Boundary Preserved Salient Object Detection Using Guided Filter Based Hybridization Approach of Transformation and Spatial Domain Analysis

BHAGYASHREE V. LAD[1], (Student Member, IEEE),
MOHAMMAD FARUKH HASHMI[2], (Senior Member, IEEE),
AND AVINASH G. KESKAR[1], (Senior Member, IEEE)

[1]Department of Electronics and Communication Engineering, Visvesvaraya National Institute of Technology, Nagpur 440010, India
[2]Department of Electronics and Communication Engineering, National Institute of Technology, Warangal 506004, India

Corresponding author: Bhagyashree V. Lad (bhagyashreevl@students.vnit.ac.in)

ABSTRACT Techniques for detecting salient objects mimic human behavior by recognizing the most noticeable parts of images as objects. Salient object detection has attracted many researchers' attention nowadays for various computer vision and pattern recognition applications. In this paper, a unique approach is proposed based on the global and local saliency detection using wavelet transform and hybridizing it with learning-based saliency detection using a guided filter. First, the input image is subjected to superpixel segmentation to achieve visually uniform regions and to reduce the computational cost. The global and local saliency maps are then generated using global and local features extracted by the wavelet transform of the segmented image, as the wavelet transform gives a multiscale analysis of images in frequency as well as in spatial domain. The learning-based saliency maps are generated using random forest regression which considers the location, color, and textural features of the segmented image. The global and local saliency maps are fused to generate the wavelet-based saliency map which is further hybridized with the saliency map generated using random forest regression. The paper discusses the novel technique for hybridizing wavelet-based and learning-based saliency maps using a guided filter-based attention map generation. Several experiments are conducted on five different saliency datasets containing images with complex backgrounds, multiple objects, and low contrast. To evaluate the efficacy of the proposed method, extensive qualitative and quantitative performance analysis is carried out. Experimental results validate the significant improvement in the detection of salient regions as compared to the state-of-the-art methods.

INDEX TERMS Discrete wavelet transform (DWT), guided filters, learning based saliency map, random forest regression, salient object detection, wavelet based saliency map.

I. INTRODUCTION

Salient object detection is a technique where some objects draw more visual attention than its surrounding. The human visual system (HVS) is a very complex process and it can easily detect salient objects. But it is very difficult to apply HVS in the applications of computer vision. Salient object detection methods are mostly based on the development of human attention mechanisms. The main aim of salient object detection is to localize the objects of interest as per HVS and

suppress the other surrounding part. It detects to what extent the objects are distinct from its surrounding and based on that the saliency maps are calculated. Salient object detection is considered to be a very effective step as it reduces a lot of computational time in a selection of computer vision applications like image segmentation, object recognition, image compression [1], image thumbnailing, image quality assessment, image retrieval, video summarization, object tracking, and medical imaging, etc. [2]. Zhu *et al.* [3] used perceptual information using visual saliency detection for image style transfer which is an artwork application of saliency detection. Many deep learning techniques for salient object

The associate editor coordinating the review of this manuscript and approving it for publication was Gangyi Jiang.

detection (SOD) tasks are proposed in recent years [4] which gives very good results at the cost of large data availability, very high computational and architectural complexity. Also, their results are quantitatively good but lack in preserving the complete boundaries and edges of the objects. The work in [5] used contour information to obtain the boundary preserved saliency maps by using a twice learning strategy. This method gives good results however the CNN model has to be trained twice, first to obtain the contour map, and this contour map is used as an input for the second training to obtain the final saliency map. This twice learning strategy may result in complex architecture. As a result, the work of SOD employing handcrafted features plays an important role in today's world for applications where data availability is limited and low computing complexity is preferred with better preservation of details.

In the proposed work superpixel segmentation is used to convert an image into visually consistent regions called superpixels, with the improvement in the computational efficiency of the proposed method. Superpixels are non-overlapping and aesthetically meaningful sets of pixels that can be used in place of individual pixels. HVS is more attracted to image regions than the individual pixel, as image regions give more information. In recent years, there has been a rise in developing saliency detection algorithms in both the spatial and frequency domains. It is very important to consider an acceptable range of frequency components to detect the entire salient region consistently. In most of the spatial domain approaches, the high-frequency components are discarded which results in the blurring of edges and boundaries of the objects. So the complete salient region in an image can not be highlighted uniformly. This limitation can be addressed by using frequency-domain techniques. The wavelet transform deals with the multiresolution analysis of an image, i.e. it uses both spatial and frequency domain-based knowledge to generate the saliency maps. The majority of saliency detection algorithms consider only the local saliency features, which are based on the image's fundamental characteristics, but the overall significance is frequently overlooked. The global saliency map depicts global information based on low-level characteristics. It accurately and consistently recognizes the complete important object. As a result, combining global and local saliency maps can improve saliency detection performance. In the proposed work the local and global saliency maps based on wavelet features are fused to generate the final wavelet transform-based saliency map. The learning-based method deals with the features i.e. spatial distance and local and global color contrast between the superpixels, and textural distance to obtain the saliency map using random forest regression. With the help of the wavelet transform-based method, we are concentrating on the textural information and by using the learning-based method, location, color, and textural contrast-based features are considered.

For creating the final saliency maps, most state-of-the-art algorithms use saliency map integration of previously produced baseline saliency maps [6]–[8]. For integrating baseline saliency maps many methods [9], [10] use simple pixel-wise addition or multiplication. Edge blurring or artifacts around the object boundaries may emerge if these methods ignores the intensity variations of nearby pixels. Some of the methods integrate global and local saliency maps using a weighted average, however, the selection of weights is generally done by trial and error. Meta-heuristic optimization approaches have also been proposed, although they come at a higher cost in terms of computing. In the proposed work the novel hybridization technique using guided filters is presented, to intelligently integrate the more visually prominent regions from the wavelet transform-based saliency map and learning-based saliency map. A guided filter-based model extracts the prominent visual features by generating the decision maps of the wavelet and learning-based saliency maps and gives a more informative and perceptually appealing final saliency map. Mainly the work in video saliency detection requires computationally efficient methods to apply in real-time video salient object detection tasks. The proposed method can also be applied to video saliency detection [11] in the future as it is computationally efficient and has simple architecture.

In the era of deep learning-based methods, the employment of handcrafted feature-based algorithms and conventional computer vision techniques to solve SOD tasks is still highly significant [12], [13]. Traditional computer vision methods can solve problems faster and with fewer lines of code than deep learning-based methods. Handcrafted feature-based methods, such as simple color thresholding and pixel counting techniques, are not class-specific, which means they work with any image making it independent of the data. Deep learning features, on the other hand, are specific to the training dataset and are unlikely to work well for images other than the training set if poorly created. If the training dataset is too small, the machine may be unable to generalize to the situation at hand due to over-fitting. Manually changing the model's parameters would be too complex because a deep neural network has millions of parameters, each with complicated inter-relationships. Most recent deep learning algorithms may achieve far higher accuracy but at the cost of billions of additional math operations and a higher computing power demand. For some applications where low computational complexity and less data dependency are required, machine learning-based algorithms play a very important role. As machine learning techniques are computationally less complex and can work on less amount of data than deep learning methods. So the handcrafted feature-based and machine learning-based salient object detection techniques are not obsolete. In the proposed work machine learning-based salient object detection technique is used which makes the proposed method computationally efficient.

The following are the major contributions of the proposed work:
- A novel technique is used to obtain local and global saliency maps using wavelet transform, as earlier

research worked only on generating global saliency maps using wavelet transform.

- The learning-based saliency map is predicted by a random forest regression model using location, color, and textural features of superpixels.
- The novel hybridization method to merge the wavelet-based and learning-based saliency maps using a guided filter is proposed. Guided filters work on decision map generation by refining the attention maps which considers saliency maps using wavelet transform and learning-based method as guidance images.
- The proposed method detects more than one salient object accurately by preserving boundaries and edges of salient objects concerning other state-of-the-art methods.

The remainder of the paper is laid out as follows. Section 2 discusses, in brief, the salient object detection techniques concerning wavelet transform-based as well as learning-based approaches. Section 3 presents in detail the implementation steps of the proposed method. Section 4 presents the experiments and discussions. The run time analysis is presented in section 5. Finally, the limitations and conclusion are presented in sections 6 and 7 respectively.

II. RELATED WORK

The saliency methods are mainly based on two approaches: generation of eye fixation maps and generation of saliency maps. The work in [14] and [15] indicated saliency by capturing eye gaze points. The eye fixation point prediction is used in various applications such as advertisement placement in a video, finding fixation scan paths of an eye, etc. But the tasks in computer vision and pattern recognition mainly deal with locating and segmenting the salient objects, where locating the eye fixation points does not work. So other methods like [7], [16] provided the results as a saliency map which indicates the probability of each pixel being salient. Further, the saliency detection task is classified into bottom-up and top-down approaches. Bottom-up techniques are typically used for handcrafted low-level features that are data-driven. The bottom-up saliency approaches [2], [17], [18] employ self-information, histogram, and region-based characteristics, dissimilarities measured in local areas, weighting based on information content, and a frequency refined approach to compute the saliency. Many saliency detection [6], [17], [19] techniques use graph-based approaches with the help of handcrafted low-level features and different saliency priors which has improved the saliency detection results to great extent. All the above-discussed methods use saliency models to identify the most important object based on color contrast. As a result, items with comparable colors to the background will not be identified by these methods, and textures and edges will not be retained. Bottom-up techniques are concise to develop and converge faster, but they have limitations when dealing with low contrast and dynamic patterned images. Top-down saliency detection methods, on the other hand, are task-driven and use convolutional and deep neural networks

to recognize task-specific high-level features [4], [20]. The use of such technologies improves performance in this area but at the cost of a significant amount of data and a lot of computer power. Also recently many researchers are focusing on depth enhanced saliency detection models [8], [12], [20] which take the information from depth cues to enhance the performance of saliency detection.

Salient object detection techniques can also be categorized as supervised, semi-supervised and unsupervised techniques as suggested by Ji *et al.* in [4]. The proposed method can be treated as the hybridization of supervised and unsupervised saliency detection techniques. The transform-based saliency detection and hybridization part using guided filters of the proposed method does not involve any training process with labeled data so they are categorized as unsupervised saliency detection while the learning-based saliency map generation process of the proposed method involves training of random forest regression model using labeled data to predict the result, so it can be categorized to supervised saliency detection technique.

The wavelet transform-based methods are used in the literature on salient object detection for extracting the image features in frequency domain [21], [22]. The work in [23], [24] utilizes wavelet transform for detecting the salient objects. The main advantage of using wavelet transform for salient object detection is that it uses multi-scale analysis of the image by considering frequency domain as well as spatial domain information at the same time. The saliency model based on wavelet transform [25] detects the salient regions which consider only detail coefficients at different scales to generate multi-scale feature maps. The method proposed by [24] utilizes the visual data extraction by using a combination of wavelet transform and contrast mechanism. Here the difference of Gaussian function is used as the basis function for the wavelet transform and wavelet transform decomposition is applied to the multi-channel of the human visual system. The wavelet-based saliency detection techniques help to preserve the details of the objects to a great extent but the entire object cannot still be detected using only wavelet transform.

