

U-FIN: Unsupervised Feature Integration Approach for Salient Object Detection



Vivek Kumar Singh and Nitin Kumar

Abstract Salient object detection is a challenging research field in computer vision. The existing saliency detection methods generally focus on finding feature maps for saliency computation. However, the combination of these feature maps significantly improves salient region(s) detection. In this paper, we propose a novel feature integration approach called U-FIN in which final saliency map is obtained by a weighted combination of individual feature maps. The proposed approach works in three phases *viz.* (i) artifact reference (AR) map generation (ii) weight learning and (iii) final saliency map computation. Firstly, AR map is produced using majority voting on the individual feature maps extracted from the input image. Secondly, linear regression is employed for weight learning which is used in the next phase. Finally, the individual feature maps are linearly combined using weights learned in the second phase to generate the final saliency map. Extensive experiments are conducted on two benchmark datasets, i.e., ASD and ECSSD to validate the proposed feature integration approach. The performance is measured in terms of precision, recall, receiver operating characteristic (ROC) curve, F-measure and area under the curve (AUC). Extensive experiments demonstrate the superiority of the proposed U-FIN approach against nine state-of-the-art saliency methods on ASD dataset and comparable on ECSSD dataset with the best performing methods.

Keywords Combine · Linear regression · Majority voting · Artifact reference map

1 Introduction

The phenomenon of the human visual system is simulated in salient object detection. Salient object detection approach rapidly extracts more relevant information in a scene. It attempts to locate visually more prominent and conspicuous objects/regions

V. Kumar Singh (✉) · N. Kumar
National Institute of Technology Uttarakhand, Srinagar (Garhwal), Uttarakhand, India
e-mail: vivek.kumarsingh@nituk.ac.in

N. Kumar
e-mail: nitin@nituk.ac.in

© Springer Nature Singapore Pte Ltd. 2021
G. S. Hura et al. (eds.), *Advances in Communication and Computational Technology*, Lecture Notes in Electrical Engineering 668,
https://doi.org/10.1007/978-981-15-5341-7_89

1173

in an image. Saliency detection is a more attractive and challenging research area in various fields such as neuroscience, psychology and computer vision. Salient object detection devoted to compute a saliency map [9] that highlights most significant part(s) in an image. It has also been deemed as preprocessing step to rectify the computational time in variety of visual applications such as object detection [20], video summarization [15], visual tracking [32] and image classification [26].

In the last decade, a number of saliency detection methods have been investigated to achieve efficient performance in a robust manner. However, this problem is still challenging specifically on complex images. Salient object detection (SOD) methods are broadly divided into two categories [30] (a) bottom-up and (b) top-down methods based on the way in which visual cues are explored. Bottom-up saliency detection methods [5, 9] exploit various low-level visual cues, i.e., color, intensity, texture and contrast, while top-down methods [17] entail training model and prior knowledge for computing saliency value of image elements. Typically, a single feature is not sufficient to capture salient object in an efficient and robust manner, e.g., frequency-tuned (FT) SOD [1], and hierarchical contrast (HC) [3] methods are single feature methods. In both these methods, contrast feature is employed for finding the saliency map, which is not appropriate for complex structure images. Besides, many saliency methods exploit multiple features and heuristic features combination approaches for saliency analysis such as linear [9] and nonlinear [8].

Learning-based feature integration methods were introduced by Liu et al. [13] who fused three novel visual feature maps, i.e., (a) color spatial distribution, (b) surround histogram and (c) center-multi-scale contrast using a weight vector. This is a supervised learning method, and weights are learnt using conditional random field (CRF). Feature integration approach defines the role of each feature in the saliency computation. Hence, the performance of these kinds of methods mainly depends upon the weights which are used for combination of individual features maps. A simple approach can be to linearly combine all the features with equal weights, but the performance may be poor due to the fact that all the features may not equally highlight salient regions. Another approach can be to derive single weight vector for all the natural images similar to Liu et al. [13]. The performance may again be poor due to the diverse characteristics of natural images. Based on the above discussion, we have made an attempt to alleviate the problem of feature combination approach by deriving image-dependent weights in an unsupervised manner.

Here, we propose an unsupervised feature integration (U-FIN) approach which derives image-dependent weights by using unsupervised method. The feature integration approach has three phases: (i) artifact reference (AR) map generation (ii) weight learning and (ii) final saliency map computation. Firstly, AR map is produced using majority voting on the individual feature maps extracted from the input image. Secondly, linear regression (LR) is employed for learning weight. Finally, the individual feature maps are linearly combined to generate the final saliency map. In this paper, our contribution is twofold:

1. A novel feature integration approach is proposed which derives weights in an unsupervised manner using linear regression.
2. Extensive validation is performed on two publicly available datasets to exhibit the better performance of the proposed approach.

