Enhancing Salient Object Detection with Supervised Learning and Multi-prior Integration

Gayathri Dhara and Ravi Kant Kumar *

Department of Computer Science and Engineering, SRM University-AP, Andhra Pradesh, India
Email: gayathri_dhara@srmap.edu.in (G.D.); ravikant.k@srmap.edu.in (R.K.K.)
*Corresponding author

Abstract—**Salient Object Detection (SOD) can mimic the human vision system by using algorithms that simulate the way how the eye detects and processes visual information. It focuses mainly on the visually distinctive parts of an image, similar to how the human brain processes visual information. The approach proposed in this study is an ensemble approach that incorporates classification algorithm, foreground connectivity and prior calculations. It involves a series of preprocessing, feature generation, selection, training, and prediction using random forest to identify and extract salient objects in an image as a first step. Next, an object proposals map is created for the foreground object. Subsequently, a fusion map is generated using boundary, global, and local contrast priors. In the feature generation step, different edge filters are implemented as the saliency score at edges will be high; additionally, with the use of Gabor's filter the texture-based features are calculated. The Boruta feature selection algorithm is then used to identify the most appropriate and discriminative features, which helps to reduce the computational time required for feature selection. Ultimately, the initial map obtained from the random forest, along with the fusion saliency maps based on foreground connectivity and prior calculations, is merged to produce a saliency map. This map is then refined using post-processing techniques to acquire the final saliency map. The approach we propose surpasses the performance of 17 cutting-edge techniques across three benchmark datasets, showcasing superior results in terms of precision, recall, and f-measure. The proposed method performs well even on the DUT-OMRON dataset, known for its multiple salient objects and complex backgrounds, achieving a Mean Absolute Error (MAE) value of 0.113. The method also demonstrates high recall values (0.862, 0.923, 0.849 for ECSSD, MSRA-B and DUT-OMRON datasets, respectively) across all datasets, further establishing its suitability for salient object detection.**

Keywords—**computer vision, salient object detection, random forest, classification, visual attention, visual saliency, video surveillance**

I. Introduction

Computer Vision and Video Surveillance have emerged as crucial research areas in the rapidly developing field of technology. Computer Vision, is a subset of artificial intelligence, enables computers with the ability to interpret and discern visual data. This is achieved through the use of sophisticated cameras and deep learning algorithms, mirroring human vision abilities. This allows machines to accurately recognize and classify objects in digital images and videos. In video surveillance, the capability to recognize salient objects emphasizes items of interest in a scene, such as people, vehicles, or any unusual objects or activities. This feature allows the surveillance system to concentrate on potentially important events, thus improving the surveillance system's efficiency and effectiveness. For more details on the evolution of salient object detection methods, please refer to the related survey papers [1, 2]. The Human Vision System (HVS) uses various sensors to identify and process visual information. It primarily relies on the visual sensor, the eye, which contains specialized sensory cells known as photoreceptors. These photoreceptors translate light into electrical signals that the brain can elucidate as images. The eye's photoreceptors, rods and cones, are responsible for detecting light and colour, respectively. The HVS also comprises the optic nerve, which conveys visual data from the eye to the brain. This data is then interpreted by the visual cortex, transforming the signals received from the eye into images. Our daily visual perceptions are formed through the collaboration of sensors and neural pathways [3]. Visual saliency is when an object, person, or pixel catches our attention by standing out from its surroundings in the visual field. Machines employ a technique known as salient object detection to address the challenge of visual attention that humans can easily handle. The significance of Salient Object Detection (SOD) in computer vision applications stems from its ability to minimize computing complexity [1].

The human brain can quickly process information about the surrounding environment. As we acquire knowledge via our senses, it is important to note that the more profound regions of the brain do not fully analyze every single piece of sensory data that enters. The reason for this is that our perception of information varies in terms of the level of attention and engagement, leading the brain to selectively exclude the majority of the incoming sensory input. Identifying all the exciting targets in the visual field

Manuscript received September 25, 2023; revised December 14, 2023; accepted December 26, 2023; published June 20, 2024.

doi: 10.18178/joig.12.2.186-198

would be a difficult task, even for a highly sophisticated biological brain. Humans address this challenge by breaking the visual field into smaller segments [3]. In visual scene analysis, visual attention mechanisms facilitate the serialization of processing by breaking up the scene into smaller regions. Visual saliency plays a critical role in this process, as pixels, objects, or persons with high saliency are likelier to capture our attention than their neighbours. Detecting relevant information from cluttered visual scenes whilst filtering out irrelevant information is known as visual attention. It suggests that visual attention requires at least the following fundamental elements in [3, 4]: Visual attention involves several processes, including selecting the region of interest within the visual field, identifying important feature characteristics and values, regulating information flow across the network of neurons in the visual system, and gradually shifting attention to other locations over time.

A general-purpose vision is only possible with attention, which is continuous and automatic. Visual attention is derived from two distinct sources: (1) bottom-up, which is pre-attentive and based on the saliency of the retinal input, and (2) the top-down approach, characterized by its slower pace and reliance on memory and conscious intention, is propelled by the specific objective at hand. For a model to effectively detect saliency, it must fulfill three fundamental requirements that are universally recognized by scholars in the field [5].

The first requirement is accurate detection, where the model should ideally distinguish between true salient areas and false salient areas in the background. The second requirement underscores the significance of high resolution, which is crucial for the precise recognition of salient objects and the maintenance of the original image details. Finally, the model should have computational efficiency, allowing it to quickly identify salient areas, given that these areas are the starting points for many complex processes. There are various applications of bottom-up methods for salient object detection, and saliency detection itself is becoming increasingly popular as a useful tool in the fields of computer vision and artificial intelligence [6–8]. This tool significantly simplifies the intricacies involved in image analysis and speeds up processing durations.

Saliency has found various applications across different domains, such as image segmentation [9–12] object recognition and detection [13–15], anomaly detection [16, 17], image retrieval [18, 19], image compression [20], object classification [21], object tracking [22, 23], image retargeting and summarization [24, 25], alpha matting [25], target detection [26], video object segmentation [27], video summarization [28], Image and video compression [29], Medical image analysis [30], Virtual reality and augmented reality [31], Human-Computer Interaction (HCI) [32] and user perception of digital video content [33]. Integrating various methods has the potential to enhance the strength of object detection. Certain methods may excel in specific aspects, such as capturing global context, handling local details, or considering spatial relationships. A more robust and comprehensive detection can be achieved by effectively exploiting the individual strengths of these techniques. With this motivation, we merged the unique features of popular methods, resulting in a more comprehensive approach that can effectively handle a wide range of image features.

The rest of this document is outlined as follows. The related work is given as part of Section II, the description of popular datasets used for SOD is given under Section III, and the proposed method of saliency is introduced in detail under Section IV. Saliency assignment and refinement of spatial data are covered under Section V. The combined saliency map with enhancement is given in Section VI. Details of evaluation metrics are part of Section VII. A comparative study of qualitative and quantitative results is available in Section VIII. The quintessence of this work has been summarized in Section IX.

II. Related Work

In the realm of computer vision, saliency detection has witnessed noteworthy advancements in recent times. There are three general categories of saliency estimation methods: biological, purely mathematical, and a fusion of both. These methods typically use a combination of pixel intensities, colours, and orientations to identify salient regions in an image and understand their relationship with the surrounding areas.

Contrast is the degree of differentiation between two or further pixels or regions in an image. Saliency values can be computed based on the contrast by measuring the distance between two features. When using local-contrast methods, saliency scores are typically higher at the edges of salient objects, as Jian *et al.* [34] demonstrated in their saliency detection work. This approach emphasizes the entire salient object to enhance its saliency. Local contrast-based: The detection of salient regions in an image using the local contrast-based approach entails determining the uncommonness of image features within a limited spatial extent. Koch and Ullman's [35] biologically plausible architecture formed the basis for previous method [36].