In recent research, machine learning-based bottom-up saliency detection methods are also extensively accepted. For saliency identification, Pang *et al.* [16] employed a bagging-based distributed learning strategy that uses training samples based on center prior and background prior information. Singh and Kumar [26] and Lei *et al.* [27] used a framework based on bagging and Bayesian decision, respectively to enhance the basic rough saliency map extracted using several saliency techniques. Tong *et al.* [9] used a bootstrap learning strategy to create a powerful classifier that can differentiate between prominent and background objects. Yang and Yang [28] used a conditional random field (CRF) and a visual dictionary to quantify saliency. Huang *et al.* [29] proposed a saliency metric that takes into account both object suggestions and multiple instances of learning. The suggested method's learning-based saliency detection is also
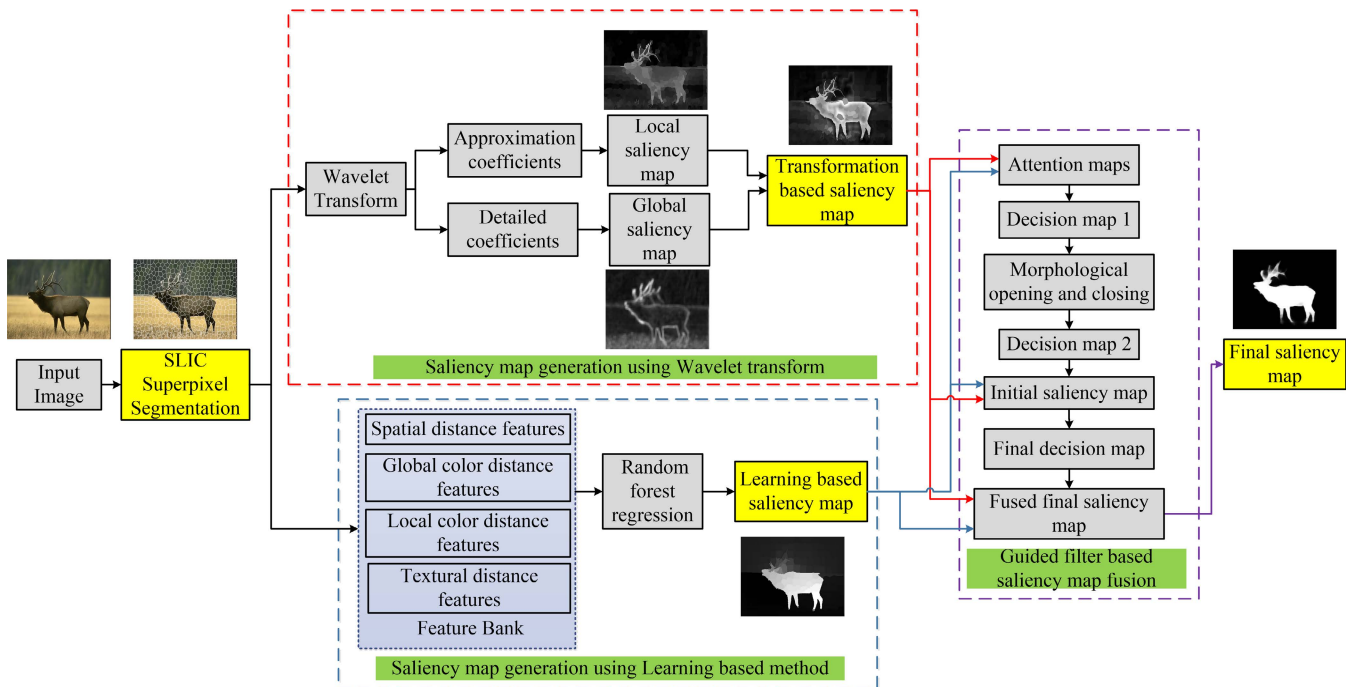
FIGURE 1. Workflow of proposed algorithm.

based on a machine learning-based bottom-up salient object recognition methodology. All the above-discussed models are accurate and simple but they require more computational time. So in this paper, we have used superpixel-wise computation to overcome the limitation of high computational time. Superpixel segmentation is generally used to reduce the computational cost of many algorithms, which is largely used in salient object detection research. Superpixel-wise segmentation helps to convert an image into homogeneous regions which are considered as the image elements for that particular group of pixels, which further helps to reduce the computational overhead.

The wavelet transform-based salient object detection method helps to preserve the edges and boundaries of the objects but can not highlight the object as a whole. While learning-based techniques consider various visual features to detect the complete salient objects but can not preserve the object boundaries. So, to utilize the advantages of both wavelet-based and learning-based salient object detection methods, the proposed method merges wavelet-based and learning-based saliency maps using edge-preserving guided filters. There are many methods proposed to merge the baseline saliency maps. To incorporate several existing saliency approaches, Xu *et al.* [30] presented an arbitrator model. The method created a reference map by combining the results of numerous existing saliency methods with external knowledge to create a reference map. After that, the method picks up on current saliency methods' expertise. Finally, to merge the known saliency methods of varying expertise with the reference map, a unique integration framework based on Bayesian inference is used. Qin *et al.* [31] proposed a cellular automata

aggregation approach for incorporating diverse saliency maps created by existing saliency methods. These integration methods fail to preserve the details of existing saliency maps. So the proposed method uses boundary preserved guided filter-based integration which improves the salient object detection task by preserving the edges and boundaries of objects and also by detecting more than one salient object with better accuracy.

III. PROPOSED METHOD
The workflow of the proposed method based on hybridization of wavelet-based and learning-based saliency maps using guided filters is shown in Fig.1. The implementation details of the proposed method are given below.

A. SUPERPIXEL SEGMENTATION
Single pixels are less appealing to the human visual system than image sections. In the proposed method firstly an input image is transformed to CIELAB color space, as CIELAB color space is thought to be highly similar to human perception. An image is split into the superpixels after it has been converted to CIELAB color space. Many research [12], [32]–[35] have also shown that superpixel segmented images are computationally intensive and particularly successful for salient object detection. We obtain the superpixels SP from an input image I by over-segmenting an image, where SP is given by,

$$SP = \{SP_1, SP_2, \ldots \ldots, SP_P\} \qquad (1)$$

In the proposed work the number of superpixels P is considered to be $P = 500$. For segmenting an image into superpixels

SLIC superpixel algorithm [36] is used. It is computationally effective as the search space in SLIC superpixel segmentation is considered to be the region similar to the size of the superpixel instead of the whole image.

B. SALIENCY MAP GENERATION USING WAVELET TRANSFORM

The proposed method works on obtaining the local and global feature maps using wavelet transform. The local saliency maps detect the saliency of pixels within the fixed neighborhood and global saliency detects the saliency of pixels considering the entire image. Wavelet transform is preferred over other transform domain techniques as it deals with multiresolution analysis, while other transforms like Fourier transform, performs more local analysis. In the proposed method, the l, a, and b values of all the pixels which belong to one superpixel are replaced by the l, a, and b values of the center of that superpixel. The most unique superpixels are located by applying the Wavelet transform to the l, a, and b channels of the superpixel segmented image. The application of wavelet transform to the superpixel segmented image is given by Eq. 2.

$$\left\{ A_S^C, D_S^C \right\} = DWT_S(SP^C) \qquad (2)$$

SP^C denotes the superpixel of the image at each color channel. DWT (.) denotes the Discrete Wavelet Transform and A_S^C and D_S^C represents the approximation coefficients and detailed coefficients containing horizontal, vertical and diagonal details of an image, at different scales for each color channel $C \in \{l, a, b\}$. $S \in \{1, 2, \ldots, N\}$ indicates the decomposition levels of wavelet transform where N is the maximum decomposition level which we have considered 8 here. Based on the superiority of wavelet transform in multiscale decomposition, the wavelet transforms coefficients at each scale are used to extract the local and global features which are further used to generate the local and global feature maps for salient object detection. The local feature maps are obtained by using some statistical process on low pass sub-band and global feature maps are obtained by utilizing the high pass sub-bands of the wavelet transform.

1) LOCAL SALIENCY MAP GENERATION USING WAVELET BASED FEATURES

The local feature maps using wavelet transform are obtained by extracting features from low pass sub-bands. After each decomposition level, the low-pass sub-band is partitioned into blocks of $p \times q$ superpixels, and the local variance of each block is taken as the block's local feature map value, as indicated in Eq. 3.

$$Local_S^C (x, y) = variance \left\{ A_S^C (x + p, y + q) \right\}_{\{p,q = \pm 2, \pm 1, 0\}} \qquad (3)$$

These local feature maps describe the image's different textural features at different decomposition levels. To generate

the local saliency map, the local feature maps are linearly combined using low entropy criteria. The local saliency map is given by Eq. 4.

$$Smap_L(x, y) = \sum_{S=1}^{8} \max_{C \in \{l,a,b\}} \left(w_S^C Local_S^C \right) \qquad (4)$$

w_S^C describes the weights to corresponding features at each decomposition level for each color channel. The weight values are decided based on the entropy of the local feature maps. A low entropy value generally indicates that the data is concentrated towards one value i.e. the spread is less, which is the useful criteria for saliency detection. So the weights are assigned according to the Eq. 5.

$$w_S^C = \frac{\sum_{x,y} M \left(Nr(Local_S^C) \right)}{H \left(Local_S^C * G \right)} \qquad (5)$$

M is a two-dimensional centered Gaussian mask with elements that have the highest value of 1 and with the same size as the channel feature map. Here centered Gaussian map is considered, since saliency is generally considered at the center of an image. $Nr (\cdot)$ indicates the normalization function applied to local feature maps. $H (\cdot)$ refers to the entropy value of the smoothened local feature maps which are filtered with Gaussian kernel G.

2) GLOBAL SALIENCY MAP GENERATION USING WAVELET BASED FEATURES

The global saliency map is generated based on the selection of detailed coefficients of wavelet transform over each scale. These wavelet coefficients are further used to generate the feature maps with multiresolution analysis using superpixels. To extract the global features of an image from the three color segmented channels, inverse wavelet transform is used where the reconstruction is done only using the detailed wavelet coefficients from finer to coarser scale which is given by Eq. 6.

$$Global_S^C = \frac{(IDWT_S[D_S^C])^2}{k} \qquad (6)$$

where $Global_S^C$ denotes the global feature maps for each color channel at various decomposition levels and $IDWT (\cdot)$ denotes the inverse discrete wavelet transform which is used for reconstruction. The feature maps are scaled using the scaling factor k which is considered as $k = 10^4$. The scaling of the feature map is important to minimize the large fluctuations in the co-variance matrix while calculating inverse wavelet transform. The feature maps at different decomposition levels represent the l, a, and b color channels features from finer to coarser level i.e. it indicates the features from edges to different textures of an image. To obtain the global saliency map from the extracted features, only the maximum value from three color channels $\{l, a, b\}$ is considered and finally these maximum valued features are summed over all the decomposition levels. The Gaussian lowpass filter G is

TABLE 1. Features of each superpixel used in learning based method.

Features	Dimension
Superpixels spatial distance from K-nearest superpixels	K
Superpixels global color distance from K-nearest superpixels	8K
Superpixels local color distance from K-nearest superpixels	8K
Superpixels textural distance from K-nearest superpixels	10K

applied to this global saliency map to obtain the smooth saliency map $Smap_G(x, y)$ which is given by Eq. 7.

$$Smap_G(x, y) = (\sum_{S=1}^{8} \max_{C \in \{l,a,b\}} (Global_S^C)) * G \quad (7)$$

The saliency map based on wavelet transform is obtained by merging the local and global wavelet-based saliency maps. The saliency map using wavelet transform is given by Eq. 8.

$$Smap_W = \left(\frac{[Smap_L \times e^{Smap_G}]^{ln\sqrt{2}}}{\sqrt{2}} \right) * G \quad (8)$$

$Smap_W$ represents the final saliency map using wavelet transform by fusing local and global saliency maps. Saliency is nothing but the region of an image that stands out in the entire image. Global saliency helps to preserve the edges of the objects, while local saliency takes neighborhoods into account to detect the objects accurately. So, the merging of local and global saliency gives prominent results in the salient object detection tasks.

