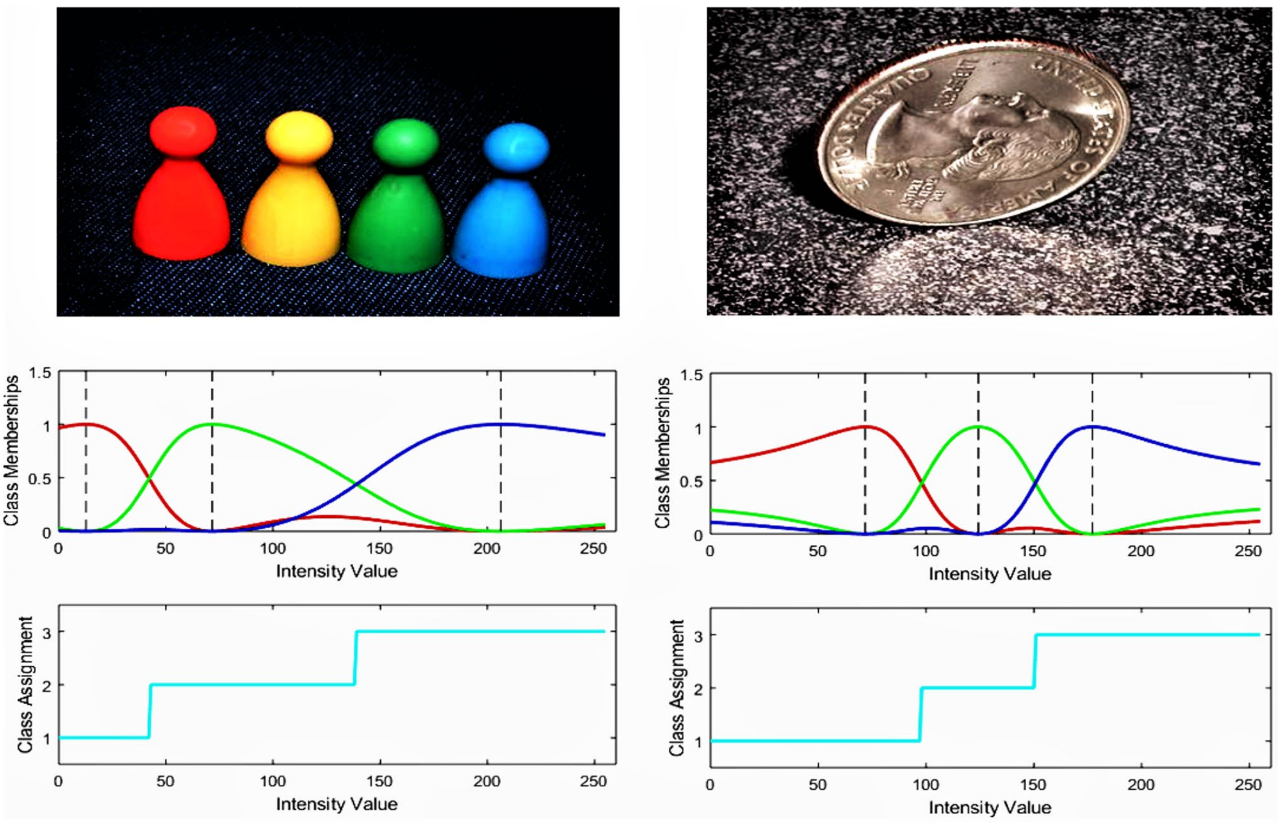


Fig. 3 An illustration of the histogram and the intensity values of an RGB image. The first row shows the original images, second and third row show the histogram and the intensity values of the image

colors communicate with each other from the foreground pixels as well as from the background pixels. The blending technique efficiently highlights the foreground pixels during composition.

- We proposed a multiscale morphological gradient reconstruction procedure to eliminate the boundary outliers in rough saliency maps. This technique helps to integrate the adjacent information of the salient pixel and decrease the number of different pixels in the saliency map.
- The proposed framework takes advantage in fusing the composed FFCM saliency maps efficiently in comparison to others. The FFCM based saliency model outperforms the superpixel-based image segmentation models because the superpixel models facilitate the pre-processing technique in saliency detection. In superpixel segmentation, all pixels in the nearest patches assign the same salience value and ignores some important information, which decreases the visual quality of the salient object in an image.
- We demonstrate the supremacy of the proposed method and conduct detailed experiments on six different benchmarks in comparison with thirteen different models to validate the efficiency of the proposed method.

3 Proposed framework

In this section, we introduce the proposed technique, including fast fuzzy c-mean clustering based saliency maps, composition and blending of saliency maps, mask construction using frequency prior, color prior, and location prior.

3.1 Saliency maps construction and composition

To find the membership maps, a FFCM clustering technique is used. By using histogram and image intensity level as illustrated in Fig. 2a, it splits several features into different clusters. The FFCM is a leading technique that is used to construct unsupervised data models. Instead of finding the absolute membership of a particular cluster data point, it decides the degree of membership (likelihood) that is discussed below.

3.2 Saliency maps

Let X be the input RGB image. Partition $X = \{x_1, x_2, x_3, x_4 \ldots x_n\}$ is a data set of n samples and assumes that each x_k sample is described by a set of f

characteristics. A X partition in C clusters is a series of X_i of X mutually disjoint subsets such as $X_i \bigcup ... \bigcup X_c = X$ and $X_i \bigcap X_j = \phi$ for any $i \neq j$. The clustering can be described by the $(c \times n)$ partition matrix U, the general term of which is $u_{ik} = 1$ if $x_k \in X_i$, and 0 otherwise, respectively. To get a partition matrix J , we generalized the objective function with $m > 1$ (m is coefficient of fuzziness):

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} U_{jk}^{m} \left\| x_i - V_j \right\|^2, \tag{2}$$

where N is the number of data points, C is the total number of clusters, V_j is the center of the j^{th} cluster, and i is the degree of membership of the i^{th} data points in the (x_i) cluster. The instruction $\|.\|$ denotes the proximity of the data points to the V_j center of the j cluster. It calculates the center of the cluster as:

$$V_j = \frac{\sum_{k=1}^{N} U_{jk}^{m} . x_k}{\sum_{k=1}^{N} U_{jk}^{m}}, \tag{3}$$

where m is the fuzziness coefficient. The membership maps U_{jk} of point k in cluster j is calculated as:

$$U_{jk} = \frac{1}{\sum_{j=1}^{C} \left(\frac{\|x_i - V_j\|}{\|x_i - V_k\|} \right)^{\frac{2}{m-1}}}. \tag{4}$$

The number of initial saliency maps depends on the total number of clusters. Let X be an RGB color image whose size is $M \times N \times K$, where M and N are the height and width of the image respectively, and K is the total number of color channels. Suppose C is the total number of classes/clusters in an image then the total number of maps can be calculated as:

$$B_{map} = K \times C. \tag{5}$$

In natural color images, we found that the optimal value of K is 3 through many experiments, which produces the best final results. For $K = 3$, the total number of saliency maps will be 9. In Fig. 4, different membership maps of *Baby* images are generated by using fast fuzzy c-mean clustering based technique that relies on image histogram and intensity values. Figure 4a shows the nine different saliency maps of baby image with $Source - In$ blending mode. Figure 4b shows the saliency maps with $Destination - In$ blending mode. Figure 4c shows the saliency maps with $Source - Atop$ blending mode and Fig. 4d shows the saliency maps of baby image with $Destination - Atop$ blending mode. Figure 5 shows the matching score matrix that is used to differentiate the good saliency maps.

3.3 Maps composition and blending

The Porter–Duff method contains two basic steps; composition and blending Duff (2015). The method of integrating the graphic elements of the foreground with the graphic elements of the background in maps is known as composition. The blending process is defined, how different colors communicate with each other from the foreground and the background graphic element. In composition, the resulting color is first determined from the graphical element of the foreground and the background maps, and then the foreground color is replaced with the resulting color using a particular composition operator. The composition of Porter–Duff is a pixel-based model in which two maps communicate and generate the final salient map (source and destination), as shown in Fig. 2c. Blending is the factor that measures map blending, where the foreground and the background component overlap. The colors are blended between the background and the foreground pixels. The foreground aspect is compounded with the background pixels after blending, as shown in Fig. 2b. In general, There are 12 distinct operators of the Porter–Duff composition that have a different combination of source to destination used. Table 1 only addresses four composition operators, of which the alpha values of the source (foreground) and background pixels are the as and ab. The f_a and f_b are the fractional terms of source and destination map. The source and background maps color are presented by C_s and C_b. The output pixel values of saliency maps are C_o and A_o for the mixing mode. The entire mixing process is basically carried out in one stage. In the proposed method, the following blending modes are used.