The Difference of Gaussians method is used for the calculation of centre-surround contrast. Frintrop *et al.* [37] created a technique that utilizes the Itti's approach but computes centre-surround disparities with square filters and integrated images to speed up calculation. These are purely computational methods [13, 38, 39] and do not adhere to biological vision principles. Saliency was estimated using centre-surround feature distances as described in [13, 38]. To estimate saliency, a heuristic saliency measure is applied to the histogram thresholding of feature maps obtained by Hu *et al.* [39]. Models of the biological world and simulations of the computational world make up two different categories of methods. In Ref. [40], the creation of maps is illustrated with the help of Itti's state-of-the-art, but the normalization is performed via a graph-based approach.

As a biologically plausible saliency detection model, the maximization of information using a computational model is another method [41]. Jiang *et al.* [42] achieved salient object segmentation using boundary and super pixel techniques. According to Perazzi *et al.* [43], an image's salient region can be determined by its uniqueness and space distribution. In Ref. [44], contrast features were computed across different scales of an image using a hierarchical model, and fused into one map with a graphical model. In global-contrast-based models, salient regions are detected based on the colour contrast over the entire image, enabling an object to be separated from its surroundings. These models can identify prominent parts of an image in a uniform and operationally simple manner. The low-level features (colour and brightness) are used. Achanta *et al.* [45] described a frequency-tuned strategy for detecting centre-surround contrast employing colour and brightness in the frequency domain as characteristics. With a low-rank matrix and sparse noise, Shen and Wu [46] decompose an image into two factors, where the first one indicates regions that fall as background, and the next shows the salient regions.

Cheng *et al.* [47] proposed an abstract representation method based on a Gaussian Mixture Model (GMM) that simultaneously calculates global contrast and spatial coherence differences to detect salient regions at a perceptional homogeneous level. Using light fields, Li *et al.* [48] suggested a method to tackle complex saliency detection challenges, such as comparable foreground and background in an image. An energy-efficient framework for detecting salient regions was enhanced by a method proposed by Wang *et al.* [49], which estimated segmentation with an auto-context model. With graph-based manifold ranking based on affinity matrices, Yang *et al.* [50] ranked the similarity between images and foreground and background cues and successfully detected saliency between the two. Using an unsupervised approach, according to Shiva *et al.* [51], unlabeled images identify patches the most likely to contain salient objects. The objects are located by sampling the regions in the saliency maps. Using a set of background templates as the basis for reconstruction, Li *et al.* [52] propose a saliency measure using dense and sparse representation errors of each image region, and they create a saliency map by integrating the multiscale reconstruction errors. As deep learning advanced, many researchers began to enhance neural networks for saliency detection. Kuen *et al.* [53] proposed a recursive attention neural network that used spatial transformation and recursive network components to recognize salient objects. It can improve the saliency findings of the sub-regions and handle context-aware information in that framework. Cholakkal *et al.* [54] suggested a top-down technique for saliency detection that included linked image classification blocks and a class-aware sparse coding scheme. Murabito *et al.* [55] suggested a top-down saliency map generation using a deep architecture guided by object categorization.

Saliency detection using supervised learning: Saliency detection methods have made use of high-level characteristics through the utilization of supervised machine learning techniques. These methods create regional descriptors by extracting complex image features and predicting saliency scores at the regional level using a classifier or regressor [56]. Kim *et al.* [57] suggested a saliency detection method based on learning, which estimates global saliency by utilizing high-dimensional colour transformation and local contrast through regression. Srivatsa and Babu [58] estimate the regions that fall foreground in an RGB image, which utilizes objectness proposals to obtain smooth and correct saliency maps. In terms of saliency maps, deep learning models based on Convolutional Neural Networks (CNNs) acquire robust features and produce better quality saliency maps [59–62].

Using a deep neural network trained on multiscale features from CNNs, salient regions can be detected with the help of multiscale features [59]. Additionally, deep learning-based mechanisms combined with global and local context cues result in better salient object detection [60]. Object saliency is evaluated at the pixel level and segment-wise for deep contrast network detection [61]. Full Convolutional Networks (FCNs) are used to identify human gaze saliency [62]. Cao *et al.* [63] suggested an improved model based on the You Only Look Once V3 (YOLO V3) Algorithm for object recognition in remote sensing photos in 2020. Recently, Jian *et al.* [64] designed a model to detect a salient region based on location and background cues. A practical method based on sparse background features and spatial position prior to attractive objects is proposed in [65]. An unsupervised method, Topo-Prior-Guided saliency detection System (TOPS), is proposed by Peng *et al.* [66]. Yu *et al.* [67] proposed a local coherence loss to propagate the labels to unlabeled regions based on image features and pixel distance to predict integral salient regions with complete object structures; also, designed a saliency structure consistency loss as a consistent mechanism to ensure consistent saliency maps are predicted with different scales of the same image as input, which could be viewed as a regularization technique to enhance the model generalization ability.

A weakly supervised salient object detection method using point supervision is proposed in [68]. An additional supervision mechanism, a self-Supervised Equivariant Attention Mechanism (SEAM), is proposed in [69]. Using diverse weak supervision sources, a unified framework is proposed for training saliency detection models [70]. The framework in [71] generates the learning curriculum and pseudo ground truth for supervising the training of deep salient object detectors based on combining an intra-image fusion stream and an inter-image fusion stream. A simple GateNet is proposed in [72] to solve interference and disparity between encoder blocks. Qin *et al.* [73] use a predict-refine architecture, BASNet, and a hybrid loss to detect boundary-aware salient objects. U-Net [74], a representative example of a fully convolutional neural network, has achieved remarkable success in medical image segmentation. Its effectiveness in suppressing background noise is achieved through a two-step process: jointly employing an encoder and decoder to process image data, and then integrating the information using skip connections.

III. Datasets

The development of innovative approaches for identifying salient features has resulted in the development of new sets of data for the purpose of evaluating various cutting-edge techniques. A brief description of the different datasets used in this paper for result analysis is provided below.

A. MSRA-B

MSRA-B [75] is a dataset for the detection of salient objects. More than 5,000 images are included in this dataset. The majority of the images in the dataset primarily consist of a single dominant entity. Natural scenes, animals, indoor and outdoor images are included in the collection. Sample images of MSRA-B with our results are depicted in Fig 1.

B. ECSSD

The results are also evaluated using a popular and challenging data set, the ECSSD [44]. It has been designed to enhance image segmentation and research in complex scene saliency. These images are paired with corresponding ground-truth masks. Several images in the ECSSD dataset include complex scenes, varied textures, and low-contrast colours. A total of 1,000 natural images were used in the analysis. Sample images of ECSSD with our results are depicted in Fig. 2.

C. DUT-OMRON

The dataset has hand-picked 5,168 images from over 140,000 nature photographs. These photographs have one or more prominent items and backgrounds that are relatively complicated. Our proposed method provides good result even for such type of challenging images. A sample images of DUT-OMRON [50] with our results are shown in Fig. 3.

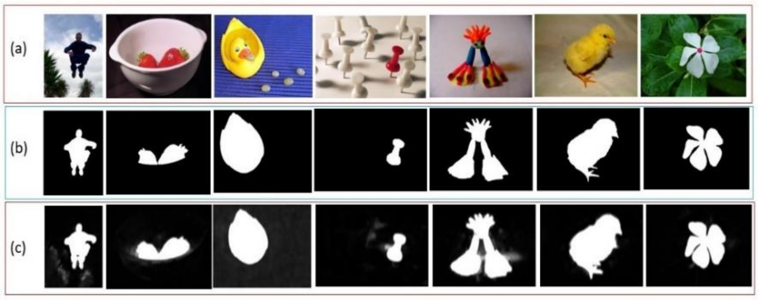

Fig. 1. MSRA dataset: (a) Input images (b) Ground truth (c) Our results.

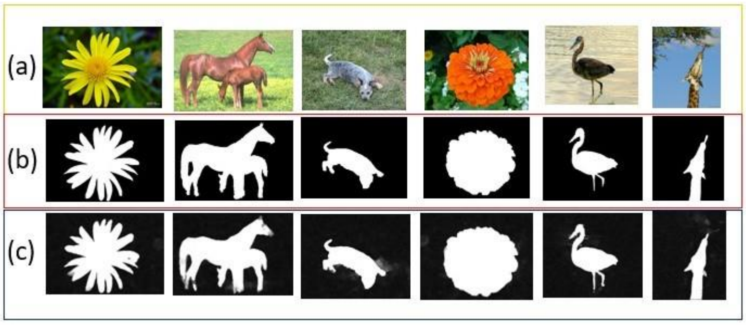

Fig. 2. ECSSD dataset: (a) Input images (b) Ground truth (c) Our results.