C. SALIENCY MAP GENERATION USING LEARNING BASED METHOD

The generation of saliency map based on learning-based method deals with spatial, color, and texture difference of the superpixel with K neighboring superpixels. Here we used features as the euclidean distance of each superpixel from K-nearest superpixels, a global and local color distance of each superpixel from K nearest superpixels, and textural distance of superpixels from K nearest superpixels. The value of K we have set to 25 as it was giving best results considering the F-measure value. For every l^{th} superpixel SP_l, we first obtain K-nearest superpixels $\{SP_{l_1}, SP_{l_2}, SP_{l_3}, \ldots \ldots SP_{l_K}\}$.

1) SPATIAL DISTANCE FEATURES

The euclidean distance feature vector of lth superpixel from K-nearest superpixels $(D_l \in \mathbb{R}^{K \times 1})$ is given by Eq. 9.

$$D_l = d(y_l, y_{l_i})_{\{i \in 1,2,..,K\}} \quad (9)$$

Here, y_l and y_{l_i} denotes the center location of the lth superpixel and its K-nearest superpixels. $d(y_l, y_{l_i})$ denotes the euclidean distance of the lth superpixel from its K-nearest superpixels.

2) GLOBAL COLOR DISTANCE FEATURES

The global color contrast feature vector of lth superpixel from K-nearest superpixels $(D_{GC_l} \in \mathbb{R}^{8K \times 1})$ is given by Eq. 10.

Here 8 color channels i.e. CIELAB, RGB, hue and saturation are used to obtain the color contrast features. c_l and c_{l_i} are 8 dimensional color vectors. $d(c_l, c_{l_i})$ denotes the color distance of lth superpixel from K-nearest superpixels.

$$D_{GC_l} = d(c_l, c_{l_i})_{\{i \in 1,2,..,K\}} \quad (10)$$

3) LOCAL COLOR DISTANCE FEATURES

The local color contrast feature vector of lth superpixel from K-nearest superpixels $(D_{LC_l} \in \mathbb{R}^{8K \times 1})$ is given by Eq. 11. Here also 8 color channels i.e. CIELAB, RGB, hue, and saturation are used to obtain the color contrast features.

$$D_{LC_l} = w_{l_i} d(c_l, c_{l_i})_{\{i \in 1,2,..,K\}} \quad (11)$$

where w_{l_i} is given by Eq. 12.

$$w_{l_i} = \frac{1}{Z_l} exp\left(-\frac{1}{2\sigma^2} d(p_l, p_{l_i})\right)_{\{i \in 1,2,..,K\}} \quad (12)$$

p_l indicates the normalized position of lth superpixel and Z_l is the normalization term. This function of weight is adopted to give more weight to immediate neighboring superpixels. We have set the value of $\sigma^2 = 0.25$. The dimension of the feature vector for each superpixel concerning spatial distance is K and concerning color contrast is 8K.

4) TEXTURAL DISTANCE FEATURES

The texture distance features indicate the distance between feature vectors that represent region textures which are used to calculate texture deviation. The texture feature vector of lth superpixel from K-nearest superpixels $(D_{T_l} \in \mathbb{R}^{10K \times 1})$ is given by Eq. 13.

$$D_{T_l} = d(t_l, t_{l_i})_{\{i \in 1,2,..,K\}} \quad (13)$$

where $t(\cdot)$ indicates the textural attributes of the superpixel region, such as gradient mean, gradient direction, and histogram of the gradient. $d(t_l, t_{l_i})$ indicates the Euclidean distance between the textural attributes of lth superpixel from K-nearest superpixels. The dimension of the feature vector for each superpixel concerning textural distance is 10K.

5) SALIENCY MAP GENERATION USING RANDOM FOREST REGRESSION

The feature vectors for all the superpixels are obtained as per Table 1 and the saliency map $Smap_{RF}$ is generated using these feature vectors by using random forest regression [37] algorithm, as it is very effective for large dimensional feature vector. For training the random forest, we have used 3000 images from MSRA-B dataset [38] and annotated ground truth images we have used as labels. From MSRA-B dataset, 1500 images are used for testing and remaining images are used for validation. The trees used for the random forest regression model are 200 with the maximum level of the tree as 10 [39]. Random forest regression output decides the saliency of the superpixel according to the extent to which the particular superpixel belongs to the foreground or background region and a corresponding saliency map is produced for an image.

D. FINAL SALIENCY MAP GENERATION USING GUIDED FILTER BASED INTEGRATION

The final saliency map is created by fusing the attentive regions of the wavelet domain-based saliency map $Smap_W$ and learning-based saliency map $Smap_{RF}$ using guided filters [40]. The stepwise description of attention-based saliency map fusion using guided filters is as follows.

Step 1: The saliency maps $Smap_W$ and $Smap_{RF}$ are considered to be the source images for obtaining the final saliency map using guided filter-based fusion. To detect the attention regions, the source images are first blurred using mean filters. The textural or edge information is then obtained by finding the difference between source images and their mean filtered images with window size w. f_a is mean filter with window size w as 3×3. The high pass edge maps of the corresponding source images are given in Eq. 14 and Eq. 15.

$$Emap_1(x, y) = |Smap_W(x, y) - (Smap_W(x, y) * f_a)| \quad (14)$$

$$Emap_2(x, y) = |Smap_{RF}(x, y) - (Smap_{RF}(x, y) * f_a)| \quad (15)$$

Step 2: The high pass information maps are further refined using guided filters [40] by considering $Smap_W$ and $Smap_{RF}$ as guidance images. The refined high pass maps are given by Eq. 16 and Eq. 17.

$$REmap_1 = GF_{R,\theta}(Smap_W(x, y), Emap_1(x, y)) \quad (16)$$

$$REmap_2 = GF_{R,\theta}(Smap_{RF}(x, y), Emap_2(x, y)) \quad (17)$$

where $GF_{R,\theta}()$ represents the guided filter having local window radius R and degree of blur θ of kernel function. In the proposed work we have considered $R = 5$ and $\theta = 0.3$ by trial and error approach.

Step 3: The attention measures of the source images $Smap_W$ and $Smap_{RF}$ are represented by $REmap_1$ and $REmap_2$, respectively. As a result, the pixel-wise maximum rule of the matching accurate attention maps $REmap_1$ and $REmap_2$ is used to generate an initial decision map. The initial decision map is given by Eq. 18.

$$IDmap_1(x, y) = \begin{cases} 1, & \text{if } REmap_1(x, y) > REmap_2(x, y) \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

Step 4: The non-attentive regions of the initial decision map $IDmap_1$ contain some non-required spots or burrs which can be removed using morphological opening and closing operation.

$$O = (IDmap_1 \ominus D) \oplus D \quad (19)$$

$$IDmap_2 = (O \oplus D) \ominus D \quad (20)$$

Eq. 19 indicates the morphological opening operation on the initial decision map using disk structure window D of radius 10. The attentive regions in O have some blurs which can be further eliminated by applying the morphological closing operation with disk structure window D. Eq. 20 generates the decision map $IDmap_2$ using morphological opening and closing operations.

Step 5: Following the tiny region elimination technique using morphological opening and closing operation, an initial fused saliency map $Smap^I$ is obtained as stated in Eq. 21 using decision map $IDmap_2$ and source images $Smap_W$ and $Smap_{RF}$.

$$Smap^I = IDmap_2(x, y) Smap_W(x, y)$$
$$+ (1 - IDmap_2(x, y)) Smap_{RF}(x, y) \quad (21)$$

Step 6: By utilizing the processed initial decision map $IDmap_2$, the edges between attention and non-attention regions of the source images cannot be retained, which may result in some distortions in generating the final saliency map. So, the guided filter is employed once again to obtain boundary preserved final decision map by considering the initial fused saliency map $Smap^I$ as the guidance image which helps to improve the weights of the final decision map. The desired final decision map $FDmap$ is obtained by Eq. 22 by filtering decision map $IDmap_2$ and considering initial fused saliency map $Smap_I$ as the guidance image.

$$FDmap(x, y) = GF_{R,\theta}\left(Smap^I(x, y), IDmap_2(x, y)\right) \quad (22)$$

Step 7: The final fused saliency map $Smap_F$ is thus obtained by pixel-wise weighted averaging of source images and final decision map which is stated in Eq. 23.

$$Smap^F = FDmap(x, y) Smap_W(x, y)$$
$$+ (1 - FDmap(x, y)) Smap_{RF}(x, y) \quad (23)$$

IV. EXPERIMENTS AND DISCUSSIONS

This section compares the proposed guided filter based hybridization of wavelet based and learning based saliency maps approach to 17 state-of-the-art and contemporary saliency detection methods in terms of quantitative and qualitative performance. The quantitative and qualitative analysis is done with the methods proposed by Goferman *et al.* (CA) [50], Yang *et al.* (GR) [51], Rahtu *et al.* (SEG) [52], Yang *et al.* (MR) [35], Jiang *et al.* (MC) [53], Li *et al.* (LPS) [54], Li *et al.* (RR) [55], Tong *et al.* (LGF) [56], Zhou *et al.* (DPSG) [57], Fu *et al.* (NCUT) [58], Yuan *et al.* (RCRR) [59], Peng *et al.* (SMD) [60], Liu and Yang (FCB) [61], Zhang *et al.* (LSP) [62], Pang *et al.* (BSDL) [16], Wang and Peng (CDHL) [7] and Sun *et al.* (TMR) [13].

A. DATASETS FOR SALIENT OBJECT DETECTION

For the performance evaluation of the proposed technique, five key object detection datasets are used, as shown in Table 2. Images with a complicated and messy background, poor contrast, and many objects can be found in the datasets. The suggested method's performance is assessed across the full dataset to demonstrate its capacity to operate efficiently and reliably across a diverse group of images.

TABLE 2. Salient object detection datasets.

Sr. No.	Dataset	No.of images	Description
1	SOD [41]	300	Many images include multiple prominent items with little color contrast to the background
2	ECSSD [42]	1000	The dataset contains multiple salient objects and also having a complex background, which makes the dataset more challenging for the salient object detection task
3	DUT-OMRON [35]	5168	Images of high quality with multiple significant objects, while the backgrounds are relatively cluttered
4	HKUIS [43]	4447	Majority of the images are low contrast with several salient objects
5	MSRA10K [44]	10000	Large number of diverse set of images, generally used for training the deep learning methods

TABLE 3. Evaluation metrics for salient object detection.