3.3.1 Normal blend mode

The default blending mode does not specify blending. The blending formula only selects the foreground, in which the blending function of the background pixels $B(C_b, C_s)$ is defined as:

$$B(C_b, C_s) = C_s. \tag{6}$$

3.3.2 Multiply blend mode

In multiple blending, the foreground pixel color replaces the background pixel color. The resulting map is always as dark as either the foreground or the background color at least. Here, the saliency maps are combination of both foreground and background maps. So, the resultant map has color gray in this blending mode.

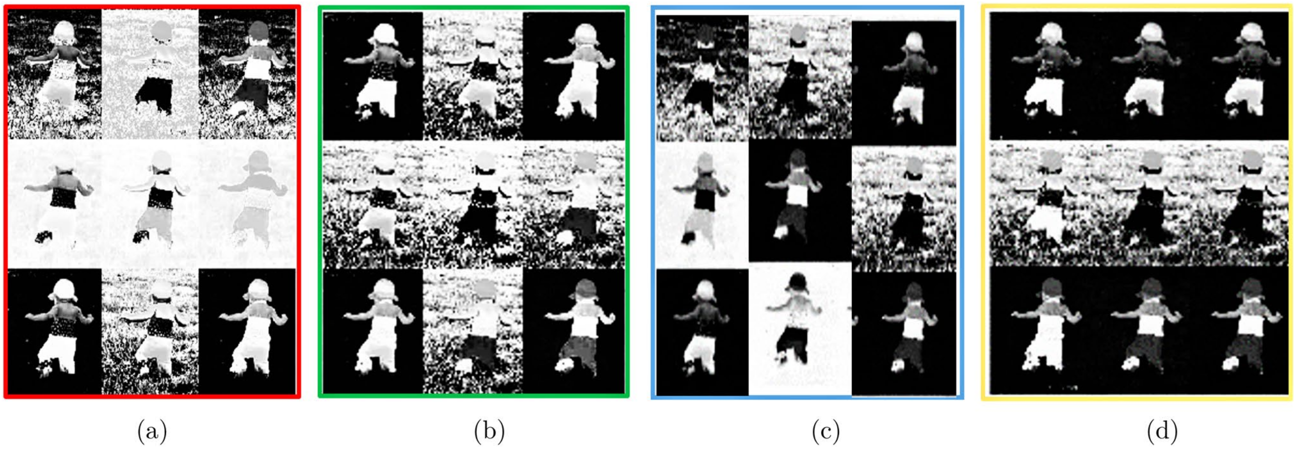
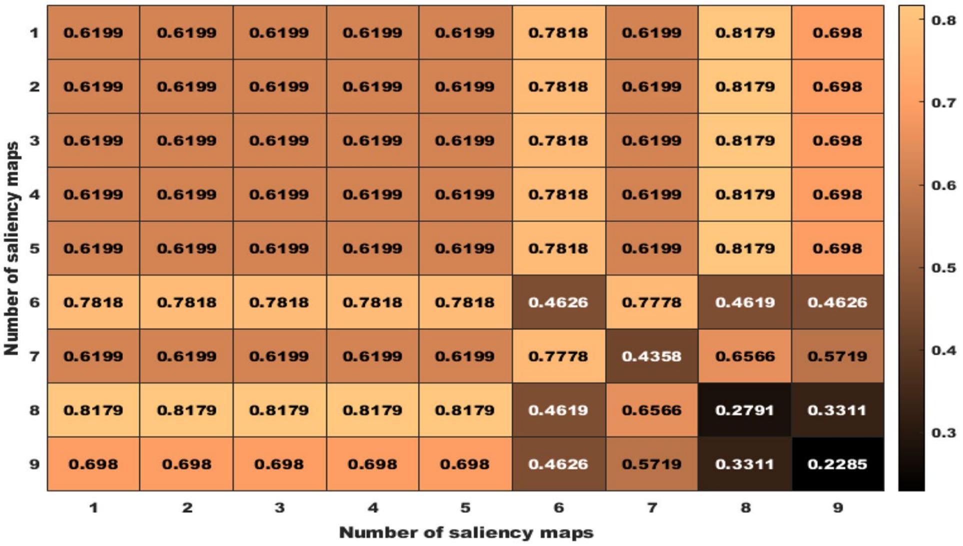


Fig. 4 The results of four blending modes of nine baby saliency maps. **a** Is “Source In” blend mode, where foreground pixels overlap the background pixels and replace the background, **b** is “Destination In” blend mode, where background pixels overlap the foreground pixels and replace the foreground, **c** is “Source Atop” blend mode, where the foreground pixels overlap the background pixels and keep both overlap pixels and background, and **d** is “Destination Atop” blend mode, where the background pixels overlap the foreground pixels and keep both overlap pixels and foreground

Fig. 5 An illustration of the similarity matrix of nine baby images. The maximum similarity values are shown in light orange color. The proposed method choose the high values maps and fused them together. The final fusion process is described in Eq. (16)



	1	2	3	4	5	6	7	8	9
1	0.6199	0.6199	0.6199	0.6199	0.6199	0.7818	0.6199	0.8179	0.698
2	0.6199	0.6199	0.6199	0.6199	0.6199	0.7818	0.6199	0.8179	0.698
3	0.6199	0.6199	0.6199	0.6199	0.6199	0.7818	0.6199	0.8179	0.698
4	0.6199	0.6199	0.6199	0.6199	0.6199	0.7818	0.6199	0.8179	0.698
5	0.6199	0.6199	0.6199	0.6199	0.6199	0.7818	0.6199	0.8179	0.698
6	0.7818	0.7818	0.7818	0.7818	0.7818	0.4626	0.7778	0.4619	0.4626
7	0.6199	0.6199	0.6199	0.6199	0.6199	0.7778	0.4358	0.6566	0.5719
8	0.8179	0.8179	0.8179	0.8179	0.8179	0.4619	0.6566	0.2791	0.3311
9	0.698	0.698	0.698	0.698	0.698	0.4626	0.5719	0.3311	0.2285

$$B(C_b, C_s) = C_b \times C_s. \qquad (7)$$

3.3.3 Screen blend mode

The color of the foreground pixels and the color of the background pixels are multiplied and the result is then complemented. The resulting map is always divided into two integral colors. The white screening of any color produces white in this blending mode.

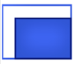
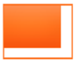
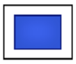
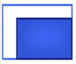
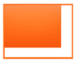
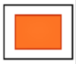
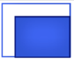
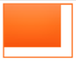
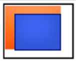
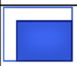
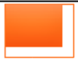
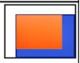
$$B(C_b, C_s) = 1 - \left[(1 - C_b) \times (1 - C_s)\right]. \qquad (8)$$

$$B(C_b, C_s) = C_b + C_s - (C_b \times C_s). \qquad (9)$$

3.3.4 Overlay blend mode

Depending upon the background pixel value, it multiplies or screens the foreground and the background maps. Due to retaining highlights and shadows, the foreground pixel

Table 1 The mathematical and graphical representation of four compositing methods

Methods	Mathematical Expression	Foreground	Background	Results
Source In	$F_a = \alpha b, \quad F_b = 0$ $C_o = \alpha s \times C_s \times \alpha b$ $A_o = \alpha s \times \alpha b$			
Destination In	$F_a = 0, \quad F_b = \alpha s$ $C_o = \alpha b \times C_b \times \alpha s$ $A_o = \alpha b \times \alpha s$			
Source Atop	$F_a = \alpha b, \quad F_b = 1 - \alpha s$ $C_o = \alpha s \times C_s \times \alpha b + \alpha b \times C_b \times (1 - \alpha s)$ $A_o = \alpha s \times \alpha b \times (1 - \alpha s)$			
Destination Atop	$F_a = 1 - \alpha b, \quad F_b = \alpha s$ $C_o = \alpha s \times C_s \times (1 - \alpha b) + \alpha b \times C_b \times \alpha s$ $A_o = \alpha s \times (1 - \alpha b) \times \alpha s)$			

In "Source In", foreground pixels overlap the background pixels and replace the background. In "Destination In", background pixels overlap the foreground pixels and replace the foreground. The "Source Atop", foreground pixels overlap the background pixels and keep both overlap pixels and background. The "Destination Atop", background pixels overlap the foreground pixels and keep both overlap pixels and foreground

overlays the background pixel. The background pixel values are not replaced, but are combined to represent the lightness or darkness of the background with the foreground pixel values.

$$B(C_b, C_s) = HardLight(C_b, C_s). \tag{10}$$

4 Mask construction and salient region extraction

We used a mask obtained by the SDSP (Saliency Detection Simple Priors) process (Liu et al. 2014), to find the appropriate salient object from the blending maps array. It is a combination of three basic priors, such as frequency, color, and image location as shown in Fig. 2c.