2 Related Work

In the last few years, numerous saliency detection methods have been developed, and fabulous performance has been achieved. The early saliency computation work was prompted by capturing the visual attention process of human visual system (HVS). First computational on salient object detection was proposed by Itti et al. [9] in which feature integration theory [22] with biologically plausible visual attention system [10] was explored to generate saliency map. Itti et al. [9] proposed model extract various contrast feature maps, namely orientation, luminance and color based on center-surround approach across multiple scales, and after that normalize all the features and aggregate for generating the saliency map. A number of methods have extended Itti et al. [9] work in different disciplines such as Walther et al. [23] had extended it to highlight proto-object and Han et al. [7] extended it with Markov random field (MRF) and region growing approach to identify salient objects. The center-surround contrast has been extensively used either locally or globally in many existing saliency detection methods since it clearly highlights salient region from their surrounds regions. The center-surround mechanism is studied across variety of visual features, *viz.* color, shape and texture [14]. Zhang et al. [29] measures saliency based on information theory where the uniqueness is represented using self-information of local image features. Seo and Milanfar [21] proposed saliency method in which local regression kernels-based self-resemblance is utilized for saliency estimation. Rahtu et al. [19] proposed saliency method that integrates saliency measures obtained by jointly consideration of a statistical framework and local feature contrast with a conditional random field (CRF). Murray et al. computed weighted center-surround maps and applied inverse wavelet transform (IWT) for generation of saliency map.

Furthermore, global knowledge of visuals has been exploited in different directions to compute saliency map. Context-aware saliency detection approach proposed by Goferman et al. [5] which incorporates local center-surround difference along the global distinctive few visual organization principles and color feature to compute saliency map. The statistical information of image has been exploited to build foreground/background model which assigns saliency value to image elements based on posterior probability of foreground model to background model [30]. Li et al. [12] learnt the prior information for saliency estimation. In [11], saliency is analyzed in the frequency domain that which part of the frequency spectrum significantly contributed to saliency estimation. Additionally, many saliency detection methods decomposed the image into regions by applying either segmentation or clustering approach. Such participation of images is helpful for incorporating global knowledge at region level [3]. Ren et al. [20] proposed effective region-based

saliency computation approach that decomposed input image into the perceptually and semantically meaningful regions, and saliency of each region is measured based on spatial compactness using Gaussian mixture model (GMM). Furthermore, Fang et al. [4] suggested an approach in which discriminative subspaces are learnt for image saliency computation. Zeng et al. [28] proposed saliency estimation based on an unsupervised game-theoretic approach which does not depend on labeled training data.

Recently, deep learning-based methods have been proposed that influence performance greatly, but performance of these models entirely depends on large number of training data for optimizing network learnable parameters which increase computational time. Wang et al. [25] suggested saliency measure approach in which two deep neural network (DNNs) are trained to extract local features and global search, respectively. Context-based DNN is suggested by Zhao et al. [31] that constructs multi-context DNN with the consideration of local and global context. Pan et al. [18] proposed various saliency estimation approaches using convolutional neural networks (CNN) which greatly reduce computational cost. Guan et al. [6] proposed edge-aware CNN in which global contextual knowledge is combined along with the low-level edge features for saliency measure. Wang et al. [24] exploited recurrent fully convolutional networks (RFCNs) that incorporated saliency priors to generate saliency map.

3 Proposed Approach

In this section, we illustrate the framework of the proposed model in which features are integrated using a three-phase approach, i.e., (i) artifact reference map generation (ii) weight learning using linear regression and (iii) final saliency map generation.

In the first phase, more than one visually distinguishing feature maps are extracted from an image. In this proposed model, we have employed three features maps, *viz.* color spatial distribution, multi-scale contrast and center-surround histogram as suggested by Liu et al. [13]. A Gaussian image pyramid is employed for multi-scale contrasts which are linearly added to derive multi-scale contrast feature map. This is local feature that perseveres high-contrast boundaries (i.e., edges) while suppressing homogeneous regions. The center-surround histogram is regional features which significantly highlight salient object that is distinctive with its surroundings. This feature is calculated with the consideration of surroundings for salient object and measures the distinctiveness as the distance between histograms of RGB color of salient object and its surroundings. The global information of image is captured using color spatial distribution. Larger a color is scattered in the image, then it is less likely to be contained by salient object. Hence, the global spatial distribution of a certain color is utilized to compute saliency of regions. The spatial distribution of color can be calculated as spatial variance of the color. The Gaussian mixture model (GMM) is used statistically to describe all colors of image and assign all color belongingness probability to each pixel. Then, variance of each color is computed,

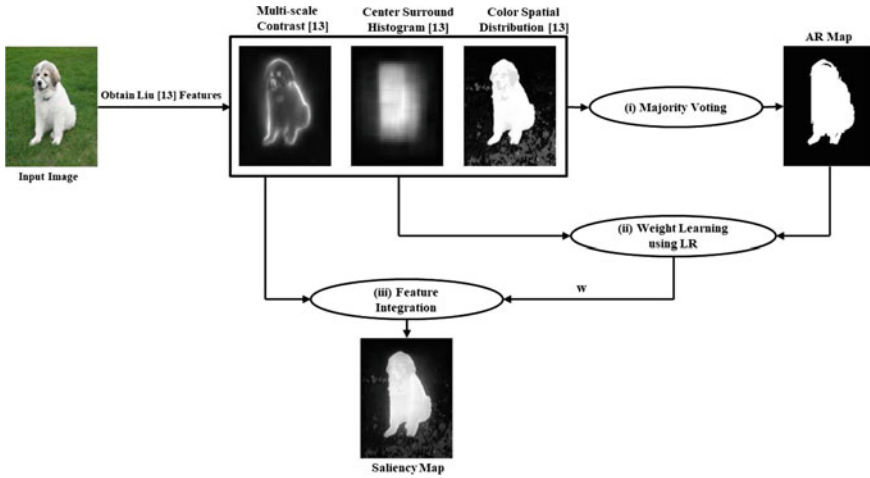


Fig. 1 A schematic representation of the proposed approach. Several features like color spatial distribution, center-surround histogram and multi-scale contrast are extracted using Liu [13] saliency method