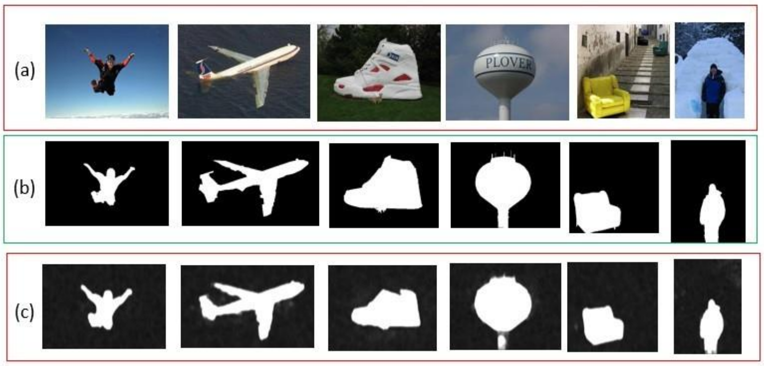

Fig. 3. DUT-OMRON: (a) Input images (b) Groundtruth (c) Final saliency map of our method.

IV. Proposed Method for Saliency Detection

The proposed approach for salient object detection combines several computer vision techniques and machine learning algorithms. Our method consists of several steps, including preprocessing the input image and applying various edge filters such as Sobel, Prewitt, Scharr, Gaussian, and Gabor to extract relevant features. We then use a feature selection algorithm to select the unique features and apply the Random Forest classifier to predict the initial saliency map. Ultimately, the saliency map is subjected to a thresholding technique to achieve the initial salient object. In the subsequent stage, local and global contrast and prior background are computed to obtain the prior calculation.

The SOD scheme is depicted in Fig. 4.

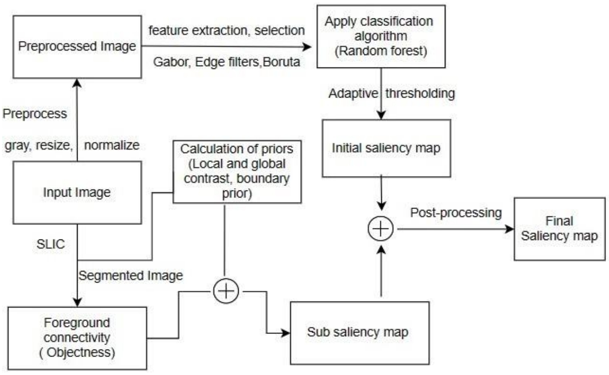
Preprocessed Image
feature extraction, selection
Gabor, Edge filters,Boruta
Apply classification algorithm (Random forest)
Adaptive thresholding
Preprocess
gray, resize, normalize
Calculation of priors (Local and global contrast, boundary prior)
Initial saliency map
Input Image
Post-processing
Final Saliency map
SLIC
Segmented Image
Foreground connectivity (Objectness)
Sub saliency map

Fig. 4. Proposed method for saliency detection.

The saliency values of an image can be obtained through the application of the suggested method, which encompasses the subsequent procedures.

Each step of the Algorithm 1 is explained below.

Algorithm 1: Proposed method

```
1)  Input: RGB Image
2)  Output: Final saliency map
3)  S = RF (Boruta (Gabor(edge((I))))
4)  2-D superpixel over segmentation of images using SLIC
5)  Calculate sub-saliency map using foreground connectivity FG(S)
        $FG(S) = \frac{\sum_{k=1}^{N} d(P, P_k).\delta(P_k)}{\sum_{k=1}^{N} d(P, P_k).(1-\delta(P_k))}$
6)  Calculate fusion map by the calculation of priors Fmap
        $F_{map} = \frac{1}{Z}(B_{nd}Con(p) \times D_{L_i} \times D_{G_i})$
7)  Combine the initial and fusion map
        $Salmap_{final} = \frac{1}{z}(w_1 p(w_2 FG_P) + w_3 p(w_4 F_{map}))$
8)  Post-processing:
        $Spp(i, j) = 1$ if $S(i, j) \geq TH$ and $Pmax(i, j) > Pbg(i, j)$
        $= 0$ otherwise
9)  Final saliency map
10) End
```

A. Creating the Initial Saliency Map with the Random Forest Method

The steps involved in the proposed algorithm are:

1. Preprocessing steps

a) Resize the input image to a fixed size.

b) Normalized the pixel values to the range [0, 1].

c) Applied edge filters such as Sobel, Roberts, Canny, Gaussian with different sigma values, Prewitt, Scharr to generate edge-based features.

d) Applied Gabor filter orientations in different directions and scales to generate texture-based features.

2. Feature selection: Use Boruta algorithm to select the most relevant features.

3. Training: Train a random forest classifier using the selected features.

4. Saliency map generation (Prediction): Generate an intermediate saliency map by applying the trained random forest classifier to the input image.

The sequential application of this step is:

S_{rf} = Thresholding (RF (Boruta (Gabor(edge((I))))))

S stands for the intermediate saliency map. I will be the preprocessed image, the edge is the edge filter function, Gabor is the Gabor filter bank function, Boruta is the feature selection algorithm, and RF is the random forest classifier function. The parentheses denote function composition or sequential application. Texture features capture the visual patterns and structures in the image, which are essential for discriminating between salient and non-salient regions. Contrast features accentuate variations in color or intensity among adjacent pixels, facilitating the differentiation of prominent objects from their background. On the other hand, edge features capture the sudden changes in pixel values, playing a crucial role in defining the boundaries of salient objects. Gabor filters, often used in image processing, are able to catch the frequency and orientation details of an image. This makes them perfect for tasks like edge detection, texture analysis, and object recognition. These filters are designed to mimic the response regions of primary cells in the mammalian visual cortex, which are uniquely capable of detecting oriented edges and light bars. The creation of these filters involves modulating a Gaussian function with a sinusoidal wave, resulting in filters that are sensitive to edges and textures of varying orientations and scales. When applied to an image, Gabor filters can extract key features for saliency detection, such as edges, ridges, and textures. The Boruta algorithm is then employed to identify the most significant and distinguishing features for detecting prominent objects. These generated features undergo a feature selection process to find the most relevant and discriminative features for salient object detection. The selected features confirmed by the Boruta algorithm, are used to train a random forest classifier. This classifier generates a saliency map that highlights the regions in the input image most likely to contain salient objects. The model leverages its learned associations between Gabor and other edge filters, such as Roberts, Sobel, Canny, Gaussian (with sigma = 3, 7), median filter (with sigma = 3), and Prewitt and Scharr filters. The random forest algorithm, a well-known technique in machine learning, is capable of handling numerous features and non-linear relationships among them, making it effective in analyzing complex images. We employed random forests to understand the relationship between low-level visual features, such as color, texture, and edge information, and high-level semantic concepts, like objects and scenes. For training, a labeled dataset is used, with the extracted features used as input and the ground truth saliency maps as output. The algorithm learns to predict saliency maps from the input features. The hyperparameters are set with a tree count of 200 and a maximum depth of 10 for each tree. Once the random forest is trained, we use it to predict saliency maps for new images. The test images undergo identical feature extraction procedures, and the random forest algorithm utilizes the acquired knowledge to make predictions.

The continuous-valued map, indicating the likelihood of each pixel being salient, is converted into a binary mask that segments the salient object from the background. The threshold value of 0.5 is utilized for achieving the optimal detection of the salient object.