4.1 Mask construction

We use the log-Gabor filter instead of DoG filter for band-pass filtering in a color image. Our decision has some good reasons. At first, we can build an arbitrary log-Gabor filter that has no DC component. Secondly, the log-Gabor filter's transfer function is extended to the high-frequency end, making it more capable of encoding natural images than other conventional band pass filters.

$$g(v) = exp\left\{ -\left(log \frac{\|v\|_2}{\omega} \right)^2 / 2\sigma_F^2 \right\}, \tag{11}$$

where ω and $v = (i,j) \subseteq \mathbb{R}^2$ are center frequency and coordinate in the frequency domain, respectively. The σ_F^2 handles the frequency bandwidth in the filter. The optimised parameter value of ω is 1/6 and σ_F^2 is 0.3. Let X denotes an RGB image. In the first location, we are splitting it in three colors (*CIELab*): C_l, C_a and C_b. The prior frequency $P_f(X)$ is known as:

$$P_f(X) = \left((C_l * g) + (C_a * g) + (C_b * g) \right)^{\frac{1}{2}}(X), \tag{12}$$

where g is the Log-Gabor filter and $(*)$ shows the convolution function between the filter and all channels of the color image. All color channel are represented by (i.e. *Cl*, *Ca*, and *Cb*). Human visual systems are more susceptible to colors like red and yellow than colors like green and blue. Therefore, the salient color is known as:

$$P_c(X) = 1 - exp\left(-\frac{C_{an}^2(X) + C_{bn}^2(X)}{\sigma_c^2} \right), \tag{13}$$

where P_C and σ are the color priors and the parameter, respectively. The minimum value of the channel C_a is C_{an}, and the maximum value of the channel C_b is C_{bn}, which are defined as:

$$C_{an} = \frac{C_a(X) - min(a)}{max(a) - min(a)}, C_{bn} = \frac{C_b(X) - min(a)}{max(b) - min(b)}. \tag{14}$$

The object near the center is more appealing to eyes, the object is considered to be a salient object in the middle or near to the center of the image. Let C be the image center, then the saliency location in the image X under the Gaussian map is calculated as:

$$P_l(X) = exp\left(-\frac{\|X - C\|_2^2}{\omega_l^2} \right), \tag{15}$$

where, P_l indicates the position prior to the image and the location parameter is ω_l. By computing the three maps listed above, the final mask is represented as:

$$M = P_f(X) \times P_c(X) \times P_l(X). \tag{16}$$

The method used to measure the mask is shown in Fig. 2. As shown in the proposed flow map, the resulting mask is used to separate the appropriate output.

4.2 Salient region extraction

In the proposed method, two types of composition and blending are used. In the first type, the "Source Atop" composition operator with "Multiply" blend mode is used. In this mode, background pixels overlap the foreground pixels and then multiplies with overlap area of both foreground and background pixels, which is shown in Fig. 2b. In the second type, the "Source Atop" composition operator with "Screen" blend mode is used. In this mode, background pixels overlap the foreground pixels, the same as above, but the significant difference is the "Screen blend mode," which multiplies the complement of both overlap area in the background to find the final output. However, the second blending recovered the missing area created during the first blending. To remove the noisy areas from the final salient map, we use both composition and blending techniques. Some outliers are found in the salient maps, which are removed by using advanced morphological operations (closing and opening) (see Fig. 2b).

5 Experiments

We compared our method with thirteen different methods using six different data sets: MSRA (Liu et al. 2010), DUT-OMRON (Nawaz and Yan 2020), MSRA-10K Li et al. (2013a), THUR-15000 THUR (2013), HKU-IS (Nawaz and Yan 2020), and SED (Zhu 2014). These data sets are widely used to extract salient objects from color images. The performance of the proposed method and thirteen different methods, SIM (Nawaz et al. 2019), SUN (Li et al. 2013a), SEG (Itti et al. 1998), SeR (Seo and Milanfar 2009), CA (Goferman et al. 2011a), GR (Yang et al. 2013a), FES (Tavakoli et al. 2011), MC (Jiang et al. 2013), DSR (Li et al. 2013b), RBD (Zhu 2014), CYB (Chen et al. 2020), ResNet-50 (Goyal et al. 2019), and SDDF (Nawaz and Yan 2020) are shown in Figs. 7, 8, 9, 10 and 11.

5.1 Data sets

The MSRA, MSRA-10K, THUR-15K, and SED data sets are used to determine the method's performance. Both data sets include pixel-wise ground reality labelled by humans. The MSRA data set contains 5000 RGB images with a ground-mask of truth is created by Liu et al. (2010). This data collection is commonly used for the identification of salient objects. All pictures are single based object having noisy background. The MSRA-10K comprises the 10,000 pictures with the mask on the foreground (Li et al. 2013a). THUR-15000 Commonly used data set (THUR 2013) for 15,000 color images. It includes different sizes of low-contrast salient objects, which makes it very difficult for object detection. There are 100 color images with ground truth (Li et al. 2013a) in the SED data set. In this data set, all images have multiple and single salient artifacts and are labelled with pixel-wise ground truth.

5.2 Parameter settings

To maximize the performance of the proposed method, various parameters are used. To choose the optimized value of parameters is difficult some times. We have tested our method on a large number of images with optimized parameter values. In all experiments, nine different saliency maps of each color image are combined and blend one by one using porter-duff technique, as shown in Fig. 2b. To construct a such type of maps, we fixed the values of parameter C as mentioned in Eq. (5), and K is the number of color-image channels, having a value of three due to RGB image and C is the sum of clusters/classes indicated by Eq. (5). The optimized C value in all experiments is 3. In map blending, instead of 12, we use four distinct composition operators and blending modes, which are "Source Atop" with "Multiply Mode" and "Source Atop" with "Screen Mode" operators.

5.3 Performance comparisons

The qualitative and quantitative analysis of the proposed method and other methods are useful to discern accuracy with the six commonly used data sets. The results of all other methods can be obtained in this section by using publicly accessible codes. The similarity matching score matrix is shown in Fig. 5.

5.3.1 Quantitative comparison

The precision and recall rates suggested by Achanta et al. (2009) are the requirements for quantitative evaluation. The precision rate is specified as the ratio of the salient pixels to all ground truth pixels detected, and the ratio of the salient pixels to all ground truth pixels is reported correctly. The precision and recall curves are shown in Fig. 6. The proposed method is dominant in PR-curve in three data sets (MSRA, MSRA-10K, and SED) and has similar value in THUR-15000 data set. The PR-curve values find by different threshold values (0–255). In certain instances, precision and recall values are not enough to compare the results then the F-measure is found. The general F-measure is calculated by the weighted harmonic of precision and recall, which is defined as:

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

We used $\beta = 0.3$ as using in Achanta et al. (2009) to calculate the F-measure. The average precision, recall, and F-measure is shown in Fig. 7. Our approach has good results in the MSRA data set, where most of the images are single-object images with a clear background. Our approach has substantially better results than other methods with respect to the F-measure value. Furthermore, for MSRA, SED and MSRA-10K data sets, the accuracy and recall curves of our method are also high. The accuracy and recall values of the THUR-15000 dataset are similar to the RBD (Zhu 2014) and