and using these variances, color spatial distribution is calculated. Further, the color spatial distribution is refined by image center weight. These features are integrated into various disciplines such as linear summation and weighted linear summation. All these feature maps are combined using majority voting, and the resultant labeled map is termed as artifact reference (AR) map.

In phase two, linear regression (LR) is employed to learn the weights for combining initial feature maps. The AR map is used as the target map. Thus, instead of using human-annotated map of an image, our approach uses the estimated AR map. Hence, the proposed approach entails unsupervised learning mechanism and presents a novel unsupervised learning-based feature integration approach which learns integration weights for each image. In phase three, a final saliency map is found by combining the initial feature maps with corresponding weights learnt in the previous phase.

The architecture of the proposed feature integration approach is delineated in Fig. 1. First, the saliency method [13] is utilized to extract various features from the given input image. These features are incorporated with majority of vote process to obtain AR map. Further, features and AR map are fed into LR, and a set of weights (w) are learnt. Afterward, the features are linearly combined using w to generate final saliency map S . Next, we provide the mathematical formulation of the proposed approach.

3.1 Artifact Reference (AR) Map Generation

The feature maps of an image are obtained using Liu et al. [13] method. These feature maps are represented as a set of features $\mathbf{F} = \{\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_N\}$ where N is the number of feature maps. \mathbf{F} is transformed into a set of classified map in which pixel value is either 0 or 1. Suppose \mathbf{C}_i is the classified map corresponding to feature \mathbf{F}_i . Thus, all the classified map can be represented as $\mathbf{C} = \{\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_N\}$. The classified map is obtained using adaptive thresholding as suggested by Achanta et al. [1] where the threshold (T_i) for i -th feature map (\mathbf{F}_i) is computed as follows:

$$T_i = \frac{2}{I_w \times I_h} \sum_{x=1}^{I_w} \sum_{y=1}^{I_h} \mathbf{F}_i(x, y) \quad i = 1, 2, \dots, N \quad (1)$$

where I_w and I_h are width and height of the input image. Hence, classified map \mathbf{C}_i corresponding to feature \mathbf{F}_i is computed as follows:

$$\mathbf{C}_i(x, y) = \begin{cases} 1 & \text{if } \mathbf{F}_i(x, y) \geq T_i \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Here, (x, y) represents the location of the pixel under consideration such that $1 \leq x \leq I_w$ and $1 \leq y \leq I_h$.

Hence, the classified maps thus contain only two values, i.e., 1 and 0 where 1 denotes salient region and 0 denotes background region in a given image. Therefore, the classified map is annotated map which partitions the input image pixels into two parts. Further, we use these classified maps for generating artifact reference map. Since these classified maps have class labels, we apply the majority vote scheme to obtain a artifact reference map which is act like human annotation map in a better manner. To find the artifact reference map \mathbf{AR} for the input image, the following equation is used:

$$\mathbf{AR}(x, y) = \begin{cases} 1 & \text{if } \sum_{i=1}^N \mathbf{C}_i(x, y) > N/2 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

For proper working of the above equation, N must be an odd number. In this research work, we have chosen $N = 3$.

3.2 Weight Learning Using Linear Regression

Linear regression using gradient descent learns image-dependent weights for combination of various feature maps of an image. Each pixel value is described with the help of a set of features (i.e., color spatial distribution, multi-scale contrast and center-

surround histogram) given in Liu et al. [13] as a feature vector $\mathbf{x} = \{x_1, x_2, \dots, x_N\}$ where N is the number of features. Hence, the i -th feature of an image \mathbf{I} is represented as $\mathbf{A}_i = \{x_1(i), x_2(i), \dots, x_p(i)\}$, $i = 1, 2, \dots, N$, where $\mathbf{A}_i \in \mathbb{R}^p$ and $p = I_w \times I_h$. The set of features $\mathbf{A} = \{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_N\}$, where $\mathbf{A} \in \mathbb{R}^{p \times N}$ and corresponding artifact reference (AR) map $\mathbf{Y} = \{y_1, y_2, \dots, y_p\}$, where $\mathbf{Y} \in \mathbb{R}^p$. The proposed linear regression is mathematically defined as follows:

$$\Phi : (\mathbb{R}^N | \mathbf{w}) \rightarrow \mathbb{R} \tag{4}$$

where $\mathbf{w} = \{w_1, w_2, \dots, w_{N+1}\}$ is set of image-dependent weights. Initially, \mathbf{w} is set to zero and is gradually adjusted during learning in order to reduce error between combined features output and the AR map. Consequently, the linear regression is obtained fitted weights which is further used from features combination task. The linear regression predicts pixel-wise output and is denoted as \hat{y}_j for j -th pixel in the given image and mathematically represented as:

$$\hat{y}_j = \Phi(\mathbf{x}_j | \mathbf{w}) \tag{5}$$

$$\Phi(\mathbf{x}_j | \mathbf{w}) = \sum_{\substack{i=1 \\ x_i \in \mathbf{x}_j}}^N w_i x_i + w_{N+1} \tag{6}$$