Sr. No.	Evaluation Metrics	Mathematical Expression
1	Precision [45]	$Precision = \frac{\sum_{x=1}^{w}\sum_{y=1}^{h} Smap_{x,y}^{FT} \cdot GT_{x,y}}{\sum_{x=1}^{w}\sum_{y=1}^{h} Smap_{x,y}^{FT}}$
2	Recall [45]	$Recall = \frac{\sum_{x=1}^{w}\sum_{y=1}^{h} Smap_{x,y}^{FT} \cdot GT_{x,y}}{\sum_{x=1}^{w}\sum_{y=1}^{h} GT_{x,y}}$
3	F-measure [46]	$Fm = \frac{(1+\xi^2)\cdot Precision \cdot Recall}{(\xi^2 \cdot Precision)+Recall}$ where $\xi^2 = 0.3$
4	MAE [47]	$MAE = \frac{1}{w\times h}\sum_{x=1}^{w}\sum_{y=1}^{h} \left| Smap_{x,y}^{FN} - GT_{x,y} \right|$
5	TPR [45]	$TPR = \frac{\left|Smap^{FT} \cap GT\right|}{\left|GT\right|}$
6	FPR [45]	$FPR = \frac{\left|Smap^{FT} \cap \overline{GT}\right|}{\left|Smap^{FT} \cap \overline{GT}\right| + \left|\overline{Smap^{FT}} \cap \overline{GT}\right|}$
7	S-measure [48]	$Sm = \eta * S_o + (1-\eta) * S_r$ where $\eta = 0.5$
8	E-measure [49]	$Em = \frac{1}{w\times h}\sum_{x=1}^{w}\sum_{y=1}^{h} \Psi(x,y)$

B. EVALUATION METRICS

The performance of the suggested technique is evaluated using the assessment metrics listed in Table 3. Receiver operating characteristic (ROC) curve, precision-recall (PR) curve, area under the ROC curve (AUC) score, mean absolute error (MAE) score, F-measure (Fm) score, enhanced-alignment measure i.e. E-measure (Em) score and structure measure i.e. S-measure (Sm) score are some of the evaluation metrics. For measuring the overall performance of salient object detection algorithms, F-measure, recall, and precision are commonly utilized [46]. In Table 3, $Smap_{x,y}^{FT}$ and $Smap_{x,y}^{FN}$ are thresholded saliency map and normalized saliency map, respectively. $GT_{x,y}$ is the ground truth of the image. The F-measure value is a comprehensive metric for saliency detection tasks since it combines the recall and precision values. Thresholded saliency maps of the acquired grey level saliency maps are required to evaluate these metrics. For a grayscale saliency map with pixel values in the range of [0, 255], the threshold is altered from 0 to 255 to create segmented binary saliency maps. The resulting saliency map $Smap_{x,y}^{F}$ is binarised using thresholds 0 to 255 to evaluate the PR curve. The recall and precision scores are computed for every value of the threshold, which is then plotted on the precision-recall curve. FPR and TPR values are also computed at each threshold to plot the ROC curve. The ROC curve provides a two-dimensional explanation of the effectiveness of the proposed model, whereas the AUC value condenses this description into a single number. The AUC value is calculated using the area under the ROC curve. The overlap-based performance measurements do not take into consideration the true negative assignment of saliency. These measures favor methods that give high saliency to prominent

pixels while failing to detect non-salient areas. Continuous saliency maps are more important than thresholded binary saliency maps in some applications, such as content-aware image scaling. In such cases, the MAE does a thorough comparison of the saliency map and the ground truth. The MAE is calculated as the difference between the ground truth and the normalized final saliency map $Smap_{x,y}^{FN}$, which is normalized in the range [0, 1]. In Table 3 the terms S_o and S_r in Sm calculation are the object aware and region aware structural similarity which can be obtained from [48]. The term $\Psi(x, y)$ in Em calculation is enhanced alignment matrix which can be obtained using [49]. Sm score analyses structural similarity between real-valued saliency map and the binary ground-truth, instead of just pixel-wise errors. Em score takes into account the global average of the image as well as local pixel matching simultaneously.

C. QUALITATIVE ANALYSIS

The saliency maps produced by the proposed and other saliency detection methods are subjectively evaluated in the qualitative analysis based on the degree of separation between foreground and background regions, the portion of the salient object to be highlighted, uniformity in recognizing various regions, detection of salient objects in extremely complicated and poor contrast backgrounds, and correctly identifying more than one salient object. Experimental results are shown in Fig. 2 - Fig. 6 for DUT-OMRON, ECSSD, HKUIS, SOD, and MSRA10K datasets, respectively. It can be observed from Fig. 2 - Fig. 6 - Image 1 and Image 2 that CA [50] and SEG [52] methods are giving blurred results with loss of boundary preservation. Fig. 2 - Image 1 of DUT-OMRON dataset indicates that CA [50], GR [51], MR [35], MC [53], LPS [54], RR [55], LGF [56], DPSG [57], NCUT [58], RCRR [59] and SMD [60] methods are unable to detect all the objects present in the image. In most of the cases the man holding the cloth is not detected, while the proposed method is able to detect multiple salient objects present in an image accurately with preserving the boundaries. Fig. 2 - Image 2 only the proposed method can detect all the salient objects present in the image by preserving the fine details as compared to other state of the art methods. In Fig. 3 - Image 1 and Image 2 the proposed method is able to detect the salient objects with preserving the fine details of an image as compared to CA [50], GR [51], SEG [52], MR [35], LPS [54], RR [55], LGF [56], NCUT [58] and RCRR [59] methods.
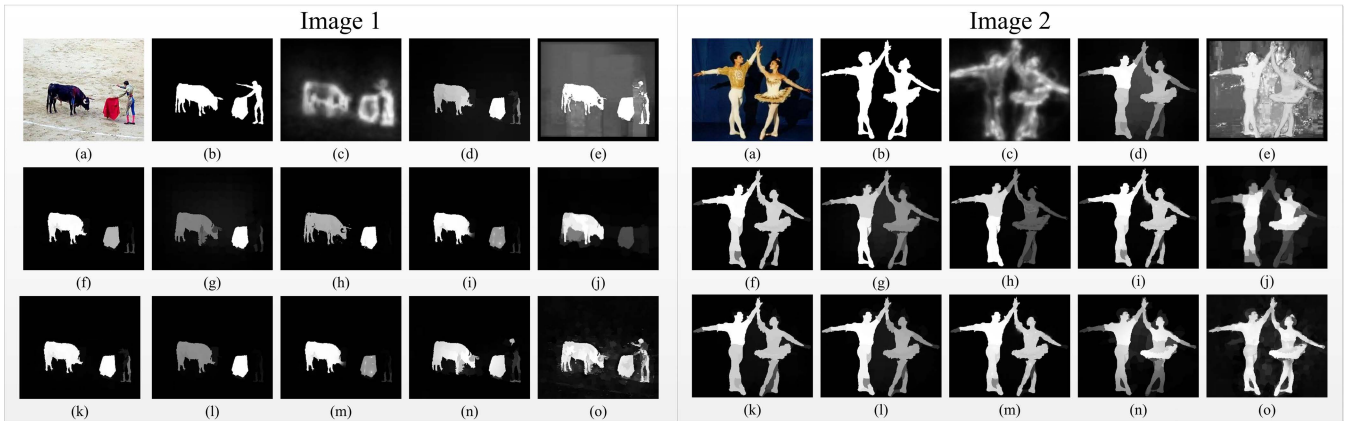
FIGURE 2. Qualitative analysis of the proposed algorithm on DUT-OMRON dataset compared to other salient object detection techniques (a) Original Image (b) Groundtruth (c) CA [50] (d) GR [51] (e) SEG [52] (f) MR [35] (g) MC [53] (h) LPS [54] (i) RR [55] (j) LGF [56] (k) DPSG [57] (l) NCUT [58] (m) RCRR [59] (n) SMD [60] (o) Proposed.

FIGURE 3. Qualitative analysis of the proposed algorithm on ECSSD dataset compared to other salient object detection techniques (a) Original Image (b) Groundtruth (c) CA [50] (d) GR [51] (e) SEG [52] (f) MR [35] (g) MC [53] (h) LPS [54] (i) RR [55] (j) LGF [56] (k) DPSG [57] (l) NCUT [58] (m) RCRR [59] (n) SMD [60] (o) Proposed.
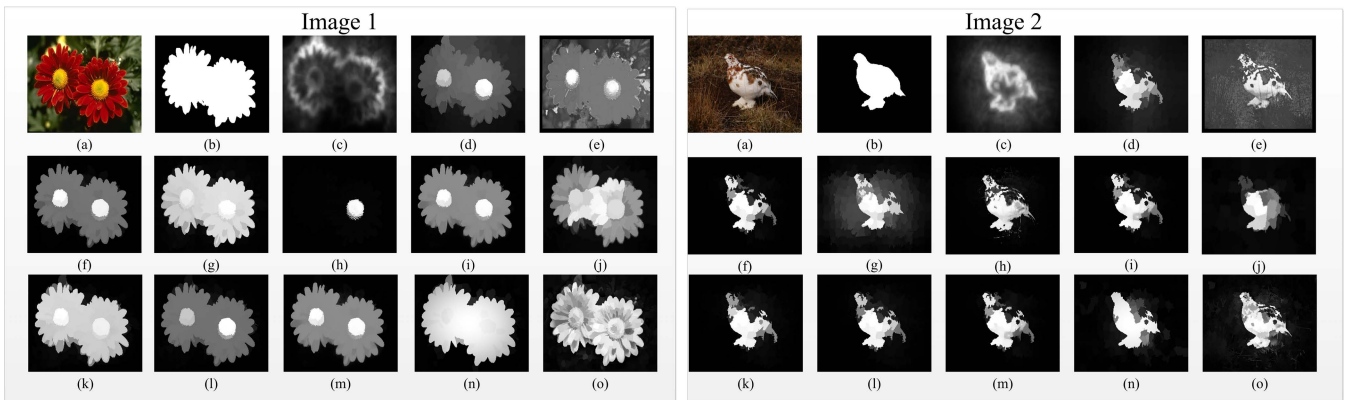

FIGURE 4. Qualitative analysis of the proposed algorithm on HKUIS dataset compared to other salient object detection techniques (a) Original Image (b) Groundtruth (c) CA [50] (d) GR [51] (e) SEG [52] (f) MR [35] (g) MC [53] (h) LPS [54] (i) RR [55] (j) LGF [56] (k) DPSG [57] (l) NCUT [58] (m) RCRR [59] (n) SMD [60] (o) Proposed.
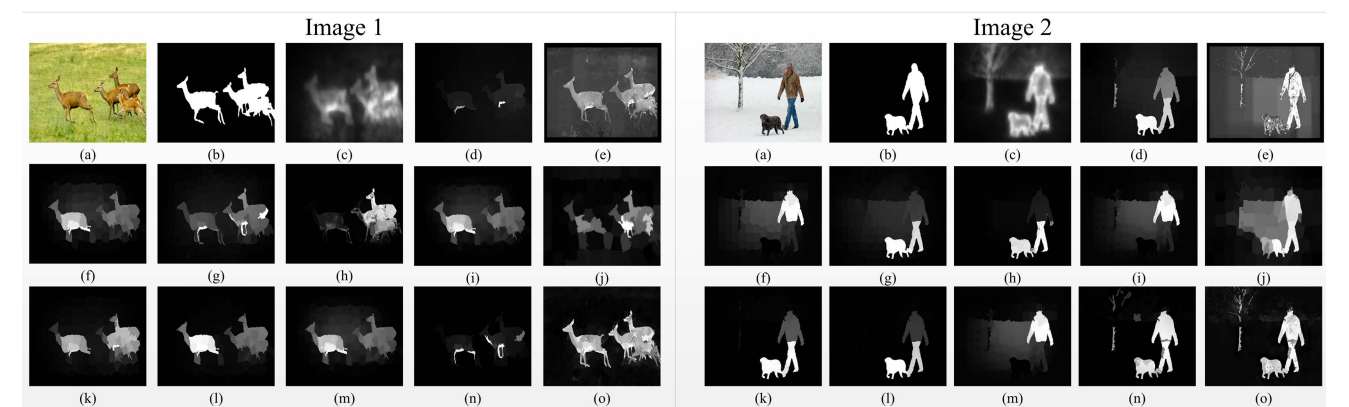
While the methods MC [53], DPSG [57] and SMD [60] gives the comparable results to the proposed method. Fig. 3 - Image 2 shows the proposed method is accurately detecting the salient object as compared to other methods even the color of the object and background is not so different. In comparison to all the mentioned state-of-the-art approaches, the
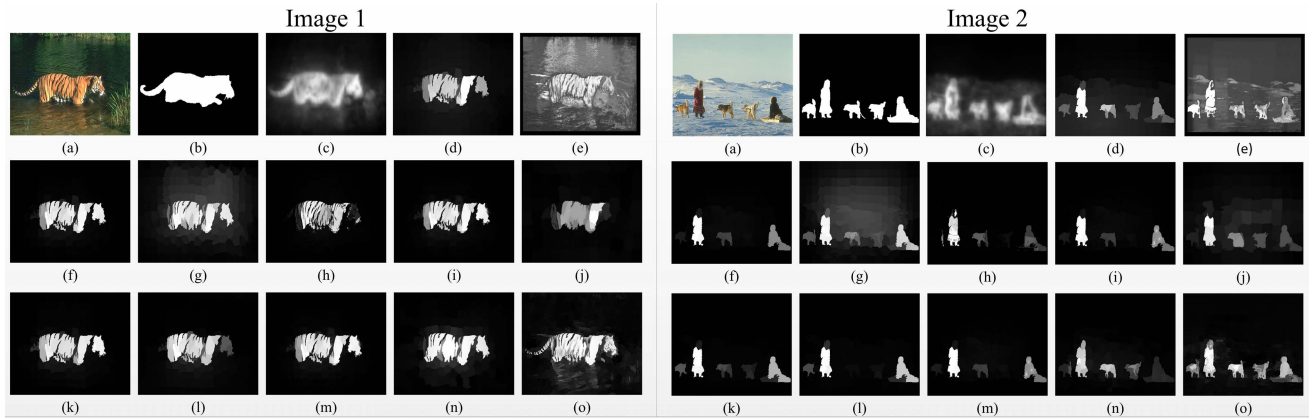
Image 1   Image 2