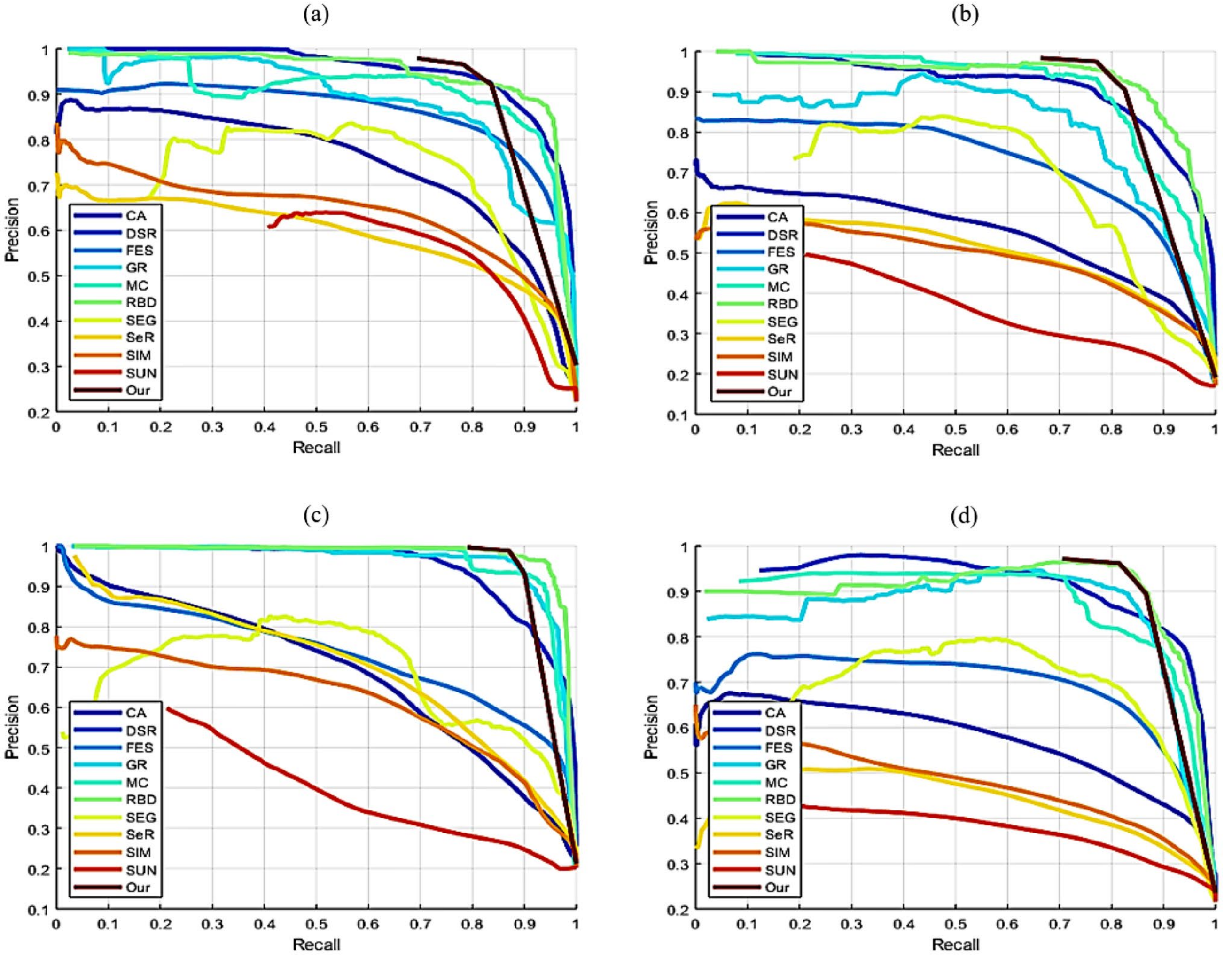


Fig. 6 The quantitative results from other approaches and proposed method in terms of the precision recall (PR) curve. **a** Shows the PR curve of the MSRA data set, **b** shows the PR curve for MSRA-10K data set, **c** shows the PR curve of the THUR-15000 data set, and **d** shows the PR curve of the SED data set. The proposed method has efficient precision to recall curve in all data sets

the DSR (Li et al. 2013b) method, but our method works better on this data set in terms of the F-measure. The proposed method obtained efficient PR curve results in a SED data collection, while the majority of images have complex and noisy backgrounds. We use the mean absolute error (*MAE*) benchmark to test our methodology to get a detailed comparison that tests the similarity of the ground truth map and the saliency maps. The average difference between ground truth pixels and pixel to the map of saliency ($S - map$) is called the MAE ($G\ truth$).

$$MAE = \frac{1}{WH} \times \sum_{x=1}^{W} \sum_{y=1}^{H} \left| S_{map} - G_{truth} \right|. \tag{18}$$

Figure 8 indicates the MAE value of all methods for four separate data sets. It shows the error between the saliency map and the ground truth map of the image. Figure 8a shows the MAE values of MSRA results, showing clearly that the proposed method's MAE values are less than other methods. The MAE values of the MSRA-10K dataset are shown in Fig. 8b. Our method has similar MAE value to RBD (Zhu 2014), but not so much as compared to other methods in this graph. Figure 8c shows the MAE value for all methods in THUR-15000 data set, where the MAE value in our system is low in comparison to other methods. The MAE of all methods in SED data set is shown by Fig. 8d. The proposed method in this data set has the lowest mean of absolute error compared to other methods. As the value of MAE increases, the effects are worse. So the overall results in terms of MAE values are good in the proposed method.

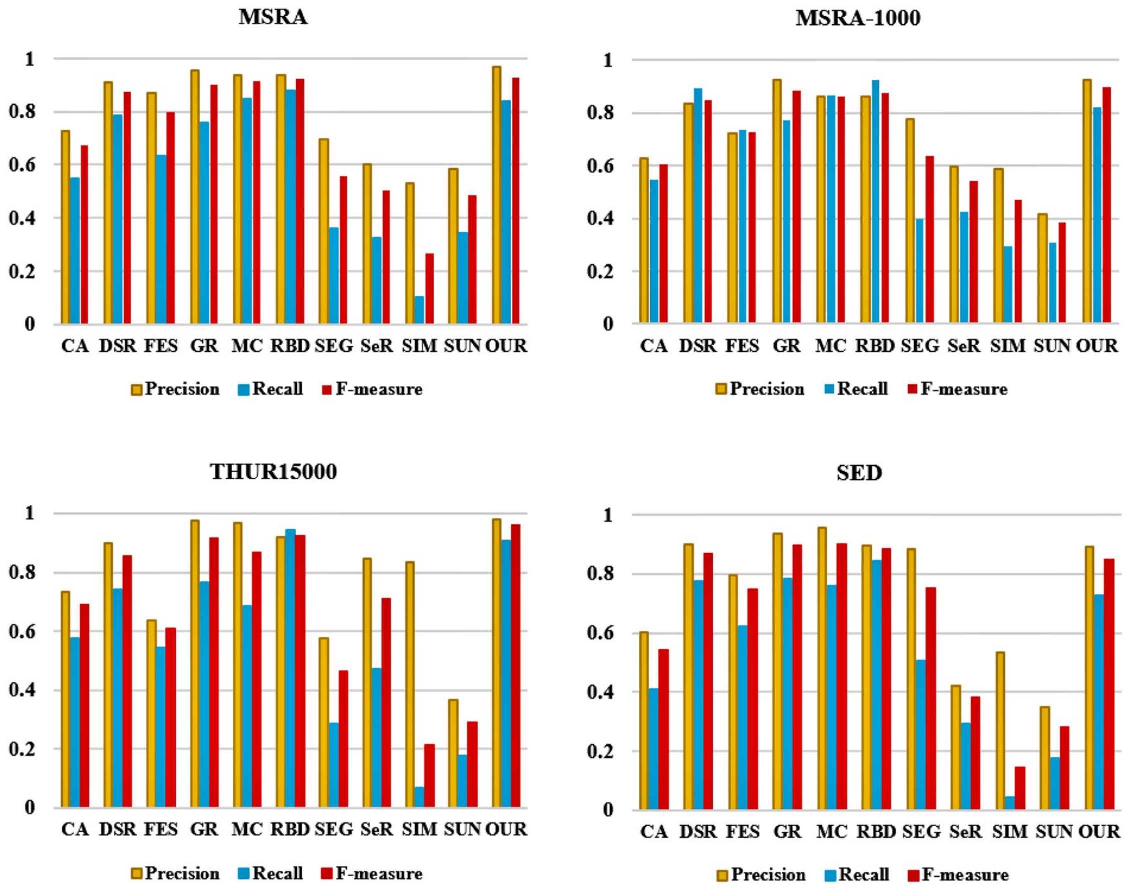


Fig. 7 Performance of other methods and proposed method in terms of precision, recall and F-measure graphs. The above left plot shows the average results for MSRA and the top right plot shows the average results for MSRA-10K; the bottom left plot shows the average THUR-15,000 results and the lower right plot shows the average SED results. The proposed approach in all data sets has been effective in terms of precision, recall and F-measure values

5.3.2 Qualitative comparison

Six widely used data sets are used for the detailed qualitative comparison between the proposed method and other methods as shown in the Figs. 9 and 10. In Fig. 9, the first column shows color RGB images, and the last column shows ground truth value of each color image. The performance of other methods is shown in columns 2–10. The results of columns 2–7 indicate that these methods were unable to carefully remove the salient object from the background, which could be influenced by the dense texture in the background of the image. The visual quality of columns 8–10 show that the MC (Jiang et al. 2013), RBD (Zhu 2014), and DSR (Li et al. 2013b) methods are comparatively better than SIM, SUN, SEG, SeR, and CA methods. If we compare the results of other methods in terms of visual quality, our method is significantly better, which is shown by the second last column of Fig. 9. Through this series of experiment, it is very easy to differentiate the visual comparison of the proposed method and other methods with a complex and noisy background. However, the proposed results have a high degree of resemblance to the ground truth, and it can uniformly highlight the salient object better as compared to other methods.