The linear regression predicts saliency map for an image as follows:

$$\Phi(\mathbf{A} | \mathbf{w}) = \{\Phi(\mathbf{x}_1 | \mathbf{w}), \Phi(\mathbf{x}_2 | \mathbf{w}), \dots, \Phi(\mathbf{x}_p | \mathbf{w})\} \tag{7}$$

$$\hat{\mathbf{Y}} = \Phi(\mathbf{A} | \mathbf{w}) \tag{8}$$

where $\hat{\mathbf{Y}}$ is predicted saliency map of the given image. Linear regression is utilized mean square error cost function between predicted saliency map $\hat{\mathbf{Y}}$ and AR map \mathbf{Y} to evaluate goodness of weights. The cost function $L(\mathbf{x}_j | \mathbf{w})$ gives the error between j -th pixel predicted output and its artifact reference value as given in Eq. 9:

$$L(\mathbf{x}_j | \mathbf{w}) = (y_j - \Phi(\mathbf{x}_j | \mathbf{w}))^2 \tag{9}$$

$$L(\mathbf{x}_j | \mathbf{w}) = (y_j - \hat{y}_j)^2 \tag{10}$$

Similarly, we can define the cost function for an input images as follows:

$$L(\mathbf{A} | \mathbf{w}) = \frac{1}{p} \sum_{j=1}^p (y_j - \hat{y}_j)^2 \tag{11}$$

Thus, our objective is to minimize the cost function $L(\mathbf{A} | \mathbf{w})$ whose solution is obtained using gradient descent algorithm.

3.3 Final Saliency Map Generation

The weight vector \mathbf{w} learnt for a specific image is used to integrate extracted features. The set of features \mathbf{A} and learnt weights are incorporated to compute final saliency map as weighted linear combination of features as follows:

$$\mathbf{S} = \sum_{i=1}^N w_i \mathbf{A}_i + w_{N+1} \quad (12)$$

Thereafter, the saliency map \mathbf{S} is normalized in the range of [0, 1] as follows:

$$\mathbf{S} = \frac{\mathbf{S} - \theta_{\min}(\mathbf{S})}{\theta_{\max}(\mathbf{S}) - \theta_{\min}(\mathbf{S})} \quad (13)$$

where θ_{\max} and θ_{\min} are operators which find maximum and minimum value from the input matrix, respectively.

4 Experimental Setup and Results

In this section, we discuss the experimental outcomes to analyze of the proposed feature integration approach across various state-of-the-art methods on two publicly available salient object datasets, i.e., ASD [1] and ECSSD [27]. ASD dataset is widely used dataset which contains 1000 natural images with variety of salient object from the MSRA-5000 saliency detection dataset [13]. ECSSD [27] dataset consists of 1000 images which shows diversity in terms of semantics and complexity constructed from the Web resources. The human annotations (i.e., ground truth labels) are obtained using five observers. Further illustrating the superiority of the proposed feature integration approach, its performance is compared with nine state-of-the-art saliency detection methods *viz.* Liu [13], SUN [29], SeR [21], CA [5], SEG [19], SIM [16], SP [12], SSD [11], LDS [4]. The validation is conducted in two different aspects: qualitative and quantitative.

The quantitative study is carried out with five performance measures, i.e., recall, precision, receiver operating characteristics (ROC), F-measure and area under the ROC curve(AUC), for validation of the proposed feature integration approach. Precision and recall are calculated by inferring an overlapped region of saliency map (\mathbf{S}) with human annotation, i.e., ground truth (\mathbf{G}). The strength of saliency methods as predicted salient regions are likely salient is depicted by precision. However, recall reveals the strength of methods in the form of completeness of real salient regions.

Besides, F-measure is illustrated as a weighted combination of precision and recall for comprehensive validation. All these metrics are mathematically represented as follows [2]:

$$\text{Precision} = \frac{|\mathbf{B} \cap \mathbf{G}|}{|\mathbf{B}|} \quad (14)$$

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

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

where \mathbf{B} is a binary map corresponding to saliency map \mathbf{S} which is generated with the help of an adaptive threshold as reported in [1]. The operator $|\cdot|$ is used to find sum of ones in the binary map in the enclosed binary labeled matrix. The β is fixed with 0.3 during all the experiments as suggested in [1] to more emphasize on precision than recall. Further, ROC is delineated using false positive rate (FPR) and true positive rate (TPR) where false positive rate (FPR) and true positive rate (TPR) on the x-axis and y-axis in the plot, respectively. The TPR and FPR are computed using a sequence of thresholds which are varied between the range of [0, 1] with equal steps and formulated as follows [2]:

$$\text{TPR} = \frac{|\mathbf{B} \cap \mathbf{G}|}{|\mathbf{G}|} \quad (17)$$

$$\text{FPR} = \frac{|\mathbf{B} \cap \mathbf{N}|}{|\mathbf{N}|} \quad (18)$$

Another most widely used metric AUC is determined as the area covered beneath the ROC curve. The experimental parameters such as learning rate ($\alpha = 0.03$) and number of iterations ($I = 25$) which are used in LR for weight leaning are set empirically.