B. Foreground Connectivity

The use of superpixels to detect salient objects has been demonstrated in recent works [43, 46, 76]. We chose the Simple Linear Iterative Clustering (SLIC) algorithm approach for superpixel segmentation since it is quick, precise, efficient and low computation cost. The segmentation on image I to form superpixels X = [X_1, ..., X_K]. The salient object boundary of an image should be well-preserved for receiving structure information. To separate the original image into K small superpixels, we use the SLIC [77]. Pixels with similar values are grouped using this algorithm. It is possible to reduce the complexity of image processing operations, such as segmentation, by using these regions. SLIC also makes the whole algorithm more efficient. Foreground extraction aims to separate foreground (desirable) information from background (undesirable) information in an image or video feed. To handle the issue of efficiently extracting a foreground object in a complex environment whose background cannot be easily subtracted. The paper is implemented with a saliency measure called foreground connectivity, which calculates the foreground connectivity of a superpixel S as illustrated in [58]. The Objectness map is created to collect superpixels that contain the salient object. We apply the foreground connectivity measure to allocate weights to the superpixels for the foreground. To generate an objectness map, we adapted the approach in [58], the objectness map is generated once the object proposals have been gathered. We can determine how likely a window is to contain an object by evaluating its objectness score. A pixel-wise objectness score (PixObj) indicates whether a pixel is part of an object. The value of the PixelObj is calculated as

$$PixObj(p) = \sum_{i=1}^{k} r_i G_i\ (y, z) \qquad (1)$$

where r_1, r_2, r_3, ..., r_n are the objectness scores of the proposals that include pixel p are denoted as G_i. Where G_i is a Gaussian window with the same dimensions as the given proposal. The relative x and y coordinates of pixel *p* with respect to the given proposal are represented by *y* and *z*, respectively. By summing up the pixel-wise object probability in a region of superpixel, we can construct the objectness map for that superpixel region.

$$Objectness(R) = \sum_{i \in R} PixObj\ (pi) \qquad (2)$$

An instance of pi is a pixel in super pixel region *R*.

The saliency measure referred to as “foreground connectivity” determines saliency values by taking into account the connectivity of superpixels to the approximated foreground. Using super-pixels as nodes, we build a graph. Each edge of an image corresponds to the Euclidean distance among the adjacent superpixels. A Superpixel P’s foreground connectivity is given by

$$FG(P) = \frac{\sum_{k=1}^{N} d(P, P_k) \cdot \delta(P_k)}{\sum_{k=1}^{N} d(P, P_k) \cdot (1 - \delta(P_k))} \qquad (3)$$

where d (P, P_k) denotes the shortest distance between P to P_k and δ(.) is 1 for a superpixel if it is estimated as foreground by the objectness map and the total number of superpixels is N. The reciprocal of FG (i.e., lower foreground weights) is used as the foreground weight w^{fg} and lower numerator value when the superpixel exhibits a greater resemblance to the estimated foreground. The initial saliency map will be estimated from the concatenated features mentioned previously.

C. Calculation of Boundary Prior

This paper is inspired by Zhu *et al.* [77], Natural images have different spatial layouts for object regions and background areas, i.e., A region with an object is much less connected to the image boundary than a region with a background. To compute how strongly a region R is associated with the boundaries of an image, known as boundary connectivity given by

$$BndCon(R) = \frac{|\{p|p \in R, p \in Bnd\}|}{\sqrt{|\{p|p \in R\}|}} \qquad (4)$$

Image boundary patches are identified by B_{nd}, whereas image patches are identified by *p*. An intuitive interpretation of the square root is that it represents the ratio between the perimeter on the boundary of a region and the perimeter on the overall boundary. In the next step, all the superpixels (*p*, *q*) are connected to create an undirected weighted graph d_{app} (*p*, *q*) reflecting the Euclidean distance between their average colors in CIE-Lab. The graph defines the geodesic distance d_{geo} (*p*, *q*) between any two superpixels as the aggregated edge weights along the shortest path between them.

$$d_{geo}(p, q) = min \sum_{i=1}^{n-1} d_{app}\ (p_i, p_{i+1}) \qquad (5)$$

The spanning area of each superpixel p as

$$Area(p) = \sum_{i=1}^{N} S\ (p, p_i) \qquad (6)$$

where the number of superpixels is *N*. Using Eq. (7), one calculates a soft area for the region where *p* lies. In the summation, S (*p*, p_i) characterizes the amount that pi contributes to the area of *p*, and is in [0, 1]. The computation of the length along the boundary is performed as:

$$Len_{bnd}(p) = \sum_{i=1}^{N} S\,(p, p_i)\;\delta\;(p_i \in B_{nd}) \qquad (7)$$

If the superpixels are on the image boundary then δ(.) is 1 otherwise 0. The boundary connectivity is given by

$$B_{nd}Con(p) = \frac{Len_{bnd}(p)}{\sqrt{Area(p)}} \qquad (8)$$

Johnson's approach [78] is used to compute the shortest pathways between all superpixel pairs efficiently.

D. Global and Local Contrast Prior

The global contrast of the ith superpixel is given by

$$D_{G_i} = \sum_{j=1}^{N} d(g_i, g_j) \qquad (9)$$

where d (g_i, g_j) denotes the Euclidean distance between the i^{th} and j^{th} superpixel's color value g_i and g_j.

Local contrast of color features is defined as

$$D_{L_i} = \sum_{j=1}^{N} w_{i,j}^{p}\; d(g_i, g_j) \qquad (10)$$

where,

$$w_{i,j}^{p} = \frac{1}{R_i}\exp\,(-\frac{1}{2\sigma_p^2} \parallel p_i - p_j \parallel_2^2 \qquad (11)$$

where $p_i \in [0, 1] \times [0, 1]$ denotes the normalized position of the i^{th} superpixel and R_i is the normalization term. The weight function in Eq. (11) is commonly employed in various applications, we adapt this function to assign more weight to neighboring superpixels. σ_p^2 is set to the value 0.25 [57].

V. Saliency Assignment of Our Proposed Method

Cheng *et al.* [79] suggested an approach to combine the two saliency maps. The method used is to multiply the two maps pixel by pixel here a fusion map is created by combining all the priors above as shown below:

$$F_{map} = \frac{1}{Z}(B_{nd}Con(p) \times D_{L_i} \times D_{G_i}) \qquad (12)$$

In general, the saliency of several cues is combined using a heuristic method involving weighted summation or multiplication. Instead, our foreground weights and background weights are merged using an existing optimization framework adopted by [58] and is defined as:

$$\sum_{i=1}^{N} w_i^{fg}\,(r_i - 1)^2 + \sum_{i=1}^{N} w_i^{bg}(r_i)^2 + \sum_{i,j} w_{ij}(r_i - r_j)^2 \qquad (13)$$

where r_i denotes the final saliency value assigned to p_i after minimizing the cost, w_i^{fg} denotes foreground weights, w_i^{bg} denotes background weights associated with superpixel p_i. High w_i^{fg} encourages p_i takes close to (0, 1) for high foreground and background, respectively. w_{ij} is the smoothness coefficient. Parameter settings are the same initial map and vice versa for pixels near the definite background. The spatial map of saliency is calculated using.

$$S_P(X_j) = \exp\,(-k\frac{A}{B}) \qquad (14)$$

where $A = min_{i\in F}(d(q_i, q_j))$ and $B = min_{i\in B}(d(q_i, q_j))$ are the Euclidean distance which is the minimum value that is calculated from the i^{th} to a particular foreground pixel and a distinct background pixel. The value of the parameter k is assigned to 0.5.