(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

(k) (l) (m) (n) (o)

FIGURE 5. Qualitative analysis of the proposed algorithm on SOD dataset compared to other salient object detection techniques (a) Original Image (b) Groundtruth (c) CA [50] (d) GR [51] (e) SEG [52] (f) MR [35] (g) MC [53] (h) LPS [54] (i) RR [55] (j) LGF [56] (k) DPSG [57] (l) NCUT [58] (m) RCRR [59] (n) SMD [60] (o) Proposed.

Image 1   Image 2

(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

(k) (l) (m) (n) (o)

FIGURE 6. Qualitative analysis of the proposed algorithm on MSRA10K dataset compared to other salient object detection techniques (a) Original Image (b) Groundtruth (c) CA [50] (d) GR [51] (e) SEG [52] (f) MR [35] (g) MC [53] (h) LPS [54] (i) RR [55] (j) LGF [56] (k) DPSG [57] (l) NCUT [58] (m) RCRR [59] (n) SMD [60] (o) Proposed.
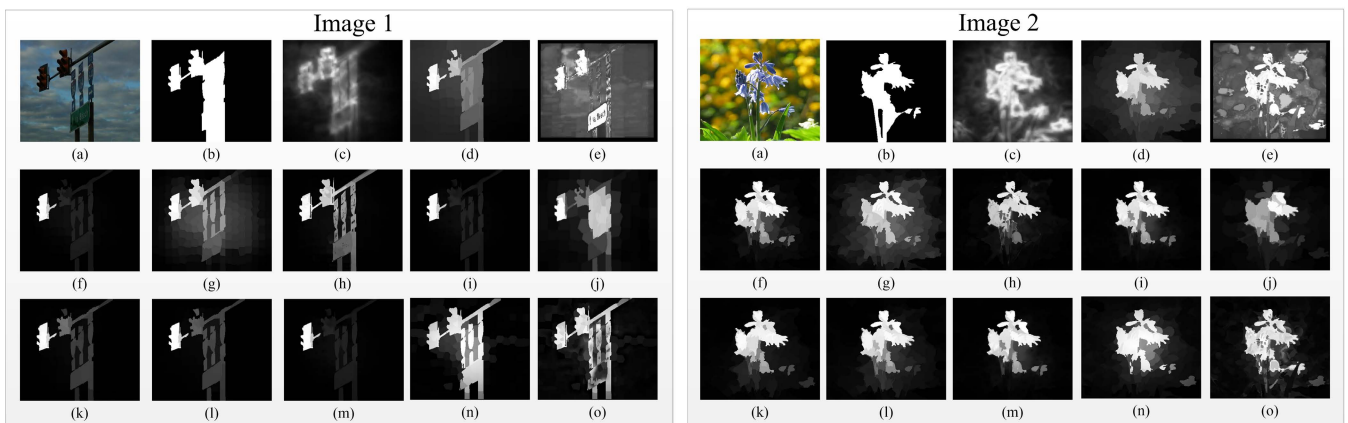
TABLE 4. Quantitative analysis of the proposed algorithm with other saliency detection techniques based on the MAE, F-measure and AUC values on five benchmark saliency detection datasets. (Red, green, and blue highlight the three leading models, respectively. AUC and Fm score higher is better. MAE score lower is better.).

Datasets	DUT-OMRON [35]			HKUIS [43]			ECSSD [42]			SOD [41]			MSRA10K [44]		
Evaluation metric	AUC	Fm	MAE	AUC	Fm	MAE	AUC	Fm	MAE	AUC	Fm	MAE	AUC	Fm	MAE
CA [50]	0.771	0.355	0.254	0.77	0.46	0.274	0.726	0.431	0.309	0.717	0.449	0.313	0.789	0.54	0.247
GR [51]	0.804	0.486	0.259	0.802	0.564	0.256	0.784	0.512	0.283	0.719	0.442	0.318	0.843	0.679	0.216
SEG [52]	0.778	0.394	0.337	0.774	0.464	0.323	0.746	0.394	0.341	0.681	0.316	0.36	0.795	0.515	0.314
MR [35]	0.781	0.48	0.187	0.787	0.517	0.178	0.79	0.691	0.189	0.714	0.57	0.259	0.823	0.833	0.126
MC [53]	0.82	0.533	0.186	0.831	0.678	0.184	0.81	0.699	0.202	0.75	0.59	0.261	0.843	0.836	0.145
LPS [54]	0.767	0.539	0.187	0.752	0.619	0.164	0.782	0.644	0.196	0.703	0.526	0.261	0.807	0.812	0.124
RR [55]	0.782	0.529	0.185	0.796	0.665	0.172	0.796	0.697	0.184	0.713	0.575	0.26	0.826	0.766	0.145
LGF [56]	0.811	0.532	0.178	0.819	0.665	0.196	0.803	0.708	0.196	0.733	0.592	0.256	0.846	0.835	0.145
DPSG [57]	0.796	0.559	0.177	0.808	0.694	0.165	0.795	0.701	0.181	0.71	0.571	0.251	0.839	0.857	0.114
NCUT [58]	0.798	0.568	0.174	0.799	0.679	0.167	0.797	0.69	0.186	0.721	0.577	0.248	0.839	0.850	0.114
RCRR [59]	0.78	0.527	0.182	0.796	0.664	0.171	0.794	0.696	0.184	0.717	0.577	0.257	0.827	0.767	0.144
SMD [60]	0.809	0.538	0.166	0.823	0.691	0.156	0.811	0.713	0.173	0.737	0.61	0.235	0.847	0.858	0.104
FCB [61]	0.668	0.480	0.149	-	-	-	0.702	0.607	0.173	0.641	0.520	0.244	-	0.759	0.121
LSP [62]	0.803	0.546	0.133	-	-	-	0.808	0.761	0.152	0.732	0.661	0.226	-	0.856	0.096
BSDL [16]	0.668	-	0.145	-	-	-	0.811	-	0.098	0.741	-	0.198	-	-	-
CDHL [7]	0.806	0.709	0.153	-	-	-	0.824	0.581	0.143	-	-	-	-	-	-
TMR [13]	-	0.592	0.159	-	-	-	-	0.708	0.143	-	-	-	0.845	0.076
Ours	0.83	0.693	0.174	0.836	0.698	0.165	0.811	0.71	0.2	0.75	0.63	0.242	0.879	0.864	0.098

suggested method detects numerous salient objects with boundary preservation. These approaches can preserve borders to some extent but fail to recognize many salient objects, as can be seen in Fig. 4 - Image 1. From Fig. 4 - Image 1 it can be seen that only the proposed method can detect all the deer present in the image with correct boundary details. The SMD
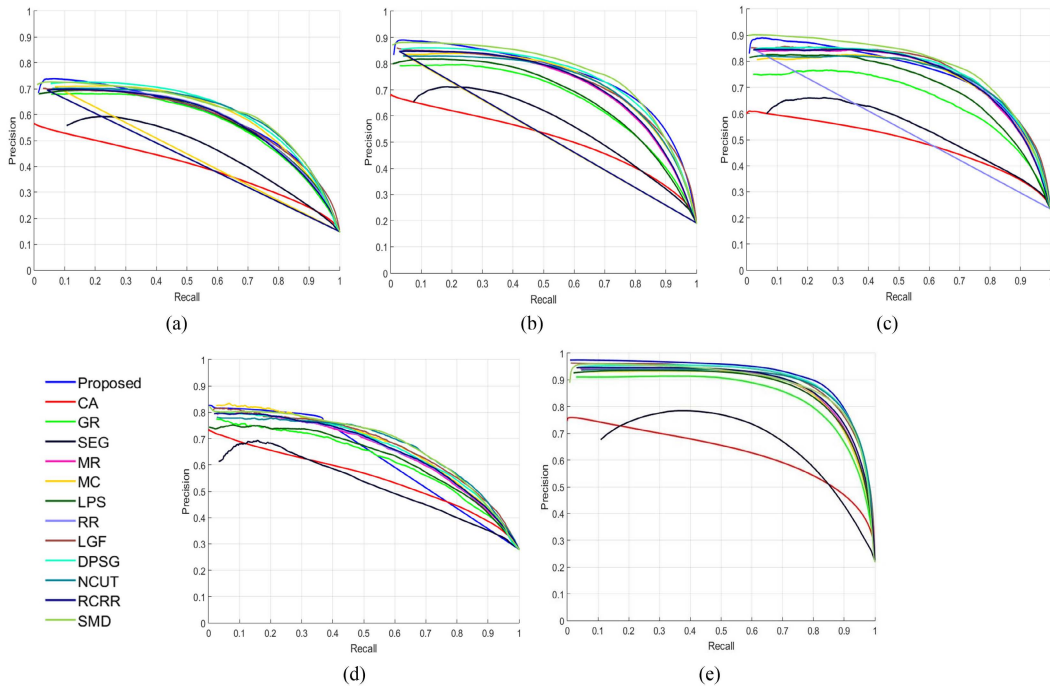
FIGURE 7. Quantitative analysis of the proposed algorithm based on the PR curve on five benchmark saliency detection datasets (a) DUT-OMRON [35] (b) HKUIS [43] (c) ECSSD [42] (d) SOD [41] (e) MSRA10K [44].