4.1 Performance Comparison with State-of-the-art Methods

We compare the proposed approach against nine state-of-the-art saliency methods qualitatively and quantitatively to illustrate the effectiveness of the proposed approach. Figure 2 demonstrates the qualitative performances of the proposed model and the compared well-performing state-of-the-art saliency methods. The columns (from left to right) show the first-third and fourth-sixth input images from ASD [1] and ECSSD [27] datasets, respectively.

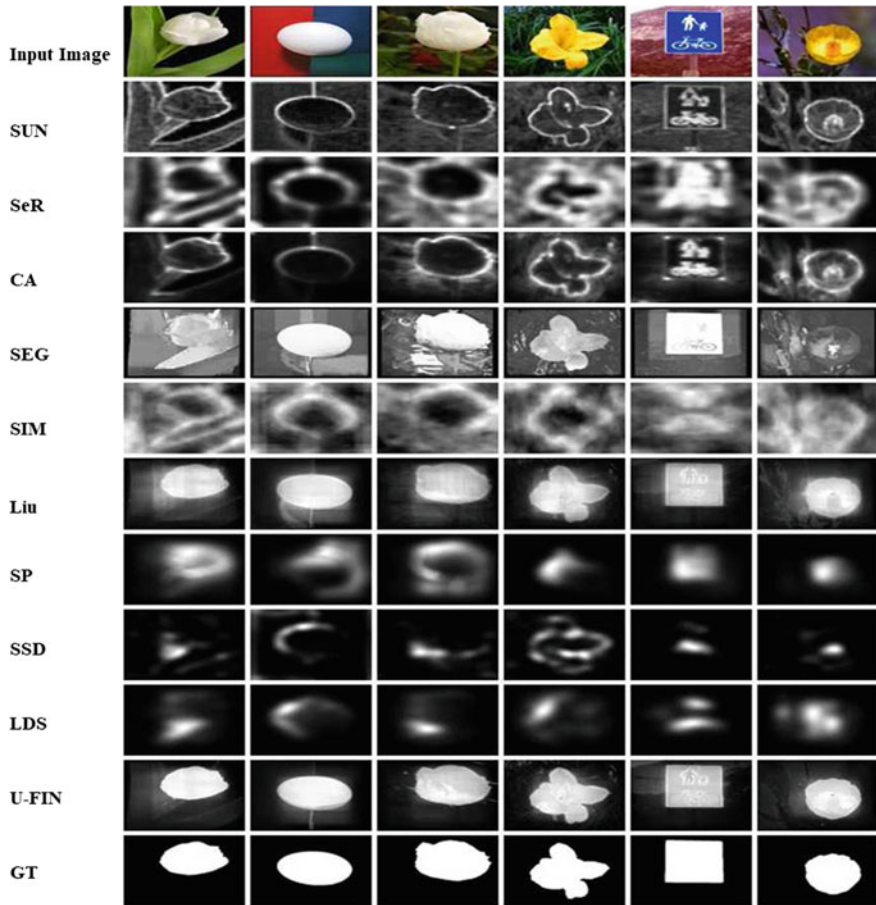


Fig. 2 Visual results of the proposed U-FIN approach with compared nine state-of-the-art methods on ASD [1] and ECSSD [27] datasets

These images represent different scenes such as single object, object near to image boundary and complex background. One can observe that some saliency methods such as SP [12], SSD [11], LDS [4], SIM [16] and SeR [21] fail to capture entire object even on simple image, e.g., fifth column. SUN [29] clearly detects the edge of object but fails to suppress background and highlights region inside object. However, Liu [13] and SEG [19] deliver better results on simple images, e.g., second and fifth columns, while fail to suppress background on complex structure images, e.g., first and sixth columns. In contrast, the proposed approach U-FIN performs uniformly on each of these images and clearly suppresses background in comparison with the second good performing saliency method, i.e., Liu [13] as shown in the first, third and sixth columns.