VI. Combined Map of Saliency and Enhancement

The initial map of saliency from Section IV.B and fusion map created from Section V are used to get the combined saliency map. We used the approach for combining saliency maps used by Kim *et al.* [57].

$$Salmap_{final} = \frac{1}{z}(Srf + w_1 p(w_2 FG_P) + w_3 p(w_4 F_{map})) \qquad (15)$$

$Salmap_{final}$ is enhanced using post-processing techniques: Non-Maximum Suppression (NMS) and Adaptive Histogram Equalization (AHE) are the post-processing steps applied to the combined saliency map. Let M be a saliency map of an input image, and $M(j, k)$ be the saliency value at the pixel position (j, k). NMS is performed by scanning the saliency map M and keeping only the maximum saliency value in a local neighbourhood. Let w and h be the width and height of the neighbourhood window, respectively. Then, the output saliency map M' after NMS can be computed as:

$$M'(j,k) = 0, if M(j,k) < max(M(j-w : j+w, k-h : k+h))$$
$$M'(i,j) = M(i,j) \text{ otherwise}$$

Adaptive Histogram Equalization is applied to enhance the contrast of the saliency map. The mathematical formula for AHE is represented as follows:

$$S_{PP}(j,k) = 1\; if\; S(j,k) \geq\; TH$$
$$and\; P_{max}(j,k) > P_{bg}(j,k) = 0\; otherwise.$$

let M' be the saliency map after NMS and H (j) be the cumulative histogram of pixel intensities up to level j. The output enhanced saliency map M'' can be computed as:

$$M''(j,k) = H(M'(j,k)) \times 255$$

where $H(M'(j, k))$ is the normalized cumulative histogram of the pixel intensity at position (j, k) in the saliency map M', and 255 is the maximum pixel value.
where S_{PP} is the final saliency map after post-processing, S is the initial saliency map obtained from the Random Forest classifier, TH is the threshold obtained from adaptive histogram equalization, P_{max} is the maximum value of the probability distribution of each superpixel, and P_{bg} is the probability of the superpixel belonging to the background. The values of P_{max} and P_{bg} are calculated using local and global contrast prior, and background prior. Finally, non-maximum suppression is applied to remove redundant salient regions, which gives the final saliency map.

VII. Details of Evaluation Metrics

The proposed method is evaluated with famous object-level evaluation metrics, i.e., Precision-Recall (PR), F-measure [80], Mean Absolute Error (MAE) [43], and weighted F_β measure (Fbw) [81].

A. Details of Metrics

Recall and Precision: Precision, also referred to as the positive predictive value, is determined by dividing the count of groundtruth pixels that are correctly identified as a salient region by the overall count of pixels that are classified as a salient region. The Recall value, or sensitivity, is directly related to the number of salient regions recovered from the ground truth. The calculation of PR entails the utilization of the binarized mask of the salient object and the ground-truth. Both the binarized mask of the salient object and the ground-truth are employed in order to determine the precision and recall.

$$Precision\ = \frac{TP}{TP+FP}, Recall = \frac{TP}{TP+FN}$$

A range of thresholds, ranging from 0 to 255, is employed to convert the prediction into binary format. Each individual threshold produces a set of precision and recall values, which are utilized to create a PR curve that characterizes the performance of the model. The ratio of accurately classified prominent pixels to all prominent pixels in the ground truth map is known as precision. As the model's performance improves, the PR curve approaches the upper left corner. The precision and recall rates are compared in the first evaluation.

F-measure [80]: The F-measure rates for the binarized saliency map are computed with a threshold range of [0, 255] and are given by

$$F_\beta = \frac{(1+\beta^2)\ Precision \times Recall}{\beta^2 \times Precision + Recall} \tag{16}$$

like the existing methods [44–46, 82], the use of β^2 is assigned with 0.3 to give more importance to precision.

Mean Absolute Error (MAE) [43]: The Precision-Recall curve does not include the fraction of pixels which have been correctly classified as non-salient. The presence of pixels mistakenly labelled as salient leads the saliency map to perform worse, despite being smooth and having greater values allocated to salient pixels. Using Mean Absolute Error (MAE) as suggested by Perazzi *et al.* [43], we can overcome the limitation of using precision and recall. We conduct an examination of the Mean Absolute Error (MAE) between the continuous saliency map M and the binary ground truth GT in order to ensure a more equitable comparison that takes into account these variables. The lower the MAE score, the model is close to ground truth, and better the performance.

$$\begin{gathered} MeanAbsoluteError \\ = \frac{1}{W \times H} \sum_{a=1}^{H} \sum_{b=1}^{W} |M(a,b) - G(a,b)| \end{gathered}$$

All the above metrics are computed for the proposed method and alternate state-of-the-art algorithms are shown under quantitative section tables. We have used the code for evaluation measures given in [5].

VIII. State-of-the-Art Comparisons

Using three distinct datasets, we have conducted a comparative analysis between our model and a total of 17 distinct state-of-the-art models. These datasets include MSRA-B [75], Extended Complex Scene Saliency Dataset (ECSSDs) [44], and DUT-OMRON [50]. Based on the evaluation results, our method is highly effective and a promising method for salient object detection.

A method with high precision but poor recall suggests that it might work better for gaze-tracking experiments, but not for salient object segmentation. So, to maintain good values for recall and precision, the thresholded value chosen is 0.5. The saliency maps of our approach outperform alternate, state-of-the-art algorithm, the results of both quantitative and qualitative results are given under subsection A and B.

A. Qualitative Results

The performance of the proposed algorithm was evaluated against multiple state-of-the-art models, which includes models such as IT [36], GB [38], SR [39], AC [13], IG [45], MZ [38], MC [83], CHC [84], DSP [85], HDCT [57], BGFG [86], DSR [52], CNS [87], FCB [88], DCLC [89], DRFI [76] DGL [90] and the approach followed by the method is mentioned in Table I.

When both the foreground and background are in similar colour, the method is robust with, the proposed approach with the addition of as a pre-processing step [91]. The proposed technique demonstrates a consistent and precise detection of salient regions across various image classes, surpassing many contemporary methods. The findings suggest that although the majority of existing methodologies excel at effectively handling relatively straightforward images containing only one or uniform objects, they encounter difficulties when it comes to image categories that feature intricate backgrounds, inadequate contrast, or multiple objects. Approaches that rely on boundary or background priors, such as BGFG, DGL, DSR, and MC, are incapable of detecting or uniformly emphasizing objects that come into contact with the image boundary, as evidenced by the accompanying illustration in the Figs. 5 and 6. The results indicate that while most existing techniques excel at processing relatively simple images with single or homogeneous objects, they struggle with image categories characterized by complex backgrounds, low contrast, or multiple objects. Techniques that employ boundary or background priors, such as BGFG, DGL, DSR, and MC fail to detect or uniformly emphasize objects that come into contact with the image boundary, as evidenced in the figure. Contrarily, our approach has effectively identified and consistently emphasized the salient objects that come into contact with the image boundaries.

TABLE I. METHOD AND THE APPROACH FOLLOWED

S:NO	Method	Approach followed
1	CHC	Contrast ratio, spatial feature, colour contrast, and central prior
2	DSP	Background seeds, distribution prior, manifold ranking
3	HDCT	Contrast features, location, histogram, texture, shape features, learning-based approach, global and local colour
4	BGFG	Background and foreground prior
5	DSR	Background prior
6	CNS	Surroundedness and global colour contrast cues
7	FCB	Foreground and background cues, centre prior.
8	DCLC	centre prior, diffusion-based, manifold ranking, compactness local contrast
9	DRFI	Colour and texture contrast features, background features
10	MC	Boundary prior, graph-based, Markov random walk
11	DSP	Manifold ranking, Chi-square distance

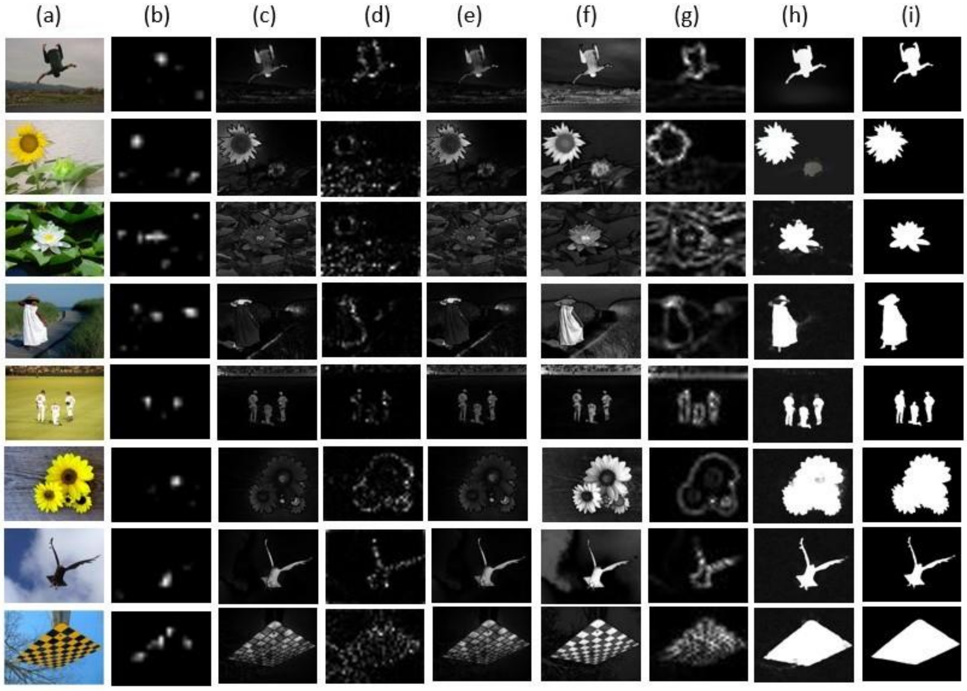

Fig. 5. Qualitative comparison of the proposed method with other state of the art methods on MSRA-B dataset. (a) input image, (b) IT [36], (c) GB [38], (d) SR [39], (e) AC [13], (f) IG [45], (g) MZ [38], (h) Ours, (i) ground truth.