5.4 Implementation

The implementation of the proposed method, including training data sets and computational time of the proposed method, is explained in this section.

Training data sets Training data greatly influences the final behavior of the saliency detection models (Wang et al. 2019). We construct the model from six data sets (SED1, MSRA, MSRA-10K, THUR-15K, DUT-OMRON, and HKU-IS). These data sets contain both single and multiple objects. The MSRA-10K dataset is randomly used for training and sampling for validation. The proposed model also uses the method of saliency map composition that has been found to improve the performance of many visual tasks effectively. From an intuitive point of view for salient object detection tasks, the sharpness of the salient area has vital

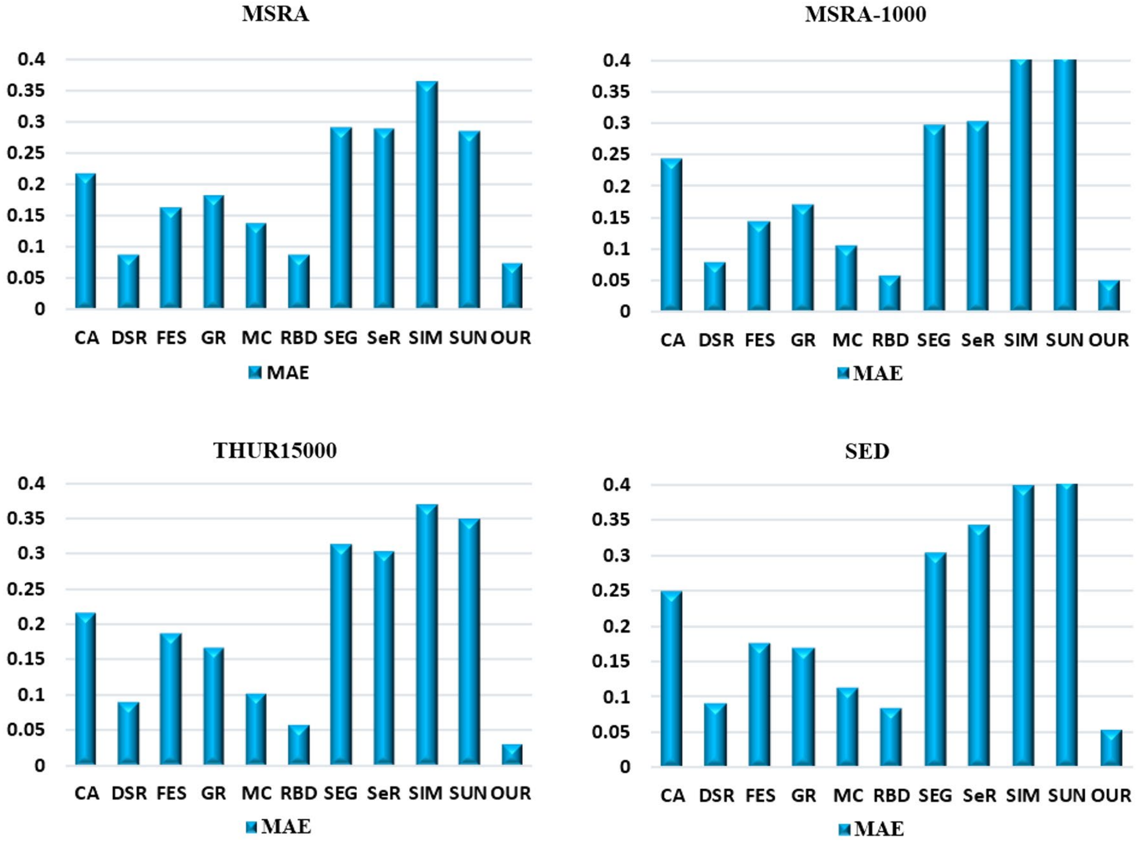


Fig. 8 Mean absolute error values between proposed method and other approaches. Top left plot represents the MAE results for MSRA data set, top right plot represents the MAE results for MSRA-10K data set, bottom left plot represents the MAE results for THUR-1500 data set, and bottom right plot represents the MAE results of SED data set. All plots show that the proposed method has the lowest MAE value for all data sets

importance in the image. Figure 11 shows object segmentation results on gray and remote sensing images. The results of third and fourth rows show that the proposed method also worked well on gray and remote sensing land images. We have executed our method on both single and multiple objects data sets of real images, which are shown in Tables 2 and 3, respectively. The quantitative results (average precision (AP), area under the curve (AUC), F-measure, and generalized F-measure) of 13 different techniques are shown in Tables 2 and 3. The mathematical form of these evaluation metrics is given in Nawaz and Yan (2020)

Running time MATLAB 2018b and Core i7 with a RAM of 8 GB are implementing the proposed method. We compare the average run time of the proposed method and the other thirteen methods for six different data sets as shown in Table 4. The size of an image is 400 × 300 in most of the data set. The running time of other methods is obtained by publicly accessible codes. Notice that our approach is implemented without code optimization on MATLAB, but C++ was used by other RBD, DSR, and MC approaches. The proposed method is faster than DSR, GR, MC, CA, but slightly slower than FES and RBD as shown in Table 4.

5.5 Failure cases

In certain cases, efficient results for single or multiple object detection could not be obtained by our proposed method. Figure 12 shows the failure cases of proposed method at six different data sets. In Fig. 12, columns 1 and 4 show the original images, columns 2 and 5 show the ground truth images, and columns 3 and 6 show the failure cases. These four images are taken from six different data sets. In the butterfly image, the ground truth map shows one butterfly, but the salient map has two butterflies, which clearly shows the failure of the proposed method, due to color similarities. In the cup image, the proposed method result is affected by object shadow and similar color appearance in the bottom. In the dog image, the background car color merges with dog color, which affects the object detection process in the proposed method. In the last image, there should be two salient objects, one is a butterfly and the second is a plant.