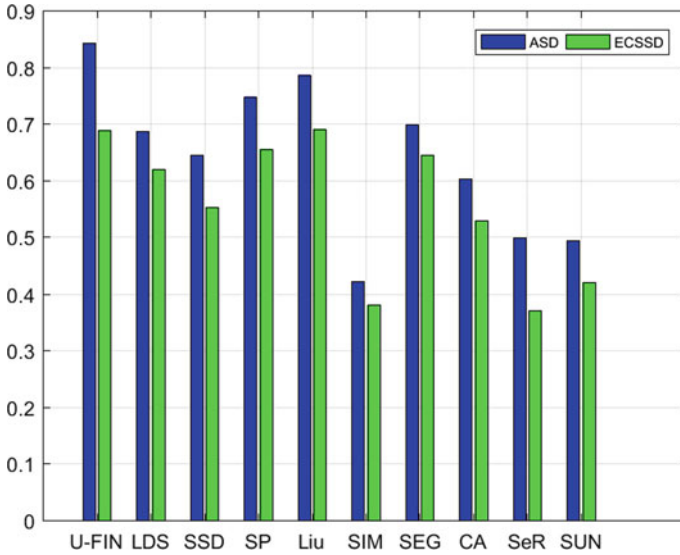


Fig. 3 Precision scores of the proposed U-FIN approach with compared nine state-of-the-art methods on ASD [1] and ECSSD [27] datasets

The quantitative analysis of the proposed U-FIN approach with compared state-of-the-art saliency detection methods in terms of *Precision*, *Recall*, *F-measure*, *AUC* and *ROC curve* is shown in Figs. 3, 4, 5, 6 and 7, respectively. It can be readily observed that on ASD [1] dataset, U-FIN outperforms compared state-of-the-art saliency detection methods. However, SIM [16] and SUN [29] are the worst performers in terms of F-measure, recall, precision, AUC and ROC curve, respectively. On ECSSD [27] dataset, U-FIN outperforms the other compared methods in terms of F-measures and equally performs with Liu [13] in terms of AUC and ROC curve. The proposed approach performs better than Liu et al. [13] in terms of recall, but LDS [4] is the best among the compared methods. In terms of precision, Liu [13] performs best, while the proposed method is comparable with the top performer.

4.2 Computational Time

The computational times of the proposed feature integration approach along with the compared saliency methods on the ASD [1] dataset are reported in Fig. 8. The dataset contains images whose size is 400×300 . The execution timings have been obtained on a desktop PC that configured with the following specification: Intel(R)Core(TM)i7-4770 CPU@3.40GHz. As shown in Fig. 8, the proposed feature

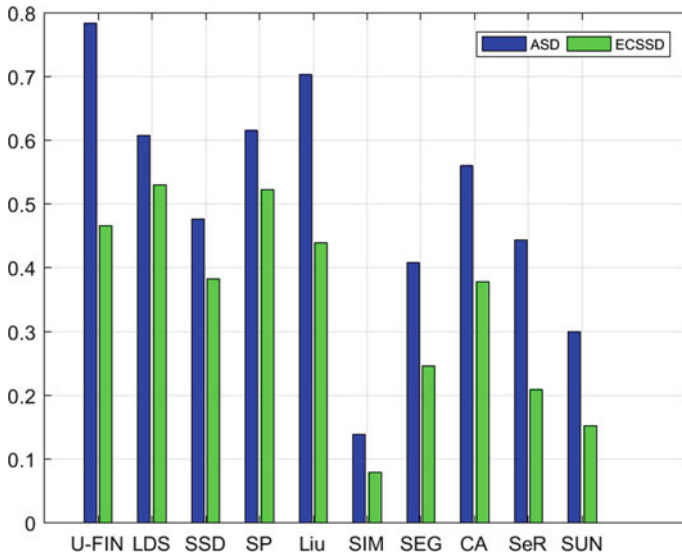


Fig. 4 Recall scores of the proposed U-FIN approach with compared nine state-of-the-art methods on ASD [1] and ECSSD [27] datasets

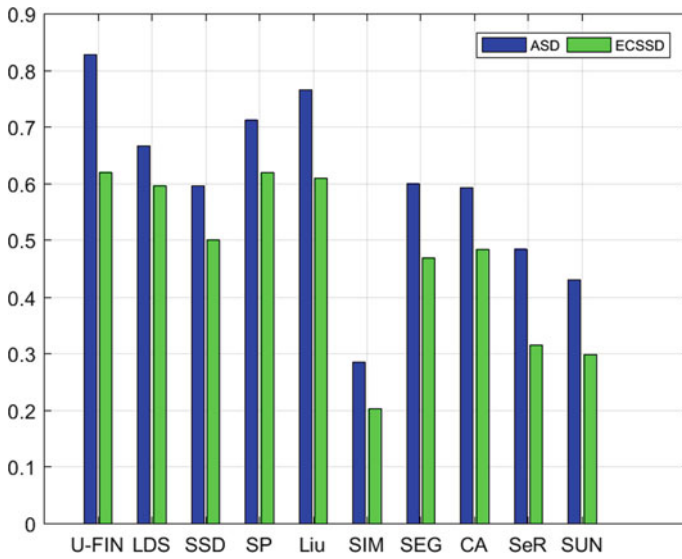


Fig. 5 F-measure scores the proposed U-FIN approach with compared nine state-of-the-art methods on ASD [1] and ECSSD [27] datasets