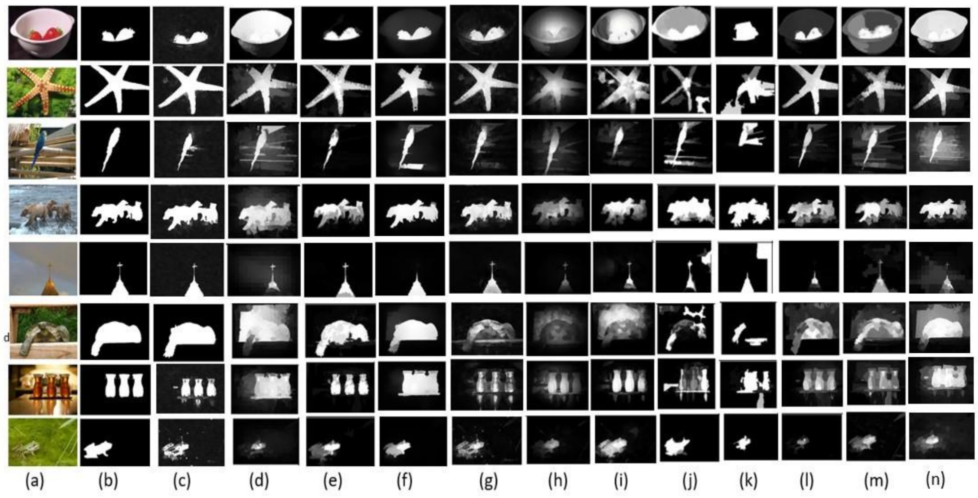

Fig. 6. Qualitative comparison of the proposed method and other state of the art methods in some challenging cases. (a) Input image (b) Groundtruth (c) ours (d) MC [83], (e) CHC [84], (f) DSP [85], (g) HDCT [57], (h) BGFG [86], (i) DSR [52], (j) CNS [87], (k) FCB [88], (l) DCLC [89], (m) DRFI [76] (n) DGL [90].

B. *Quantitative Results*

The proposed method's saliency map is compared with the state-of-the-art in Precision, Recall, F-measure and MAE. Table I gives the approach followed by the methods used for comparison. The quantitative results, compared with various state-of-the-art methods, are presented in Table II. The experimental results presented in this study offer a fair comparison, as they are derived from the implementation provided by the authors' publicly released code with its default parameters.

C. *Computation and Complexity of the Proposed Method*

For many applications, salient object detection is a pre-processing step, as it efficiently detects regions of interest and reduces the computational complexity of image analysis. For real-time applications, where computational limitations are a significant concern, the algorithm must rapidly and accurately identify the most salient regions. However, the computational complexity of many methods, including deep-learning-based approaches, can be a limiting factor that hinders their performance in real-time applications. This study utilizes computational analysis during runtime to illustrate experimentally the efficacy of the proposed method. The proposed method is run on MATLAB R2021b using an i7-10870H CPU @2.20GHz with 16 GB RAM. Fig. 7 depicts the computational time of the methods.

TABLE II. COMPARING THE THRESHOLDED SALIENCY MAP'S PRECISION, RECALL, AND F-MEASURE RATE WITH CUTTING- EDGE ALGORITHMS ACROSS THREE BENCHMARK DATASETS: ECSSD, MSRA-B, AND DUT-OMRON, THE TOP RESULTS OF EACH METRIC ARE SHOWN IN BOLD

Dataset	Metric	Ours	MC	CHC	DSP	HDCT	BGFG	DSR	CNS	FCB	DCLC	DRFI	DGL
ECSSD	Precision ↑	0.819	0.768	**0.853**	0.82	0.767	0.723	0.753	0.708	0.721	0.769	0.794	0.785
	Recall ↑	**0.862**	0.652	0.635	0.77	0.640	0.606	0.647	0.600	0.515	0.636	0.698	0.655
	F-measure ↑	**0.834**	0.738	0.790	0.80	0.733	0.692	0.726	0.680	0.660	0.734	0.769	0.750
	MAE ↓	**0.14**	0.202	0.163	0.143	0.15	0.208	0.171	0.166	0.173	0.182	0.170	0.191
MSRA-B	Precision ↑	0.913	0.84	**0.94**	0.88	0.81	0.854	0.86	0.819	0.93	0.91	0.86	0.906
	Recall ↑	**0.923**	0.79	0.88	0.83	0.80	0.850	0.70	0.903	0.615	0.915	0.81	0.909
	F-measure ↑	0.893	0.79	**0.93**	0.842	0.78	0.853	0.79	0.837	0.85	0.911	0.83	0.906
	MAE ↓	0.096	0.098	0.067	0.08	0.099	0.112	0.762	**0.058**	0.140	0.063	0.15	0.063
DUT-OMRON	Precision ↑	0.893	0.819	**0.931**	0.610	0.801	0.771	0.827	0.768	0.891	0.842	0.856	0.877
	Recall ↑	**0.849**	0.774	0.760	0.761	0.791	0.696	0.776	0.751	0.804	0.791	0.827	0.812
	F-measure ↑	0.816	0.809	0.885	0.630	0.814	0.726	0.812	0.763	**0.887**	0.834	0.847	0.861
	MAE ↓	**0.113**	0.168	0.126	0.143	0.162	0.179	0.127	0.137	0.135	0.133	0.138	0.136

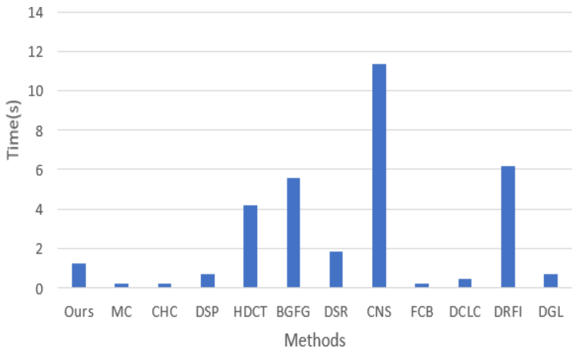

Fig. 7. Average run time comparison.

IX. CONCLUSION

This work proposes a method for estimating salient objects in an image via classification, foreground connectivity and priors. As the proposed method uses the random forest algorithm for salient object detection, it can handle numerous features, including texture-based and edge-based features, even providing a reliable and efficient way to select the most relevant and discriminative features for salient object detection. Gabor filters can extract texture information from an image, which can be helpful in identifying salient objects that have distinctive texture patterns. On the other hand, combining edge filters to detect high-contrast boundaries in an image indicates salient object edges. These features are used as inputs to a classification algorithm to predict the saliency of each pixel in the image. The technique of foreground connectivity can be used to refine the saliency map and improve the accuracy of salient object detection by considering the spatial relationships between pixels in the foreground. By incorporating local and global contrast measures, the algorithm can identify regions in the images with high contrast, which are likely to be salient objects, as the salient objects tend to have distinct edges and texture patterns that result in high contrast compared to their surroundings, so it aids in improving the accuracy of the saliency map. Findings on three benchmark datasets demonstrate that our approach outperforms the current state-of-the- art regarding precision, recall, f-measure, and MAE. Although our method does not precisely replicate human vision, it still generates output more closely aligns with ground truth than other approaches. However, our method's computational time is slightly longer than that of specific other methods; its quality still needs to be improved. In our future work, we plan to incorporate additional functionality into the saliency detection framework to improve accuracy in challenging background settings and reduce computational time by identifying optimal cues.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Gayathri Dhara: Conceptualization; Methodology; Software; Investigation; Data collection; writing; Ravi Kant Kumar: Methodology, Validation, Review. All authors had approved the final version.