FIGURE 8. Quantitative analysis of the proposed algorithm based on the ROC curve on five benchmark saliency detection datasets (a) DUT-OMRON [35] (b) HKUIS [43] (c) ECSSD [42] (d) SOD [41] (e) MSRA10K [44].
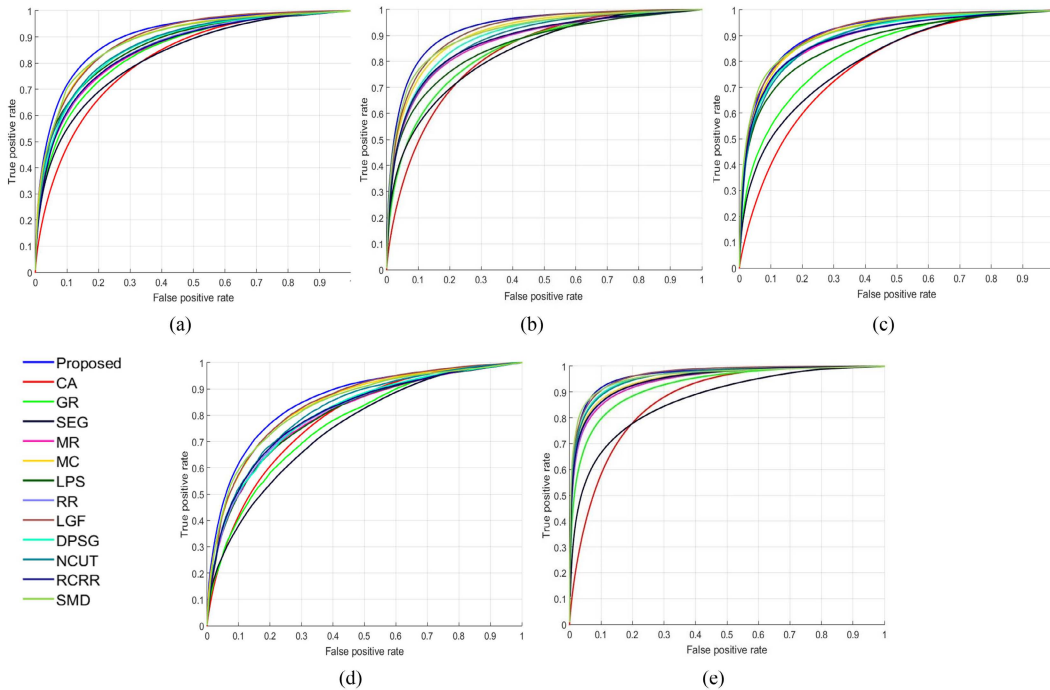
[60] technique in Fig. 4 - Image 2 can detect numerous salient objects, but not in the image displayed in Fig. 4 - Image 1. In the HKUIS dataset, the proposed technique detects several salient objects for all types of images. Fig. 5 - Image 1

indicates that the proposed technique correctly recognizes the entire salient object, unlike all previous methods, which miss the tail section of the object. Fig. 5 - Image 2 shows an image with 5 salient objects, with only the proposed method

accurately detecting all of them when compared to all other methods. Fig. 6 - Image 1 and Image 2 shows the images with complex background. It can be seen from the figures only the proposed method and SMD [60] method can provide qualitatively good results by detecting objects to a great extent. From Fig. 6 - Image 2 it can be seen that the proposed method is preserving the boundaries of the object as compared to other salient object detection techniques, specifically as compared to SMD [60] which is giving competitive results. According to the qualitative study, the proposed method surpasses existing methods for all datasets covering all conditions. The proposed method, in comparison to other saliency methods, generates a high-resolution saliency output on various difficult natural images. When compared to other approaches, the suggested method generates a saliency map that uniformly highlights salient regions while effectively suppressing background regions. The proposed method completely preserves the boundary and sharp details of images. Even in complicated background images, it accurately separates background and foreground and detects the salient sections by preserving the edges and boundaries of the objects. It also detects many important objects with more accuracy than other approaches.

D. QUANTITATIVE ANALYSIS

A quantitative evaluation of the proposed method as well as other existing salient object detection techniques is also carried out on five benchmark salient object detection datasets listed in Table 2. Figs. 7 and 8 demonstrate the relative performance of the proposed method with other saliency detection methods using the PR curve, and ROC curve, respectively. The proposed technique has the best performance in terms of PR curve on the DUT-OMRON, HKUIS, MSRA10K and SOD datasets and comparable performance on the ECSSD dataset, as shown in Fig. 7. Furthermore, in terms of ROC curves presented in Fig. 8, the proposed method beats other methods and provides the best performance for practically all datasets. In Table 4, a comparison analysis based on MAE, F-measure (Fm), and AUC scores is also done and represented. The proposed technique earns the highest AUC score for all the datasets except the ECSSD dataset where it is in the second-highest position, demonstrating its superiority in reliably differentiating the background and foreground areas in salient object identification tasks. The proposed technique also obtains the MAE score and F-measure score in the top three positions as compared to other techniques, as shown in Table 4. For all the datasets, the proposed method outperforms most other methods in terms of AUC value and obtains MAE and F-measure values that are equivalent to the best-performing methods. Table 4 shows that the proposed technique performs much better than the other approaches investigated. The following points can be interpreted by comparing the results in Table 4:

- In comparison to the proposed method, the methods CA [50], SEG [52], and GR [51] perform poorly in terms of AUC, MAE and Fm score.

- The proposed method gives the highest AUC score on all the datasets except the ECSSD dataset where it gives the second-highest performance as compared to CDHL [7] technique.
- The proposed method achieves first, second, or third highest performance in terms of Fm score, while MAE score is comparable to top-performing methods as compared to existing salient object detection methods. The overall performance of the proposed method is better or comparable to all the mentioned salient object detection techniques.
- For DUT-OMRON dataset: the proposed method improves AUC score by ≈1.22% than MC [53] and 24.25% than BSDL [16] and FCB [61] methods; the proposed method improves Fm score by ≈14.57% than TMR [13] and 44.38% than FCB [61] and MR [35] methods; the proposed method improves MAE score by ≈1.72% than DPSG [57] and 7.4% than LPS [54] and MR [35] methods.
- For HKUIS dataset: the proposed method improves AUC score by ≈0.6% than MC [53] and 11.17% than LPS [54] method; the proposed method improves Fm score by ≈0.57% than DPSG [57] and 35% than MR [35] method; the proposed method improves MAE score by ≈1.21% than NCUT [58] and 11.51% than MC [53] method.
- For ECSSD dataset: the proposed method improves AUC score by ≈0.12% than MC [53] and 15.53% than FCB [61] method; the proposed method improves Fm score by ≈0.28% than TMR [13] and LGF [56] and 22.2% than CDHL [7] method; the proposed method improves MAE score by ≈1% than MC [53] method.
- For SOD dataset: the proposed method improves AUC score by ≈1.21% than BSDL [16] and 17% than FCB [61] method; the proposed method improves Fm score by ≈3.28% than SMD [60] and 21.15% than FCB [61] method; the proposed method improves MAE score by ≈0.83% than FCB [61] and 7.85% than MC [53] and LPS [54] methods.
- For MSRA10K dataset: the proposed method improves AUC score by ≈3.78% than SMD [60] and 8.92% than LPS [54] method; the proposed method improves Fm score by ≈0.7% than SMD [60] and 13.83% than FCB [61] method; the proposed method improves MAE score by ≈6.12% than SMD [60] and 47.96% than MC [53], RR [55] and LGF [56] methods.

The metrics discussed in Table 4 are only evaluated at the pixel level and do not account for the structural similarity of the objects. In Table 5, a comparison analysis based on S-measure (Sm) score, and E-measure (Em) score is represented which can take into account structural details as well as global shape. It can be observed from Table 5, that the proposed method gives the highest Em and Sm values for HKUIS and SOD datasets. For DUT-OMRON, ECSSD and MSRA10K datasets, the proposed method achieves performance in the top three positions as compared to other

TABLE 5. Quantitative analysis of the proposed algorithm with other saliency detection techniques based on the S-measure and E-measure values on five benchmark saliency detection datasets. (Red, green, and blue highlight the three leading models, respectively. Sm and Em score higher is better.).

Dataset	DUT-OMRON [35]		HKUIS [43]		ECSSD [42]		SOD [41]		MSRA10K [44]	
Evaluation metric	Sm	Em	Sm	Em	Sm	Em	Sm	Em	Sm	Em
CA [50]	0.536	0.654	0.572	0.696	0.543	0.65	0.538	0.613	0.605	0.739
GR [51]	0.601	0.686	0.622	0.687	0.619	0.627	0.556	0.554	0.702	0.762
SEG [52]	0.521	0.636	0.562	0.605	0.553	0.539	0.513	0.45	0.610	0.596
MR [35]	0.49	0.612	0.591	0.732	0.693	0.778	0.583	0.665	0.785	0.877
MC [53]	0.649	0.728	0.684	0.804	0.692	0.787	0.603	0.687	0.785	0.879
LPS [54]	0.652	0.732	0.63	0.753	0.643	0.745	0.536	0.641	0.755	0.85
RR [55]	0.648	0.723	0.677	0.792	0.697	0.782	0.586	0.668	0.741	0.846
LGF [56]	0.647	0.73	0.678	0.801	0.687	0.789	0.584	0.685	0.778	0.874
DPSG [57]	0.664	0.741	0.691	0.804	0.704	0.784	0.589	0.663	0.808	0.895
NCUT [58]	0.669	0.745	0.682	0.788	0.696	0.773	0.602	0.663	0.805	0.886
RCRR [59]	0.649	0.721	0.677	0.792	0.695	0.782	0.588	0.668	0.742	0.847
SMD [60]	0.681	0.728	0.724	0.815	0.734	0.799	0.628	0.699	0.839	0.905
FCB [61]	0.614	0.673	-	-	0.651	0.742	-	-	0.718	0.793
TMR [13]	0.663	0.734	-	-	0.724	0.776	-	-	0.837	0.889
Ours	0.675	0.736	0.733	0.824	0.728	0.79	0.656	0.71	0.841	0.898

TABLE 6. Quantitative analysis of the proposed algorithm with deep learning based saliency detection techniques based on the MAE and AUC values on four benchmark saliency detection datasets (Bold values indicate highest performance.).

| Datasets	DUT-OMRON [35]		HKUIS [43]		ECSSD [42]		SOD [41]	
Evaluation metric	AUC	MAE	AUC	MAE	AUC	MAE	AUC	MAE
KSR16 [63]	0.825	0.131	0.824	0.12	0.823	0.134	0.739	0.199
RSD-R [64]	0.812	0.178	0.867	0.155	0.845	0.172	0.753	0.226
AMU17 [65]	0.867	0.098	0.882	0.051	0.861	0.059	**0.789**	0.142
MSRNet [66]	**0.869**	0.073	**0.889**	0.036	0.862	0.054	0.784	0.113
PAGR18 [67]	0.813	0.071	0.867	0.047	0.849	0.061	0.744	0.147
WSS [68]	0.824	0.11	0.852	0.079	0.825	0.104	0.739	0.17
SSNet [69]	0.773	**0.056**	0.819	0.041	0.813	0.046	0.692	0.118
BANet-R [70]	0.865	0.059	0.879	**0.032**	0.863	**0.035**	0.775	0.106
EGNet-R [71]	-	-	-	-	0.865	0.037	0.782	**0.099**
HRSOD-DH [72]	0.782	0.065	0.851	0.042	0.839	0.052	0.716	0.139
JDFPR [73]	0.852	0.057	0.881	0.039	0.862	0.049	0.768	0.123
SCRN [74]	0.864	**0.056**	0.883	0.034	**0.868**	0.037	0.781	0.107
Ours	0.83	0.174	0.836	0.165	0.811	0.2	0.75	0.242
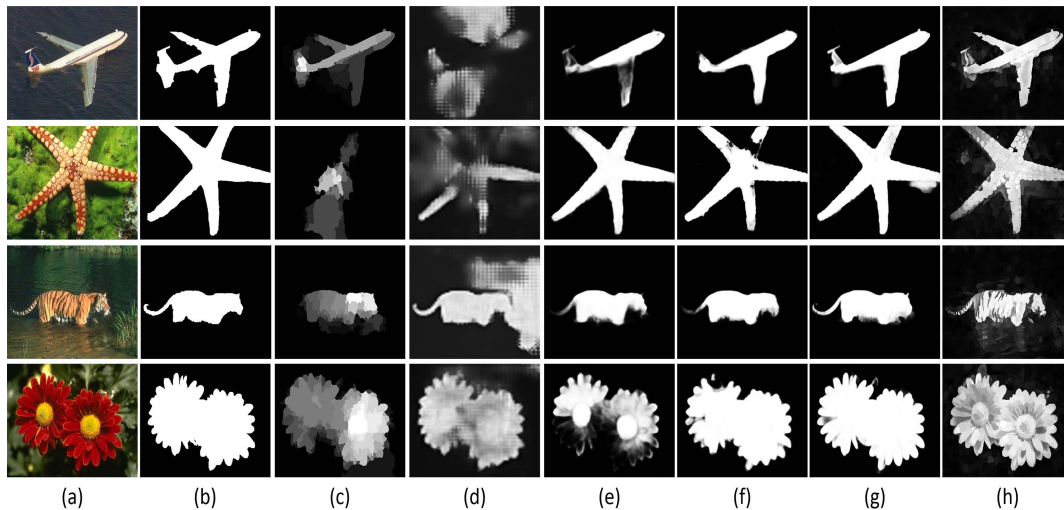
FIGURE 9. Performance analysis of the proposed method with CNN based methods (a) Original (b) GT (c) KSR16 [63] (d) RSD-R [64] (e) PAGR18 [67] (f) HRSOD-DH [72] (g) EGNet-R [71] (h) Proposed.

mentioned salient object detection methods. This shows the superiority of the proposed method in retaining the structural details of the objects as compared to other salient object detection techniques.