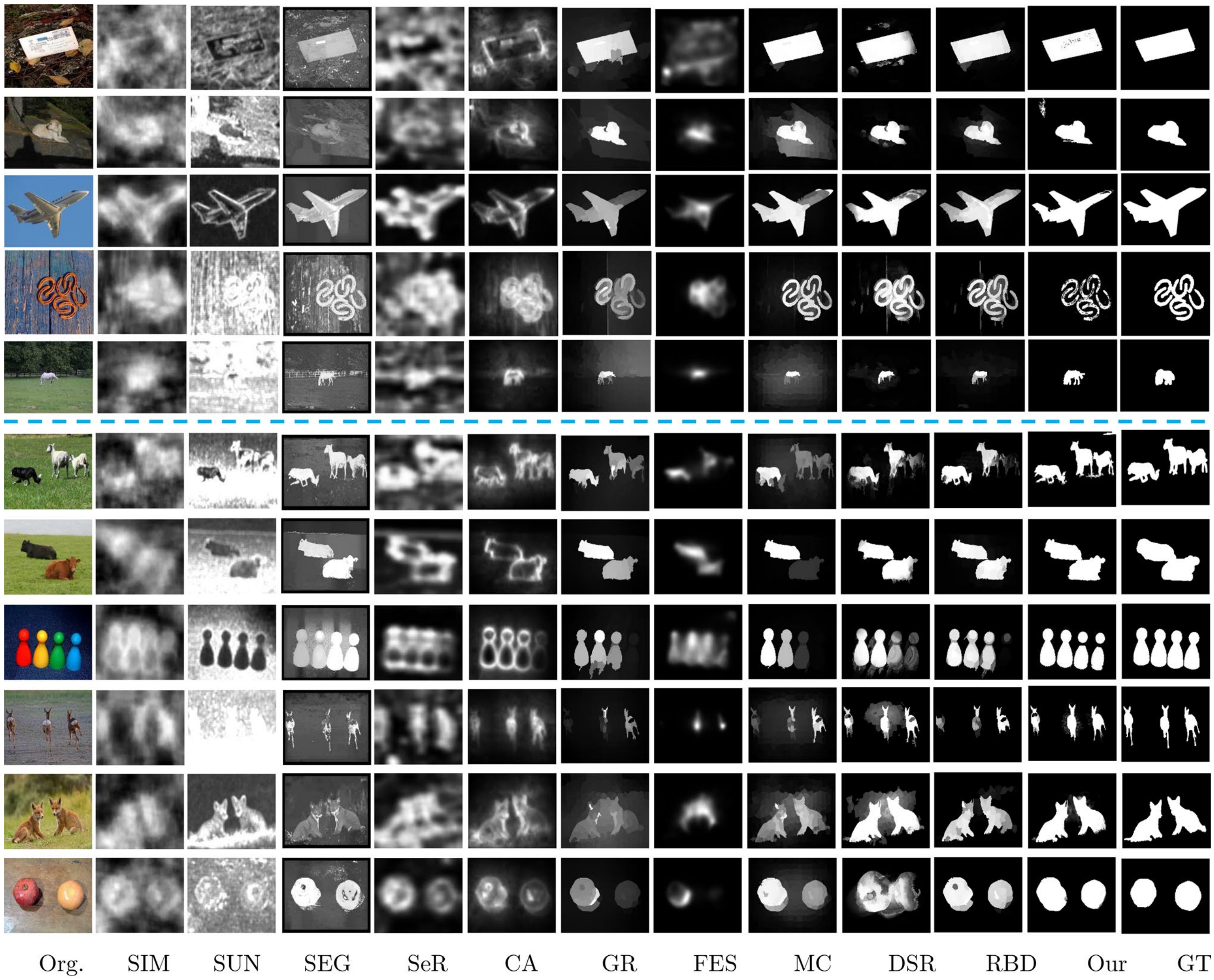


Fig. 9 Visual results on six different data sets. We divide the chosen images into two groups and each group has different type of salient object.The first five rows are the results of proposed method on single based object of images and the last six rows are the results of multiple object based images. Columns 2–11 show the results of different state of the art methods. The second last column shows the results of the proposed method. Last column shows the ground truth values of each data set. The visual efficiency of the proposed method on different data sets show the supremacy of the proposed method

According to the given ground truth map, the butterfly is only the salient object. In this image, our proposed method fails to detect the butterfly due to the high contrast of the plant image. The results of the failure cases, indicates, that in some low and related background color pictures, the proposed method is not to much efficient to extract the salient regions.

6 Conclusion

We proposed a clustering-based segmentation approach, where the fast fuzzy c-mean clustering maps are blended using the Porter–Duff (Duff 2015) composition method. In this method, one foreground pixels map is blended with the second background pixels map. The composite frequency prior, color prior, and location prior is used to select the final saliency map (Zhang et al. 2013) from the list of initially constructed saliency maps. The results of proposed method are effective and precise compared with other method, when the foreground and background pixels have same color appearance. The efficiency of the proposed method can be improved, through optimized context subtraction and effective morphological pixel-based techniques. The boundary smoothing techniques can also be used to improve the visual efficiency of the constructed saliency maps.

The goal of salient object detection is to develop a detection method based on the human visual perception model.

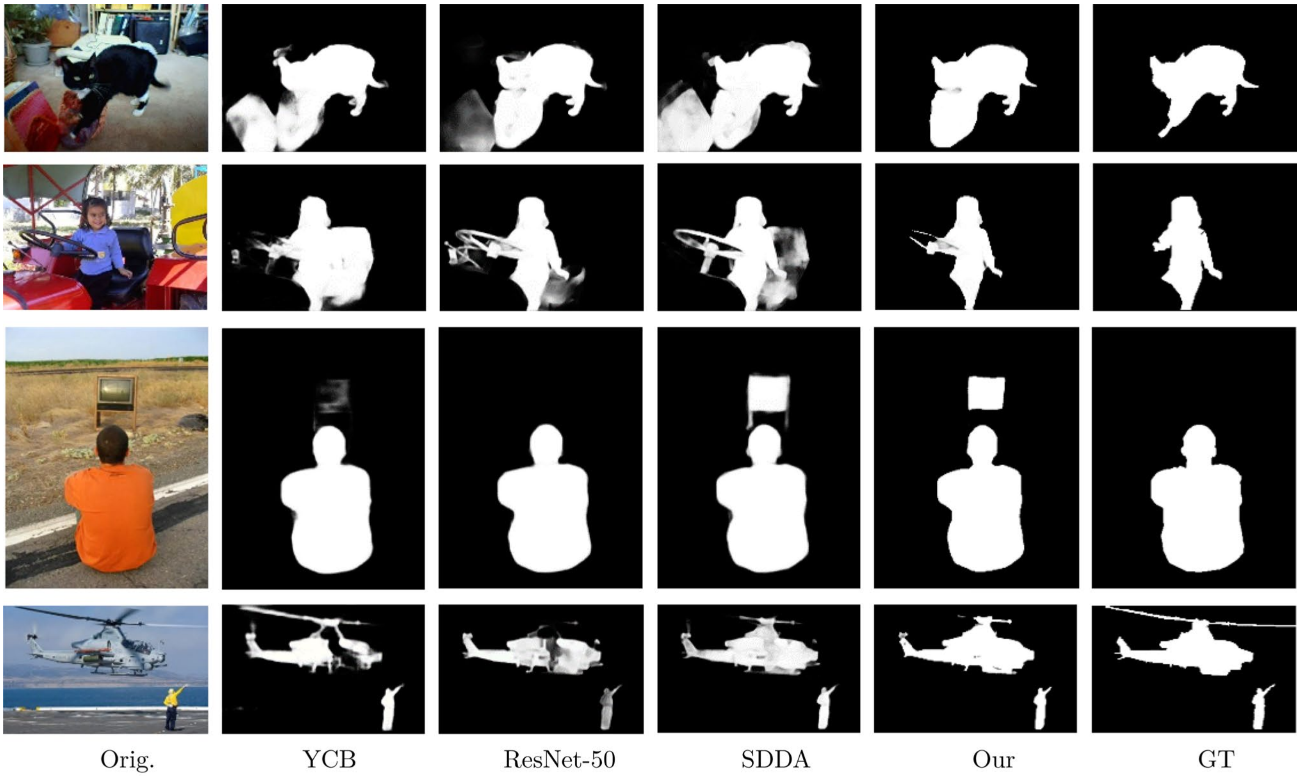


Fig. 10 Saliency detection results on DUT-OMRON and HKU-IS data sets. The columns 2–4 are the results of recent saliency detection methods and the column 5 shows the results of the proposed method. The ground truth values of these images are shown in the last column. The visual results in this figure show that the proposed method is very close to the ground truth as compared to other methods