Acknowledgment

We thank SRM University–AP for providing facilities in research lab for completing this research work.

References

[1] K. Ahmed, M. A. Gad, and A. E. Aboutabl, "Performance evaluation of salient object detection techniques," *Multimedia Tools and Applications*, pp. 1–37, 2022.

[2] I. Ullah, M. Jian, S. Hussain, J. Guo, H. Yu, X. Wang, and Y. Yin, "A brief survey of visual saliency detection," *Multimedia Tools and Applications*, vol. 79, pp. 34605–34645, 2020.

[3] J. K. Tsotsos, S. M. Culhane, W. Y. K. Wai, Y. Lai, N. Davis, and F. Nuflo, "Modeling visual attention via selective tuning," *Artificial Intelligence*, vol. 78, no. 1–2, pp. 507–545, 1995.

[4] R. K. Kumar, J. Garain, D. R. Kisku, and G. Sanyal, "Guiding attention of faces through Graph Based Visual Saliency (GBVS)," *Cognitive Neurodynamics*, vol. 13, no. 2, pp. 125–149, 2019.

[5] A. Borji, M.-M. Cheng, Q. Hou, H. Jiang, and J. Li, "Salient object detection: A survey," *Computational Visual Media*, vol. 5, pp. 117–150, 2019.

[6] Y. Wu, T. Jia, Y. Pang, J. Sun, and D. Xue, "Salient object detection via a boundary-guided graph structure," *Journal of Visual Communication and Image Representation*, vol. 75, 103048, 2021.

[7] Z. Wang, G. Xu, Z. Wang, and C. Zhu, "Saliency detection integrating both background and foreground information," *Neurocomputing*, vol. 216, pp. 468–477, 2016.

[8] X. Zhang, Y. Wang, Z. Chen, J. Yan, and D. Wang, "Saliency detection via image sparse representation and color features combination," *Multimedia Tools and Applications*, vol. 79, pp. 23147–23159, 2020.

[9] H. Fan, F. Xie, Y. Li, Z. Jiang, and J. Liu, "Automatic segmentation of dermoscopy images using saliency combined with otsu threshold," *Computers in Biology and Medicine*, vol. 85, pp. 75–85, 2017.

[10] S. Yuheng and Y. Hao, "Image segmentation algorithms overview," arXiv preprint, arXiv:1707.02051, 2017.

[11] O. O. Olugbara, T. B. Taiwo, and D. Heukelman, "Segmentation of melanoma skin lesion using perceptual color difference saliency with morphological analysis," *Mathematical Problems in Engineering*, vol. 2018, pp. 1–19, 2018.

[12] A. Joshi, M. S. Khan, S. Soomro, A. Niaz, B. S. Han, and K. N. Choi, "Sris: Saliency-based region detection and image segmentation of covid-19 infected cases," *IEEE Access*, vol. 8, pp. 190487–190503, 2020.

[13] R. Achanta, F. Estrada, P. Wils, and S. Su, "Salient region detection and segmentation," in *Proc. International Conference on Computer Vision Systems*, 2008, pp. 66–75.

[14] V. John, K. Yoneda, Z. Liu, and S. Mita, "Saliency map generation by the convolutional neural network for real-time traffic light detection using template matching," *IEEE Transactions on Computational Imaging*, vol. 1, no. 3, pp. 159–173, 2015.

[15] H. Li, X. Su, J. Wang, H. Kan, T. Han, Y. Zeng, and X. Chai, "Image processing strategies based on saliency segmentation for object recognition under simulated prosthetic vision," *Artificial Intelligence in Medicine*, vol. 84, pp. 64–78, 2018.

[16] W. Liu, X. Feng, S. Wang, B. Hu, Y. Gan, X. Zhang, and T. Lei, "Random selection-based adaptive saliency-weighted rxd anomaly detection for hyperspectral imagery," *International Journal of Remote Sensing*, vol. 39, no. 8, pp. 2139–2158, 2018.

[17] M. Al-Gabalawy, "Removed: Detecting anomalies within Unmanned Aerial Vehicle (UAV) video based on contextual saliency," *Applied Soft Computing*, vol. 96, 106715, 2020.

[18] Y. H. Tsai, "Hierarchical salient point selection for image retrieval," *Pattern Recognition Letters*, vol. 33, no. 12, pp. 1587– 1593, 2012.

[19] E. Giouvanakis and C. Kotropoulos, "Saliency map driven image retrieval combining the bag-of-words model and PLSA," in *Proc. 2014 19th International Conference on Digital Signal Processing*, IEEE, 2014, pp. 280–285.

[20] C. Zhu, K. Huang, and G. Li, "An innovative saliency guided roi selection model for panoramic images compression," in *Proc. 2018 Data Compression Conference*, IEEE, 2018, pp. 436–436.

[21] N. Li, X. Zhao, Y. Yang, and X. Zou, "Objects classification by learning-based visual saliency model and convolutional neural network," *Computational Intelligence and Neuroscience*, vol. 2016, 2016.

[22] M. Guo, Y. Zhao, C. Zhang, and Z. Chen, "Fast object detection based on selective visual attention," *Neurocomputing*, vol. 144, pp. 184–197, 2014.

[23] L. Gao, B. Liu, P. Fu, M. Xu, and J. Li, "Visual tracking via dynamic saliency discriminative correlation filter," *Applied Intelligence*, vol. 52, no. 6, pp. 5897–5911, 2022.

[24] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1915–1926, 2011.

[25] R. Li, J. Cai, H. Zhang, and T. Wang, "Aggregating complementary boundary contrast with smoothing for salient region detection," *The Visual Computer*, vol. 33, pp. 1155–1167, 2017.

[26] M. Wan, K. Ren, G. Gu, X. Zhang, W. Qian, Q. Chen, and S. Yu, "Infrared small moving target detection via saliency histogram and geometrical invariability," *Applied Sciences*, vol. 7, no. 6, 569, 2017.

[27] G. Lin and W. Fan, "Unsupervised video object segmentation based on mixture models and saliency detection," *Neural Processing Letters*, vol. 51, pp. 657–674, 2020.

[28] S. Marat, M. Guironnet, and D. Pellerin, "Video summarization using a visual attention model," in *Proc. 2007 15th European Signal Processing Conference*, 2007, pp. 1784–1788.

[29] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1304–1318, 2004.

[30] P. K. Sran, S. Gupta, and S. Singh, "Integrating saliency with fuzzy thresholding for brain tumor extraction in MR images," *Journal of Visual Communication and Image Representation*, vol. 74, 102964, 2021.

[31] H. Duan, W. Shen, X. Min, D. Tu, J. Li, and G. Zhai, "Saliency in augmented reality," *in Proc. the 30th ACM International Conference on Multimedia*, 2022, pp. 6549–6558.

[32] P. K. Pook, "Saliency in human-computer interaction," *Assistive Technology and Artificial Intelligence: Applications in Robotics, User Interfaces and Natural Language Processing*, pp. 73–83, 2006.

[33] T. Adeliyi and O. Olugbara, "Detecting salient objects in non-stationary video image sequence for analyzing user perceptions of digital video contents," *Multimedia Tools and Applications*, vol. 78, pp. 31807–31821, 2019.

[34] M. Jian, Q. Qi, J. Dong, X. Sun, Y. Sun, and K.-M. Lam, "Saliency detection using quaternionic distance based weber local descriptor and level priors," *Multimedia Tools and Applications*, vol. 77, no. 11, pp. 14343–14360, 2018.

[35] C. Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry," *in Matters of Intelligence*. 1987, pp. 115–141.