- For DUT-OMRON dataset: the proposed method improves Sm score by ≈0.89% than NCUT [58] and

37.75% than MR [35] method. The proposed method improves Em score by ≈0.27% than TMR [13] and 20.26% than MR [35] method.

- For HKUIS dataset: the proposed method improves Sm score by ≈1.24% than SMD [60] and 24.03% than MR [35] method. The proposed method improves Em

score by ≈1.1% than SMD [60] and 12.57% than MR [35] method.

- For ECSSD dataset: the proposed method improves Sm score by ≈0.55% than TMR [13] and 11.83% than FCB [61] method. The proposed method improves Em score by ≈0.13% than LGF [56] and 6.47% than FCB [61] method.
- For SOD dataset: the proposed method improves Sm score by ≈4.46% than SMD [60] and 22.39% than LPS [54] method. The proposed method improves Em score by ≈1.57% than SMD [60] and 10.76% than LPS [54] method.
- For MSRA10K dataset: the proposed method improves Sm score by ≈0.24% than SMD [60] and 17.13% than FCB [61] method. The proposed method improves Em score by ≈0.34% than DPSG [57] and 13.24% than FCB [61] method.

E. PERFORMANCE COMPARISON WITH DEEP LEARNING BASED METHODS

Furthermore, the overall performance of the proposed method is compared with recently proposed deep leaning-based approaches as KSR16 [63], RSD-R [64], AMU17 [65], MSRNet [66], PAGR18 [67], WSS [68], SSNet [69], BANet-R [70], EGNet-R [71], HRSOD-DH [72], JDFPR [73] and SCRN [74]. The comparison of the mentioned deep learning-based techniques with the proposed technique based on MAE and AUC scores is presented in Table 6. The proposed approach achieves equal or better results than deep learning-based techniques, as shown in Table 6. Considering the dataset's reputation for challenging images with crowded backgrounds, low contrast images, and images with many objects, the proposed method can give significantly better results than some of the deep learning-based techniques like HRSOD-DH [72], SSNet [69], PAGR18 [67], KSR16 [63] methods in terms of AUC score. For the DUT-OMRON dataset the proposed method gives ≈6.14%, 7.37%, 0.73%, 2.1%, 2.22%, and 0.61% higher AUC value as compared to HRSOD-DH [72], SSNet [69], WSS [68], PAGR18 [67], RSD-R [64], and KSR16 [63] methods respectively. For the DUT-OMRON dataset, the proposed method gives ≈2.3% lower MAE value as compared to RSD-R [64] method. For the HKUIS dataset the proposed method gives ≈2.08%, and 1.46% higher AUC value as compared to SSNet [69], and KSR16 [63] methods respectively. For the SOD dataset the proposed method gives ≈4.75%, 8.38%, 1.5%, 0.81%, and 1.5% higher AUC value as compared to HRSOD-DH [72], SSNet [69], WSS [68], PAGR18 [67], and KSR16 [63] methods respectively. It is evident from Fig. 9 that the proposed method is more efficient in detecting complete objects in images with cluttered backgrounds by preserving the fine details and boundaries of objects as compared to many of the deep learning-based methods. From Fig. 9 it can be observed that though deep learning-based methods give a good score in terms of MAE score and AUC score, in some cases they fail to detect the complete object in an image. While the

TABLE 7. Computational performance of proposed method with other saliency integration methods. Bold value is used to indicate the lowest computational time.

Methods	SIHCA [31]	SIAM [30]	Ours
Time (seconds)	0.26	1.07	**0.06**

proposed method can preserve the boundaries and fine details of the complete object as compared to other deep learning-based methods. So the proposed method is efficient in the task where less complex saliency architecture with accurate detection of objects by preserving complete boundary is more important. Also, the computational time required to train the deep learning-based model is very high as compared to the machine learning-based model. The method proposed by Qin *et al.* [75] which is based on saliency detection with boundary preservation takes 145 hours to train the model, whereas the model based on random forest regression takes only 3 to 4 hours for training which is comparatively very less. So the proposed boundary preserved saliency detection model is more computationally efficient as compared to deep learning-based techniques.

V. RUN TIME

The proposed method generates a saliency map in 1.92 seconds on a 64-bit PC with an i7-4770 3.40 GHz processor and 32.0 GB RAM. The test is performed on a 300 × 400 image. All of the procedures are executed in MATLAB 2017a. The integration of transformation and learning-based saliency maps using guided filters takes around 0.06 seconds. The computational performance of saliency integration time (which include only the integration time of different saliency maps) has been presented in Table 7 where the proposed method minimizes the average integration time by ≈76.92% and 94.39% as compared to SIHCA [31] and SIAM [30] methods, respectively. So Table 7 indicates that the proposed method is computationally very fast and efficient in integrating transformation-based and learning-based saliency maps than all compared integration methods in Table 7. The overall final saliency map generation time of the proposed method is also very less.

VI. LIMITATIONS

In this paper, the saliency detection task as saliency integration using attention map-based guided filters for integrating transformation-based and learning-based saliency maps are incorporated. In the proposed saliency detection task, when the transformation-based and learning-based saliency maps are unable to detect the particular salient region at the same time, the proposed model fails to detect that region as well, which is illustrated in Fig. 10. Also, it can be observed from Fig. 10 that the proposed method sometimes detects the background as an object which affects the performance of the proposed method. The limitations of the proposed algorithm can be avoided by adding some class-specific objects like humans or vehicles etc. in the saliency models to improve
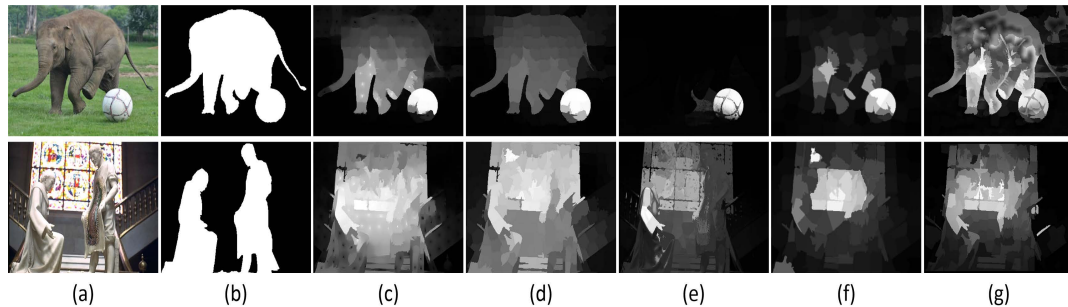
FIGURE 10. Limitations (a) Original (b) GT (c) RCRR [59] (d) DPSG [57] (e) LPS [54] (f) LGF [56] (g) Proposed.

the accuracy for specific vehicle objects or human object detection tasks. Also, concatenation of different deep features along with the features used in the proposed method to train the random forest regression can improve the performance of the proposed technique.

VII. CONCLUSION

This paper presents a novel salient object detection technique that integrates the wavelet-based and learning-based saliency maps using the edge-preserving guided filter model. SLIC superpixel segmentation is employed initially for the input image to speed up the computations. The local and global wavelet features are used to generate local and global saliency maps, which are finally integrated to produce the wavelet-based saliency map. While color, spatial distance, and textural distance features are used to generate the learning-based saliency map. The proposed method effectively merges the wavelet-based and learning-based saliency maps which takes into account the human visual consistent features of the wavelet-based and learning-based saliency maps by preserving the boundaries of objects. The use of guided filters takes into account the neighbor-hood pixel variations and helps to preserve the object boundaries without introducing blurring and artifacts. The outcome of substantial experiments carried out on a variety of datasets demonstrates that the suggested combination of wavelet-based and learning-based saliency maps outperforms other existing salient object detection approaches.

REFERENCES

[1] C. Zhu, K. Huang, and G. Li, "An innovative saliency guided ROI selection model for panoramic images compression," in *Proc. Data Compress. Conf.*, Mar. 2018, p. 436.

[2] R. Cong, J. Lei, H. Fu, M.-M. Cheng, W. Lin, and Q. Huang, "Review of visual saliency detection with comprehensive information," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 10, pp. 2941–2959, Oct. 2018.

[3] C. Zhu, W. Yan, X. Cai, S. Liu, T. H. Li, and G. Li, "Neural saliency algorithm guide bi-directional visual perception style transfer," *CAAI Trans. Intell. Technol.*, vol. 5, no. 1, pp. 1–8, Mar. 2020.

[4] Y. Ji, H. Zhang, Z. Zhang, and M. Liu, "CNN-based encoder–decoder networks for salient object detection: A comprehensive review and recent advances," *Inf. Sci.*, vol. 546, pp. 835–857, Feb. 2021.

[5] C. Zhu, W. Yan, S. Liu, T. Li, and G. Li, "Salient contour-aware based twice learning strategy for saliency detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 2541–2548.

[6] H. Wang, C. Zhu, J. Shen, Z. Zhang, and X. Shi, "Salient object detection by robust foreground and background seed selection," *Comput. Electr. Eng.*, vol. 90, Mar. 2021, Art. no. 106993.

[7] F. Wang and G. Peng, "Saliency detection based on color descriptor and high-level prior," *Mach. Vis. Appl.*, vol. 32, no. 6, pp. 1–12, Nov. 2021.

[8] C. Zhu, G. Li, W. Wang, and R. Wang, "An innovative salient object detection using center-dark channel prior," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 1509–1515.

[9] N. Tong, H. Lu, X. Ruan, and M.-H. Yang, "Salient object detection via bootstrap learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1884–1892.

[10] Z. S. Shariatmadar and K. Faez, "Visual saliency detection via integrating bottom-up and top-down information," *Optik*, vol. 178, pp. 1195–1207, Feb. 2019.

[11] Y. Ji, H. Zhang, Z. Jie, L. Ma, and Q. M. J. Wu, "CASNet: A cross-attention Siamese network for video salient object detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 6, pp. 2676–2690, Jul. 2020.

[12] G. Li and C. Zhu, "A three-pathway psychobiological framework of salient object detection using stereoscopic technology," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 3008–3014.

[13] X. Sun, X. Zhang, C. Xu, M. Xiao, and Y. Tang, "Tensorial multiview representation for saliency detection via nonconvex approach," *IEEE Trans. Cybern.*, early access, Jan. 13, 2022, doi: 10.1109/TCYB.2021.3139037.

[14] W. Wang, J. Shen, X. Dong, and A. Borji, "Salient object detection driven by fixation prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1711–1720.

[15] W. Wang, J. Shen, X. Dong, A. Borji, and R. Yang, "Inferring salient objects from human fixations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 1913–1927, Aug. 2019.