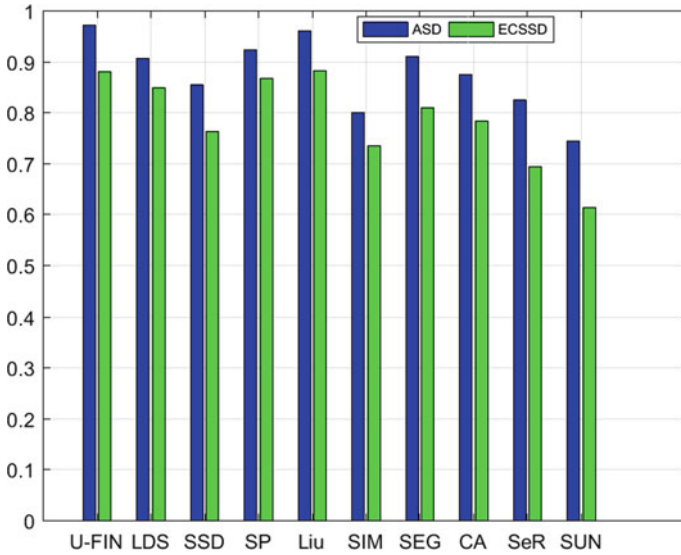


Fig. 6 AUC scores of the proposed U-FIN approach with compared nine state-of-the-art methods on ASD [1] and ECSSD [27] datasets

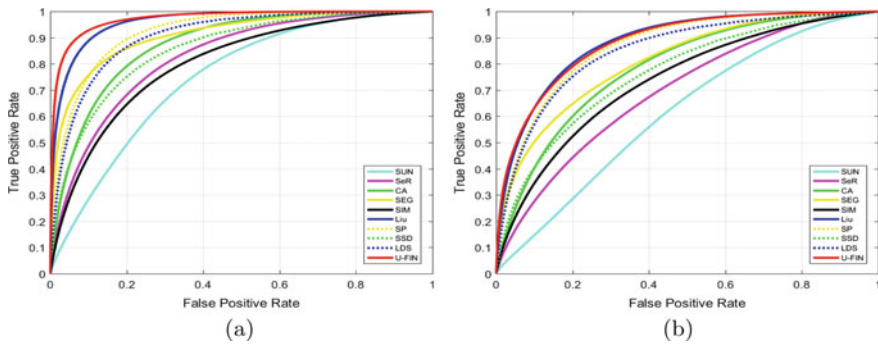


Fig. 7 ROC on the two widely used datasets: **a** ASD [1], **b** ECSSD [27]

integration approach is faster than CA [5], while SSD [11], LDS [4], SP [12], SeR [21] and SIM [16] are better than the proposed feature integration approach, and from other, it is comparable. Although the proposed method is computationally more expensive than several methods, the same can be mitigated with the improvement noticed in performance.

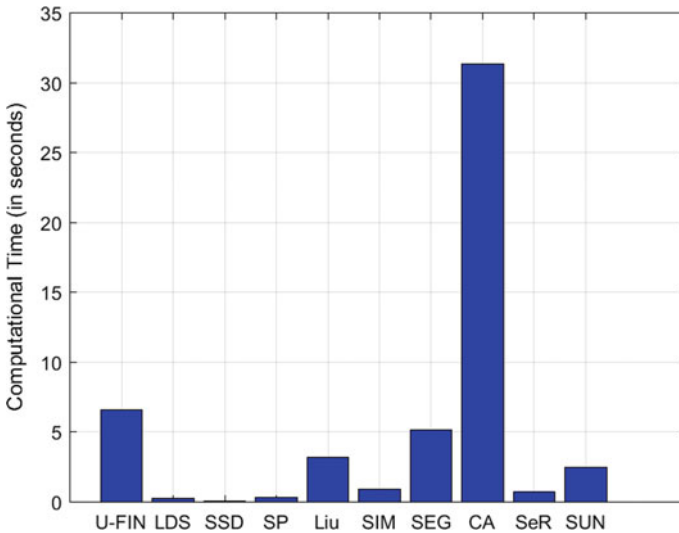


Fig. 8 Computational time analysis of the proposed U-FIN approach with compared nine state-of-the-art methods on ASD [1] dataset

5 Conclusion and Future Work

In this paper, we have presented a novel feature integration approach U-FIN in which image-dependent weights are learnt using linear integration of features extracted from the input image in an unsupervised manner. Initially, artifact reference (AR) map is produced from a set of features extracted from the image. This map assists in leaning the appropriate weights to combine specific image features. Further, linear regression (LR) model is built using gradient descent to learn weights for specific image features. Finally, these weights are used to linearly combine features to generate final saliency map. A comprehensive evaluation has been shown on two publicly available benchmark datasets, i.e., ASD and ECSSD, that show the effectiveness of the proposed U-FIN approach. It is also found that U-FIN is superior than nine state-of-the-art saliency methods on ASD dataset and comparable on ECSSD dataset. In our future work, we will extend the current feature integration approach with the selection of efficient feature maps and alternative feature integration approach(s).