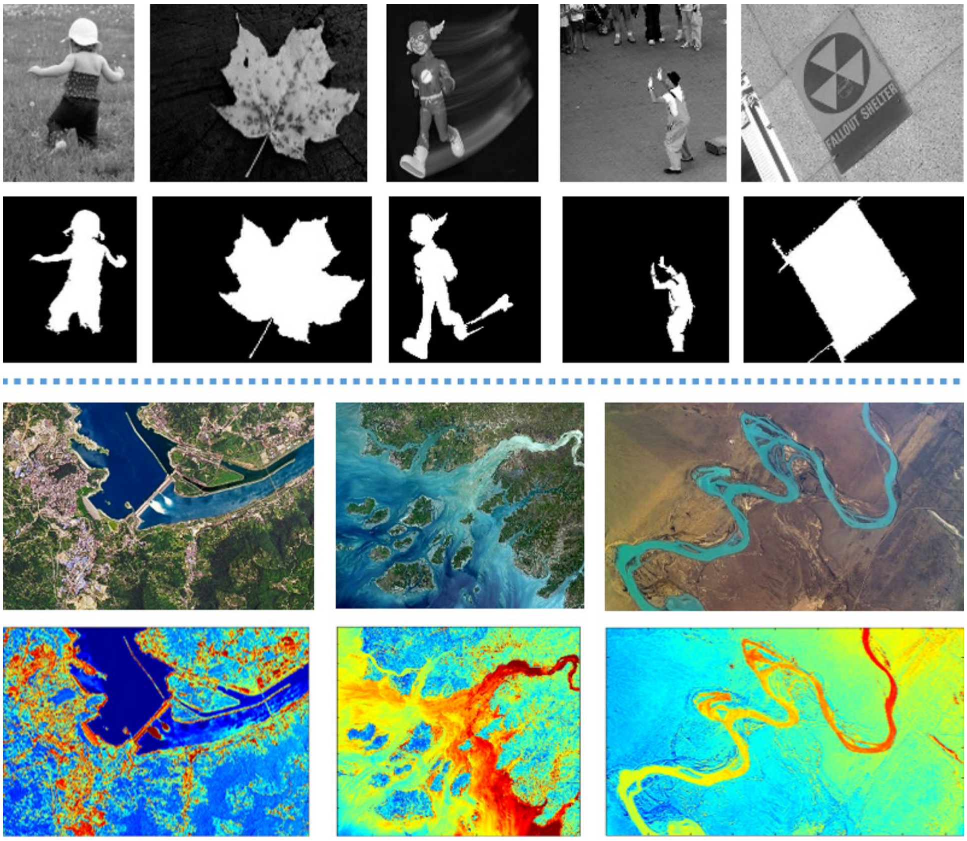


Fig. 11 Object segmentation results on gray and remote sensing images. First and second rows show the gray images and their results, respectively. Third and fourth rows show the remote sensing land images, and their segmentation results, respectively. These results show that our proposed method has efficient performance on these type of data sets

Table 2 Saliency detection results on three single object-based benchmarks

Methods	SED1				MSRA				MSRA-10K/ECSSD			
	AP	AUC	F_m	F_m^w	AP	AUC	F_m	F_m^w	AP	AUC	F_m	F_m^w
SIM	0.6419	0.4692	0.5269	0.4609	0.5819	0.4819	0.4869	0.4209	0.6069	0.4839	0.5119	0.4559
SUN	0.6547	0.4777	0.5397	0.4737	0.5947	0.4787	0.4997	0.4437	0.6097	0.4967	0.5147	0.4570
SEG	0.7470	0.5702	0.6320	0.5660	0.6870	0.5670	0.5920	0.5360	0.7020	0.5790	0.6070	0.5610
SeR	0.6654	0.4902	0.5604	0.4844	0.6054	0.4824	0.5104	0.4544	0.6203	0.5004	0.5253	0.4697
CA	0.7869	0.6589	0.6819	0.6058	0.7269	0.6039	0.6319	0.5660	0.7421	0.6289	0.6497	0.5999
GR	0.8530	0.7189	0.7114	0.4935	0.8199	0.6898	0.7514	0.4814	0.8200	0.7264	0.7249	0.3973
FES	0.8089	0.5719	0.7171	0.3595	0.5967	0.4798	0.5682	0.2763	0.7223	0.5969	0.6882	0.3612
MC	0.9020	0.7329	0.8392	0.5807	0.8010	0.6579	0.7579	0.4655	0.8374	0.7517	0.8124	0.5186
DSR	0.8708	0.7063	0.8087	0.6246	0.7430	0.6464	0.7210	0.5930	0.8171	0.7292	0.7945	0.5999
RBD	0.8220	0.7356	0.7765	0.6581	0.8039	0.7011	0.7698	0.6547	0.8076	0.7516	0.7951	0.6128
CYB	0.8650	0.7786	0.8195	0.7011	0.8469	0.7441	0.8128	0.6977	0.8506	0.7946	0.8381	0.6558
ResNet-50	0.8867	0.7685	0.8175	0.6854	0.7996	0.7393	0.7672	0.7108	0.8151	0.7655	0.8004	0.7346
SDDF	0.9080	0.7898	0.8388	0.7067	0.8209	0.7606	0.7885	0.7321	0.8364	0.7868	0.8217	0.7559
Our	0.9293	0.8111	0.8601	0.7280	0.8422	0.7819	0.8098	0.7534	0.8577	0.8081	0.8430	0.7772

AP, AUC, F_m, and F_m^w denote the average precision, Area under the ROC curve, F-measure, and generalized F-measure, respectively. The top three quantitative results are highlighted in red, green, and blue color, respectively

The results of the six different data sets using thirteen different detection methods indicate the superiority of our method. The difference between our approach and the other approaches is clearly seen in Figs. 6, 7, 8, 9, and 10. For the single and multiple objects, our proposed approach outperforms in both quantitative and qualitative comparisons as shown in Tables 2 and 3. Our proposed method produces precise salient detection results with complete edge information, while other algorithms ignore the boundary information due to high contrast and humiliating contexts in images. We will continue with more powerful superpixel-based techniques to look for more reliable edge information for an image in future. It is also helpful to improve the background detection process to produce better results. Deep learning patterns and structural properties will also be studied in order to detect low colour, where weakly supervised learning patterns may be used.

Table 3 Saliency detection results of different techniques on three multiple object-based benchmarks

Methods	THUR-15K				DUT-OMRON				HKU-IS			
	AP	AUC	F_m	F_m^w	AP	AUC	F_m	F_m^w	AP	AUC	F_m	F_m^w
SIM	0.5319	0.5009	0.5250	0.3209	0.6469	0.5399	0.6179	0.4419	0.4109	0.4369	0.4629	0.3309
SUN	0.5847	0.5537	0.5778	0.3737	0.6997	0.5927	0.6707	0.4947	0.4637	0.4897	0.5157	0.3837
SEG	0.6970	0.6660	0.6901	0.4860	0.7520	0.6450	0.7230	0.5570	0.5760	0.6020	0.6280	0.4960
SeR	0.6054	0.5780	0.5985	0.3944	0.7204	0.6134	0.6914	0.5254	0.4844	0.5104	0.5364	0.4044
CA	0.7269	0.6994	0.7200	0.5159	0.7871	0.6801	0.7581	0.5921	0.6059	0.6319	0.6579	0.5259
GR	0.8003	0.6832	0.6998	0.3458	0.8890	0.7504	0.8120	0.4734	0.6620	0.7188	0.6209	0.2886
FES	0.6250	0.5797	0.6059	0.3264	0.7522	0.6073	0.7106	0.4018	0.6801	0.5959	0.4998	0.2839
MC	0.7542	0.7137	0.7436	0.4548	0.8941	0.7558	0.8558	0.5754	0.7192	0.6190	0.6060	0.3552
DSR	0.7510	0.7068	0.7435	0.5423	0.8493	0.7382	0.8162	0.6476	0.7364	0.6059	0.5902	0.4257
RBD	0.7476	0.7418	0.7451	0.5608	0.8596	0.7738	0.8425	0.6852	0.7137	0.6436	0.6039	0.4419
CYB	0.7745	0.7844	0.7689	0.6361	0.8600	0.7731	0.8510	0.7458	0.6511	0.7592	0.6077	0.5306
ResNet-50	0.7877	0.7976	0.7821	0.6493	0.8732	0.7863	0.8642	0.7590	0.6643	0.7724	0.6209	0.5438
SDDF	0.8005	0.8104	0.7949	0.6621	0.8860	0.7991	0.8770	0.7718	0.6771	0.7852	0.6337	0.5566
Our	0.8265	0.8364	0.8209	0.6881	0.9120	0.8251	0.9030	0.7978	0.7031	0.8112	0.6597	0.5826

AP, AUC, F_m, and F_m^w are as defined in Table 2. The top three quantitative results are highlighted in red, green, and blue color, respectively

Table 4 Comparison of average running time (s) per image on different data sets

Data sets	SIM	SUN	SEG	SeR	CA	GR	FES	MC	DSR	RBD	CYB	RN50	SDDF	Our
MSRA	1.230	3.270	10.28	1.284	40.42	1.358	0.069	0.286	10.29	0.285	1.653	1.329	0.027	1.006
MSRA10K	1.110	3.561	10.81	1.311	40.90	1.351	0.096	0.192	10.20	0.269	1.235	1.035	–	0.772
THUR15K	1.259	3.254	10.39	1.265	39.52	1.269	0.080	0.267	9.268	0.845	1.112	1.064	–	1.016
SED	1.265	3.316	10.25	1.256	39.49	1.291	0.098	0.384	9.326	0.265	1.547	1.263	0.583	0.984
DUT-MN	1.259	3.254	10.39	1.265	39.52	1.269	0.080	0.267	9.268	0.845	1.398	0.857	0.198	0.489
HKU-IS	1.265	3.316	10.25	1.256	39.49	1.291	0.098	0.384	9.326	0.265	1.645	1.152	0.542	0.452