[36] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[37] S. Frintrop, M. Klodt, and E. Rome, "A real-time visual attention system using integral images," in *Proc. International Conference on Computer Vision Systems*, 2007.

[38] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proc. the Eleventh ACM International Conference on Multimedia*, 2003, pp. 374–381.

[39] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. 2007 IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[40] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Advances in Neural Information Processing Systems*, vol. 19, 2006.

[41] N. Bruce and J. Tsotsos, "Attention based on information maximization," *Journal of Vision*, vol. 7, no. 9, pp. 950–950, 2007.

[42] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li, "Automatic salient object segmentation based on context and shape prior." *BMVC*, vol. 6, no. 7, 2011, p. 9.

[43] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. 2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 733–740.

[44] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1155–1162.

[45] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. 2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1597–1604.

[46] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proc. 2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 853–860.

[47] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and N. Crook, "Efficient salient region detection with soft image abstraction," in *Proc. the IEEE International Conference on Computer Vision*, 2013, pp. 1529–1536.

[48] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu, "Saliency detection on light field," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2806–2813.

[49] L. Wang, J. Xue, N. Zheng, and G. Hua, "Automatic salient object extraction with contextual cue," in *Proc. 2011 International Conference on Computer Vision*, 2011, pp. 105–112.

[50] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3166–3173.

[51] P. Siva, C. Russell, T. Xiang, and L. Agapito, "Looking beyond the image: Unsupervised learning for object saliency and detection," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3238–3245.

[52] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *Proc. the IEEE International Conference on Computer Vision*, 2013, pp. 2976–2983.

[53] J. Kuen, Z. Wang, and G. Wang, "Recurrent attentional networks for saliency detection," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3668–3677.

[54] H. Cholakkal, J. Johnson, and D. Rajan, "A classifier-guided approach for top-down salient object detection," *Signal Processing: Image Communication*, vol. 45, pp. 24–40, 2016.

[55] F. Murabito, C. Spampinato, S. Palazzo, D. Giordano, K. Pogorelov, and M. Riegler, "Top-down saliency detection driven by visual classification," *Computer Vision and Image Understanding*, vol. 172, pp. 67–76, 2018.

[56] A. K. Gupta, A. Seal, M. Prasad, and P. Khanna, "Salient object detection techniques in computer vision—A survey," *Entropy*, vol. 22, no. 10, 1174, 2020.

[57] J. Kim, D. Han, Y.-W. Tai, and J. Kim, "Salient region detection via high-dimensional color transform and local spatial support," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 9–23, 2015.

[58] R. S. Srivatsa and R. V. Babu, "Salient object detection via objectness measure," in *Proc. 2015 IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 4481–4485.

[59] G. Li and Y. Yu, "Visual saliency based on multiscale deep features," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5455–5463.

[60] R. Zhao, W. Ouyang, H. Li, and X. Wang, "Saliency detection by multi-context deep learning," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1265–1274.

[61] G. Li and Y. Yu, "Deep contrast learning for salient object detection," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 478–487.

[62] N. D. Bruce, C. Catton, and S. Janjic, "A deeper look at saliency: Feature contrast, semantics, and beyond," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 516–524.

[63] C. Cao, J. Wu, X. Zeng, Z. Feng, T. Wang, X. Yan, Z. Wu, Q. Wu, and Z. Huang, "Research on airplane and ship detection of aerial remote sensing images based on convolutional neural network," *Sensors*, vol. 20, no. 17, 4696, 2020.

[64] M. Jian, J. Wang, H. Yu, and Y. Ju, "Visual saliency detection based on object-locat ion cues and background features," in *Proc. 2019 25th International Conference on Automation and Computing (ICAC)*, 2019.

[65] M. Jian, J. Wang, H. Yu, G. Wang, X. Meng, L. Yang, J. Dong, and Y. Yin, "Visual saliency detection by integrating spatial position prior of object with background cues," *Expert Systems with Applications*, vol. 168, 114219, 2021.

[66] P. Peng, K.-F. Yang, F.-Y. Luo, and Y.-J. Li, "Saliency detection inspired by topological perception theory," *International Journal of Computer Vision*, vol. 129, no. 8, pp. 2352–2374, 2021.

[67] S. Yu, B. Zhang, J. Xiao, and E. G. Lim, "Structure-consistent weakly supervised salient object detection with local saliency coherence," in *Proc. the AAAI Conference on Artificial Intelligence*, 2021, vol. 35, no. 4, pp. 3234–3242.

[68] S. Gao, W. Zhang, Y. Wang, Q. Guo, C. Zhang, Y. He, and W. Zhang, "Weakly-supervised salient object detection using point supervison," arXiv preprint, arXiv:2203.11652, 2022.

[69] Y. Wang, J. Zhang, M. Kan, S. Shan, and X. Chen, "Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12275–12284.

[70] Y. Zeng, Y. Zhuge, H. Lu, L. Zhang, M. Qian, and Y. Yu, "Multi-source weak supervision for saliency detection," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6074–6083.

[71] D. Zhang, J. Han, and Y. Zhang, "Supervision by fusion: Towards unsupervised learning of deep salient object detector," in *Proc. the IEEE International Conference on Computer Vision*, 2017, pp. 4048–4056.

[72] X. Zhao, Y. Pang, L. Zhang, H. Lu, and L. Zhang, "Suppress and balance: A simple gated network for salient object detection," in *Proc. European Conference on Computer Vision*, 2020, pp. 35–51.

[73] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, and M. Jagersand, "Basnet: Boundary-aware salient object detection," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 7479–7489.

[74] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich*, Germany, October, 2015, pp. 234–241.

[75] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353–367, 2010.

[76] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2083–2090.

[77] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2814–2821.

[78] D. B. Johnson, "Efficient algorithms for shortest paths in sparse networks," *Journal of the ACM (JACM)*, vol. 24, no. 1, pp. 1–13, 1977.

[79] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2014.

[80] D.-P. Fan, M.-M. Cheng, J.-J. Liu, S.-H. Gao, Q. Hou, and A. Borji, "Salient objects in clutter: Bringing salient object detection to the foreground," in *Proc. the European Conference on Computer Vision (ECCV)*, 2018, pp. 186–202.

[81] R. Margolin, L. Zelnik-Manor, and A. Tal, "How to evaluate foreground maps?" in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 248–255.

[82] J. Kim, D. Han, Y.-W. Tai, and J. Kim, "Salient region detection via high-dimensional color transform," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 883–890.

[83] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang, "Saliency detection via absorbing markov chain," in *Proc. the IEEE International Conference on Computer Vision*, 2013, pp. 1665–1672.

[84] S. Joseph and O. O. Olugbara, "Detecting salient image objects using color histogram clustering for region granularity," *Journal of Imaging*, vol. 7, no. 9, p. 187, 2021.

[85] S. Chen, L. Zheng, X. Hu, and P. Zhou, "Discriminative saliency propagation with sink points," *Pattern Recognition*, vol. 60, pp. 2–12, 2016.

[86] J. Wang, H. Lu, X. Li, N. Tong, and W. Liu, "Saliency detection via background and foreground seed selection," *Neurocomputing*, vol. 152, pp. 359–368, 2015.

[87] J. Lou, H. Wang, L. Chen, F. Xu, Q. Xia, W. Zhu, and M. Ren, "Exploiting color name space for salient object detection," *Multimedia Tools and Applications*, vol. 79, no. 15, pp. 10873–10897, 2020.

[88] G.-H. Liu and J.-Y. Yang, "Exploiting color volume and color difference for salient region detection," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 6–16, 2018.

[89] L. Zhou, Z. Yang, Q. Yuan, Z. Zhou, and D. Hu, "Salient region detection via integrating diffusion-based compactness and local contrast," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3308–3320, 2015.

[90] X. Wu, X. Ma, J. Zhang, A. Wang, and Z. Jin, "Salient object detection via deformed smoothness constraint," in *Proc. 2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 2815–2819.

[91] Y. Duan, X. Zhou, J. Zou, J. Qiu, J. Zhang, and Z. Pan, "Mask-guided noise restriction adversarial attacks for image classification," *Computers & Security*, vol. 100, 102111, 2021.