References

1. Achanta R, Hemami S, Estrada F, Süsstrunk S (2009) Frequency-tuned salient region detection. In: IEEE international conference on computer vision and pattern recognition (CVPR 2009), No. conf, pp 1597–1604
2. Borji A, Cheng MM, Jiang H, Li J (2015) Salient object detection: a benchmark. *IEEE Trans Image Process* 24(12):5706–5722

3. Cheng MM (2011) Global contrast based salient region detection. In: Conference on computer vision and pattern recognition (CVPR), IEEE, pp 409–416
4. Fang S, Li J, Tian Y, Huang T, Chen X (2017) Learning discriminative subspaces on random contrasts for image saliency analysis. *IEEE Trans Neural Netw Learn Syst* 28(5):1095
5. Goferman S (2010) Context-aware saliency detection. In: Proceedings of IEEE conference computer vision and pattern recognition, pp 2376–2383
6. Guan W, Wang T, Qi J, Zhang L, Lu H (2018) Edge-aware convolution neural network based salient object detection. *IEEE Signal Processing Letters* 26(1):114–118
7. Han J, Ngan K, Li M, Zhang HJ (2006) Unsupervised extraction of visual attention objects in color images. *IEEE Transactions on Circuits and Systems for Video Technology* 16(1):141–145
8. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: Advances in neural information processing systems. pp. 545–552 (2007)
9. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 11:1254–1259
10. Koch C, Ullman S (1987) Shifts in selective visual attention: towards the underlying neural circuitry. *Matters of intelligence*. Springer, Berlin, pp 115–141
11. Li J, Duan LY, Chen X, Huang T, Tian Y (2015) Finding the secret of image saliency in the frequency domain. *IEEE Trans Pattern Anal Mach Intell* 37(12):2428–2440
12. Li J, Tian Y, Huang T (2014) Visual saliency with statistical priors. *Int J Comput Vis* 107(3):239–253
13. Liu T, Yuan Z, Sun J, Wang J, Zheng N, Tang X, Shum H (2011) Learning to detect a salient object. *IEEE Trans Pattern Anal Mach Intell* 33(2):353
14. Ma YF, Zhang HJ (2003) Contrast-based image attention analysis by using fuzzy growing. In: Proceedings of the eleventh ACM international conference on Multimedia. ACM, pp 374–381
15. Marat S, Guironnet M, Pellerin D (2007) Video summarization using a visual attention model. In: 2007 15th European signal processing conference. IEEE, pp 1784–1788
16. Murray N, Vanrell M, Otazu X, Parraga CA (2011) Saliency estimation using a non-parametric low-level vision model. In: 2011 IEEE conference on computer vision and pattern recognition (cvpr). IEEE, pp 433–440
17. Oliva A, Torralba A, Castelhana MS, Henderson JM (2003) Top-down control of visual attention in object detection. In: Proceedings 2003 international conference on image processing (Cat. No. 03CH37429), vol 1. IEEE, pp 1–253
18. Pan H, Wang B, Jiang H (2015) Deep learning for object saliency detection and image segmentation. arXiv preprint [arXiv:1505.01173](https://arxiv.org/abs/1505.01173)
19. Rahtu E, Kannala J, Salo M, Heikkilä J (2010) Segmenting salient objects from images and videos. In: Computer vision—ECCV 2010. Springer, pp 366–379
20. Ren Z, Gao S, Chia LT, Tsang IWH (2013) Region-based saliency detection and its application in object recognition. *IEEE Trans Circ Syst Video Technol* 24(5):769–779
21. Seo HJ, Milanfar P (2009) Static and space-time visual saliency detection by self-resemblance. *J Vis* 9(12):15–15
22. Treisman AM, Gelade G (1980) A feature-integration theory of attention. *Cogn Psychol* 12(1):97–136
23. Walther D, Koch C (2006) Modeling attention to salient proto-objects. *Neural Netw* 19(9):1395–1407
24. Wang L, Wang L, Lu H, Zhang P, Ruan X (2019) Salient object detection with recurrent fully convolutional networks. *IEEE Trans Pattern Anal Mach Intell* 41(7):1734
25. Wang L, Lu H, Ruan X, Yang MH (2015) Deep networks for saliency detection via local estimation and global search. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3183–3192
26. Wang Q, Lin J, Yuan Y (2016) Salient band selection for hyperspectral image classification via manifold ranking. *IEEE Trans Neural Netw Learn Syst* 27(6):1279–1289
27. Yan Q, Xu L, Shi J, Jia J (2013) Hierarchical saliency detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1155–1162

28. Zeng Y, Feng M, Lu H, Yang G, Borji A (2018) An unsupervised game-theoretic approach to saliency detection. *IEEE Trans Image Process* 27(9):4545–4554
29. Zhang L, Tong MH, Marks TK, Shan H, Cottrell GW (2008) Sun: a bayesian framework for saliency using natural statistics. *J Vis* 8(7):32–32
30. Zhang W, Wu QJ, Wang G, Yin H (2010) An adaptive computational model for salient object detection. *IEEE Transactions on Multimedia* 12(4):300–316
31. Zhao R, Ouyang W, Li H, Wang X (2015) Saliency detection by multi-context deep learning. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1265–1274
32. Zhou T, He X, Xie K, Fu K, Zhang J, Yang J (2015) Robust visual tracking via efficient manifold ranking with low-dimensional compressive features. *Pattern Recogn* 48(8):2459–2473