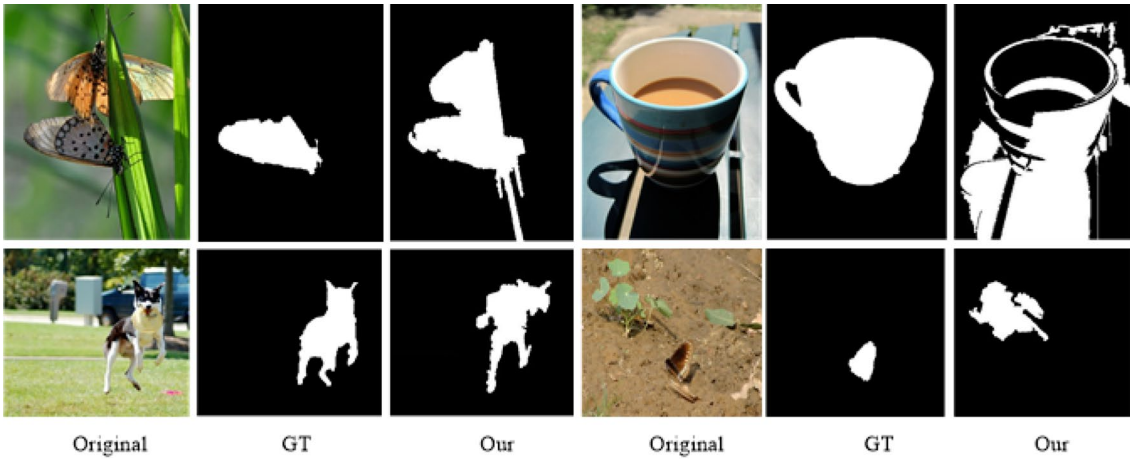


Fig. 12 Failure cases of the proposed method. It shows the difficulty of generating accurate saliency map when dealing with images with small and shaded objects in the landscape context

References

Achanta R, Hemami S, Estrada F, Susstrunk S (2009) Frequency-tuned salient region detection. In: Ieee International Conference on computer vision and pattern recognition (cvpr 2009), pp 1597–1604

Azaza A, van de Weijer J, Douik A, Masana M (2018) Context proposals for saliency detection. Comput Vis Image Underst 174:111

Badoual A, Unser M, Depeursinge A (2019) Texture-driven parametric snakes for semiautomatic image segmentation. Comput Vis Image Underst 188:102793

Barranco F, Diaz J, Pino B, Ros E (2014) Realtime visual saliency architecture for fpga with top-down attention modulation. IEEE Trans Ind Inf 10(3):1726–1735

Chen L-Q, Xie X, Fan X, Ma W-Y, Zhang H-J, Zhou H-Q (2003) A visual attention model for adapting images on small displays. Multimed Syst 9(4):353–364

Chen S, Wang B, Tan X, Hu X (2020) Embedding attention and residual network for accurate salient object detection. IEEE Trans Cybern 50(5):2050–2062

Ding M, Tong R-F (2010) Content-aware copying and pasting in images. Vis Comput 26(6–8):721–729

Duff P (2015) Retrieved from https://www.w3.org/TR/compositing-1

Gao Y, Shi M, Tao D, Xu C (2015) Database saliency for fast image retrieval. IEEE Trans Multimed 17(3):359–369

Goferman S, Zelnik-Manor L, Tal A (2011a) Context-aware saliency detection. IEEE Trans Pattern Anal Mach Intell 34(10):1915–1926

Goyal P, Mahajan D, Gupta A, Misra I (2019) Scaling and benchmarking self-supervised visual representation learning. In: Proceedings of the ieee/cvf International Conference on computer vision, pp 6391–6400

Harel J, Koch C, Perona P (2007) Graph-based visual saliency. In: Advances in neural information processing systems, pp 545–552

Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20(11):1254–1259

Jian M, Lam K-M, Dong J, Shen L (2014) Visual- patch-attention-aware saliency detection. IEEE Trans Cybern 45(8):1575–1586

Jiang B, Zhang L, Lu H, Yang C, Yang M-H (2013) Saliency detection via absorbing Markov chain. In: Proceedings of the ieee International Conference on computer vision, pp 1665–1672

Kim G, Yang S, Sim J-Y (2017) Saliency-based initialisation of gaussian mixture models for fully-automatic object segmentation. Electron Lett 53(25):1648–1649

Li X, Li Y, Shen C, Dick A, Van Den Hengel A (2013a) Contextual hypergraph modeling for salient object detection. In: Proceedings of the ieee International Conference on computer vision, pp 3328–3335

Li X, Lu H, Zhang L, Ruan X, Yang M-H (2013b) Saliency detection via dense and sparse reconstruction. In: Proceedings of the ieee International Conference on computer vision, pp 2976–2983

Liu T, Yuan Z, Sun J, Wang J, Zheng N, Tang X, Shum H-Y (2010) Learning to detect a salient object. IEEE Trans Pattern Anal Mach Intell 33(2):353–367

Liu Z, Zou W, Le Meur O (2014) Saliency tree: a novel saliency detection framework. IEEE Trans Image Process 23(5):1937–1952

Murabito F, Spampinato C, Palazzo S, Giordano D, Pogorelov K, Riegler M (2018) Top-down saliency detection driven by visual classification. Comput Vis Image Underst 172:67–76

Nawaz M, Yan H (2020) Saliency detection using deep features and affinity-based robust background subtraction. IEEE Trans Multimed

Nawaz M, Khan S, Cao J, Qureshi R, Yan H (2019a) Saliency detection by using blended membership maps of fast fuzzy-c-mean clustering. In: Eleventh International Conference on machine vision (icmv 2018), Vol. 11041, p. 1104123

Nawaz M, Khan S, Qureshi R, Yan H (2019b) Clustering based one-to-one hypergraph matching with a large number of feature points. Signal Process Image Commun 74:289–298

Perazzi F, Krahenbuhl P, Pritch Y, Hornung A (2012) Saliency filters: contrast based filtering for salient region detection. In: 2012 ieee Conference on computer vision and pattern recognition, pp 733740

Rahtu E, Kannala J, Salo M, Heikkila J (2010) Segmenting salient objects from images and videos. In: European Conference on computer vision, pp 366–379

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

Shahid AR, Khan S, Yan H (2020a) Contour and region harmonic features for sub-local facial expression recognition. J Vis Commun Image Represent 73:102949

Shahid AR, Khan S, Yan H (2020b) Human expression recognition using facial shape based fourier descriptors fusion. In: Twelfth International Conference on machine vision (ICMV 2019), 11433, 114330P

Sokhandan A, Monadjemi A (2018) Visual tracking in video sequences based on biologically inspired mechanisms. Comput Vis Image Underst

Tavakoli HR, Rahtu E, Heikkila J(2011) Fast and efficient saliency detection using sparse sampling and kernel density estimation. In: Scandinavian Conference on image analysis, pp 666–675

THUR-15000 (2013) Retrieved from https://mmcheng.net/gsal/

Wang W, Lai Q, Fu H, Shen J, Ling H (2019) Salient object detection in the deep learning era: an in-depth survey. arXiv preprint arXiv:1904.09146

Wang Z, Xiang D, Hou S, Wu F (2016) Background-driven salient object detection. IEEE Trans Multimed 19(4):750–762

Xi X, Luo Y, Wang P, Qiao H (2019) Salient object detection based on an efficient end-to-end saliency regression network. Neurocomputing 323:265–276

Xue Y, Shi R, Liu Z (2011) Saliency detection using multiple region-based features. Opt Eng 50(5):057008

Yang C, Zhang L, Lu H (2013a) Graph-regularized saliency detection with convex-hull-based center prior. IEEE Signal Process Lett 20(7):637640

Yang C, Zhang L, Lu H, Ruan X, Yang M-H (2013b) Saliency detection via graph-based manifold ranking. In: Proceedings of the ieee Conference on computer vision and pattern recognition, pp 31663173

Yuan Y, Li C, Kim J, Cai W, Feng DD (2017) Reversion correction and regularized random walk ranking for saliency detection. IEEE Trans Image Process 27(3):1311–1322

Zhang L, Gu Z, Li H (2013) Sdsp: a novel saliency detection method by combining simple priors. In: 2013 ieee International Conference on image processing, pp 171–175

Zhu W, Liang S, Wei Y, Sun J (2014) Saliency optimization from robust background detection. In: Proceedings of the ieee Conference on computer vision and pattern recognition, pp 2814–2821

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.