[16] Y. Pang, X. Yu, Y. Wu, C. Wu, and Y. Jiang, "Bagging-based saliency distribution learning for visual saliency detection," *Signal Process., Image Commun.*, vol. 87, Sep. 2020, Art. no. 115928.

[17] Q. Zhang, J. Lin, W. Li, Y. Shi, and G. Cao, "Salient object detection via compactness and objectness cues," *Vis. Comput.*, vol. 34, no. 4, pp. 473–489, 2018.

[18] A. Borji, M.-M. Cheng, Q. Hou, H. Jiang, and J. Li, "Salient object detection: A survey," *Comput. Vis. Media*, vol. 5, no. 2, pp. 117–150, 2019.

[19] L. Zhang, J. Ai, B. Jiang, H. Lu, and X. Li, "Saliency detection via absorbing Markov chain with learnt transition probability," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 987–998, Feb. 2017.

[20] C. Zhu, X. Cai, K. Huang, T. H. Li, and G. Li, "PDNet: Prior-model guided depth-enhanced network for salient object detection," in *Proc. IEEE Int. Conf. Multimedia Exp. (ICME)*, Jul. 2019, pp. 199–204.

[21] C. Guo, Q. Ma, and L. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[22] J. Li, M. D. Levine, X. An, X. Xu, and H. He, "Visual saliency based on scale-space analysis in the frequency domain," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 996–1010, Apr. 2012.

[23] N. Murray, M. Vanrell, X. Otazu, and C. A. Parraga, "Saliency estimation using a non-parametric low-level vision model," in *Proc. CVPR*, 2011, pp. 433–440.

[24] F. Zhang, T.-Y. Wu, and G. Zheng, "Video salient region detection model based on wavelet transform and feature comparison," *EURASIP J. Image Video Process.*, vol. 2019, no. 1, pp. 1–10, Dec. 2019.

[25] N. Imamoglu, W. Lin, and Y. Fang, "A saliency detection model using low-level features based on wavelet transform," *IEEE Trans. Multimedia*, vol. 15, no. 1, pp. 96–105, Jan. 2012.

[26] V. K. Singh and N. Kumar, "Saliency bagging: A novel framework for robust salient object detection," *Vis. Comput.*, vol. 36, no. 7, pp. 1423–1441, Jul. 2020.

[27] J. Lei, B. Wang, Y. Fang, W. Lin, P. Le Callet, N. Ling, and C. Hou, "A universal framework for salient object detection," *IEEE Trans. Multimedia*, vol. 18, no. 9, pp. 1783–1795, Sep. 2016.

[28] J. Yang and M.-H. Yang, "Top-down visual saliency via joint CRF and dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 576–588, Mar. 2016.

[29] F. Huang, J. Qi, H. Lu, L. Zhang, and X. Ruan, "Salient object detection via multiple instance learning," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1911–1922, Apr. 2017.

[30] Y. Xu, X. Hong, F. Porikli, X. Liu, J. Chen, and G. Zhao, "Saliency integration: An arbitrator model," *IEEE Trans. Multimedia*, vol. 21, no. 1, pp. 98–113, Jan. 2018.

[31] Y. Qin, M. Feng, H. Lu, and G. W. Cottrell, "Hierarchical cellular automata for visual saliency," *Int. J. Comput. Vis.*, vol. 126, no. 7, pp. 751–770, 2018.

[32] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2083–2090.

[33] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 733–740.

[34] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 853–860.

[35] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3166–3173.

[36] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[37] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[38] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 353–367, Feb. 2010.

[39] C. Becker, R. Rigamonti, V. Lepetit, and P. Fua, "Supervised feature learning for curvilinear structure segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2013, pp. 526–533.

[40] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.

[41] V. Movahedi and J. H. Elder, "Design and perceptual validation of performance measures for salient object segmentation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2010, pp. 49–56.

[42] J. Shi, Q. Yan, L. Xu, and J. Jia, "Hierarchical image saliency detection on extended CSSD," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 4, pp. 717–729, Apr. 2015.

[43] G. Li and Y. Yu, "Visual saliency based on multiscale deep features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5455–5463.

[44] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, Mar. 2014.

[45] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, 2006, pp. 233–240.

[46] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1597–1604.

[47] A. Borji, M.-M. Cheng, H. Jiang, and J. Li, "Salient object detection: A benchmark," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5706–5722, Dec. 2015.

[48] D.-P. Fan, M.-M. Cheng, Y. Liu, T. Li, and A. Borji, "Structure-measure: A new way to evaluate foreground maps," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4548–4557.

[49] D.-P. Fan, C. Gong, Y. Cao, B. Ren, M.-M. Cheng, and A. Borji, "Enhanced-alignment measure for binary foreground map evaluation," 2018, *arXiv:1805.10421*.

[50] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915–1926, Oct. 2011.

[51] C. Yang, L. Zhang, and H. Lu, "Graph-regularized saliency detection with convex-hull-based center prior," *IEEE Signal Process. Lett.*, vol. 20, no. 7, pp. 637–640, Jul. 2013.

[52] E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä, "Segmenting salient objects from images and videos," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2010, pp. 366–379.

[53] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang, "Saliency detection via absorbing Markov chain," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1665–1672.

[54] H. Li, H. Lu, Z. Lin, X. Shen, and B. Price, "Inner and inter label propagation: Salient object detection in the wild," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3176–3186, Oct. 2015.

[55] C. Li, Y. Yuan, W. Cai, Y. Xia, and D. D. Feng, "Robust saliency detection via regularized random walks ranking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2710–2717.

[56] N. Tong, H. Lu, Y. Zhang, and X. Ruan, "Salient object detection via global and local cues," *Pattern Recognit.*, vol. 48, no. 10, pp. 3258–3267, Oct. 2015.

[57] L. Zhou, Z. Yang, Z. Zhou, and D. Hu, "Salient region detection using diffusion process on a two-layer sparse graph," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5882–5894, Dec. 2017.

[58] K. Fu, C. Gong, I. Y.-H. Gu, and J. Yang, "Normalized cut-based saliency detection by adaptive multi-level region merging," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5671–5683, Dec. 2015.

[59] Y. Yuan, C. Li, J. Kim, W. Cai, and D. D. Feng, "Reversion correction and regularized random walk ranking for saliency detection," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1311–1322, Mar. 2018.

[60] H. Peng, B. Li, H. Ling, W. Hu, W. Xiong, and S. J. Maybank, "Salient object detection via structured matrix decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 818–832, Apr. 2016.

[61] G.-H. Liu and J.-Y. Yang, "Exploiting color volume and color difference for salient region detection," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 6–16, Jan. 2019.

[62] M. Zhang, Y. Pang, Y. Wu, Y. Du, H. Sun, and K. Zhang, "Saliency detection via local structure propagation," *J. Vis. Commun. Image Represent.*, vol. 52, no. 4, pp. 131–142, 2018.

[63] T. Wang, L. Zhang, H. Lu, C. Sun, and J. Qi, "Kernelized subspace ranking for saliency detection," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 450–466.

[64] M. A. Islam, M. Kalash, and N. D. B. Bruce, "Revisiting salient object detection: Simultaneous detection, ranking, and subitizing of multiple salient objects," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7142–7150.

[65] P. Zhang, D. Wang, H. Lu, H. Wang, and X. Ruan, "Amulet: Aggregating multi-level convolutional features for salient object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 202–211.

[66] G. Li, Y. Xie, L. Lin, and Y. Yu, "Instance-level salient object segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2386–2395.

[67] X. Zhang, T. Wang, J. Qi, H. Lu, and G. Wang, "Progressive attention guided recurrent network for salient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 714–722.

[68] L. Wang, H. Lu, Y. Wang, M. Feng, D. Wang, B. Yin, and X. Ruan, "Learning to detect salient objects with image-level supervision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 136–145.

[69] Z. Yu, Y. Zhuge, H. Lu, and L. Zhang, "Joint learning of saliency detection and weakly supervised semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7223–7233.

[70] J. Su, J. Li, Y. Zhang, C. Xia, and Y. Tian, "Selectivity or invariance: Boundary-aware salient object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3799–3808.

[71] J. Zhao, J.-J. Liu, D.-P. Fan, Y. Cao, J. Yang, and M.-M. Cheng, "EGNet: Edge guidance network for salient object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8779–8788.

[72] Y. Zeng, P. Zhang, Z. Lin, J. Zhang, and H. Lu, "Towards high-resolution salient object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7234–7243.

[73] Y. Xu, D. Xu, X. Hong, W. Ouyang, R. Ji, M. Xu, and G. Zhao, "Structured modeling of joint deep feature and prediction refinement for salient object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3789–3798.

[74] Z. Wu, L. Su, and Q. Huang, "Stacked cross refinement network for edge-aware salient object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7264–7273.

[75] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, and M. Jagersand, "BASNet: Boundary-aware salient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7479–7489.

MOHAMMAD FARUKH HASHMI (Senior Member, IEEE) received the B.E. degree in electronics and communication engineering from MIT Mandsaur/RGPV Bhopal University, in 2007, the M.E. degree in digital techniques and instrumentation from the Shri Govindram Seksaria Institute of Technology and Science (SGSITS) (Autonomous State Government) Indore/RGPV Bhopal University, in 2010, and the Ph.D. degree from the Visvesvaraya National Institute of Technology (VNIT), Nagpur, in 2015, under the supervision of Dr. Avinash G. Keskar. He is currently working as an Assistant Professor with the Department of Electronics and Communication Engineering, National Institute of Technology (NIT), Warangal. He has published up to 75 articles including 25 SCI indexed research papers in International/National Journals/Conferences of publishers like IEEE, Elsevier, and Springer. He has also published one patent to his credit. He was a Principal Investigator of one research project of worth five lakhs funded by Institute Seed grant through TEQIP III. He has a teaching and research experience of 13 years. He has supervised two Ph.D. scholars. He is currently guiding four Ph.D. Scholars. His current research interests include computer vision, machine vision, machine learning, deep learning, embedded systems, the Internet of Things, digital signal processing, image processing, and digital IC design. He is a Life Member of IETE, ISTE, and IAENG societies. He is also serving as an active and potential technical reviewer for IEEE ACCESS, *IET Image Processing*, *IET Computer Vision*, *Wireless Personal Communication*, IEEE SYSTEMS JOURNAL, *Sensors* (MDPI), *Electronics* (MDPI), *Diagnostic* (MDPI), *The Visual Computer*, *Applied Soft Computing* (Elsevier), *Colour Research*, *The Journal of Super Computing*, and various other journals like Elsevier/Springer/IEEE TRANSACTIONS publishers of repute.

BHAGYASHREE V. LAD (Student Member, IEEE) was born in 1990. She received the B.Tech. degree in electronics and telecommunications engineering from the Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, India, in 2012, and the M.Tech. degree from the Visvesvaraya National Institute of Technology, Nagpur, India, in 2015, where she is currently pursuing the Ph.D. degree, under the supervision of Dr. A. Keskar. Her research interests include computer vision, machine learning, deep learning, and object detection.

AVINASH G. KESKAR (Senior Member, IEEE) was born in Nagpur, India, in 1959. He received the B.E. degree (Hons.) from the Visvesvaraya National Institute of Technology (VNIT), Nagpur, in 1979, the M.E. degree (Hons.) from the Indian Institute of Science (IISc), Bangalore, in 1983, and the Ph.D. degree from Nagpur University, in 1997. He has 30 years of teaching experience and seven years of industrial experience. He is currently working as a Professor and the Head of the Department of Electronics and Communication Engineering, VNIT. His current research interests include computer vision, soft computing, embedded systems, and fuzzy logic. He is a Senior Member of FIETE, FIE, and LMISTE.

● ● ●