Salient Region Detection via Multiple Hierarchy and Competition Mechanism

Dongtao Zhu,1,2 Xing Yang,1,2 Yihua Hu,1,2 and Zhenyu Liang1,2

1State Key Laboratory of Pulsed Power Laser Technology, National University of Defense Technology, Hefei 230037, China
2Anhui Province Key Laboratory of Electronic Restriction, National University of Defense Technology, Hefei 230037, China

Correspondence should be addressed to Xing Yang; yangxing17@nudt.edu.cn

Received 18 January 2022; Accepted 29 March 2022; Published 28 May 2022

1. Introduction

With the development of image processing technology, large amounts of image data can be quickly acquired, which is a challenge to the efficiency of image processing algorithms. According to the visual attention mechanism, the human eye can quickly select the most interesting objects and ignore useless information in the scene, and the saliency detection task aims to identify these important regions. In recent years, saliency detection has gradually become a research hotspot in machine vision, and an increasing number of researchers have focused on this problem. As an important preprocessing procedure, saliency detection contributes to many machine vision and image processing tasks such as visual tracking [1], image compression [2], object recognition [3], and image segmentation [4].

The saliency detection method consists of salient object detection and gaze predictions [5]. Existing salient object detection methods can be divided into two categories: top-down [6] and bottom-up [7], [8]. Top-up approaches are task-driven, and supervised learning frameworks typically use top-down strategies such as deep learning or convolutional neural networks. Although deep learning has achieved good results in salient object detection, it requires numerous datasets for training and involves many parameters, which complicates parameter tuning. Conversely, bottom-up methods are driven by vision; their performance relies on various low-level visual feature extraction methods, such as color features, texture, and orientation.

Many salient object detection models exist, among which graph-based methods measure the similarity between nodes when generating saliency graphs [9]. For superpixel images with superpixels as nodes, the similarity weight between nodes is regarded as a graph model, and the saliency information of the original image can be extracted through node propagation. In these methods, initial node selection is the key to the salient object detection process. Existing node selection methods introduce various types of prior knowledge to ensure the accuracy of the foreground or background node selection. For example, the boundary prior [10] considers the presence of salient objects at the boundary of an image, and the salient objects are connected to the boundary.
The center prior [11] considers that the salient object is located at the center of the image. However, this is not the case for natural images alone, where the salient object locations are randomly distributed. This is a significant challenge for existing prior-based methods.

As regards the advantages and disadvantages of saliency graph object detection methods based on node selection in graph methods, we can conclude that prior knowledge determination plays an important role in node selection; however, the saliency graphs generated by different prior knowledge determinations of initial nodes also differ. Some researchers have attempted to improve the accuracy of saliency graphs by determining the initial node. If the initial node is initially misjudged, the accuracy of the saliency graph is significantly reduced, and even its feature method is invalid. In addition, superpixel images generated using the simple linear iterative clustering (SLIC) algorithm [12] have different shapes and numbers of nodes generated by images of different sizes. In other words, image layering enriches the input features of salient objects, thereby ensuring the robustness of the detection results. It also provides a basis for subsequent feature selection.

Therefore, image multiple hierarchy and competition mechanisms are particularly important for saliency detection. In this study, we focus on bottom-up methods to compute saliency detection models. In contrast to traditional methods, the proposed scheme improves the accuracy and consistency of saliency results by fusing multilevel saliency priors and an integration mechanism of global similarity cues. The main contributions of the proposed method are as follows:

1. The use of multilayered images overcomes the nonrobustness problem of saliency measurements because of the size of the salient objects. Simultaneously, the shapes and number of nodes generated by images of different scales are different, thus enriching the input features of the salient objects.

2. Three different feature methods are proposed in parallel to prevent prior knowledge from judging wrong nodes at the beginning, which is beneficial to the selection of later features.

3. A novel feature saliency map fusion mechanism was established. Firstly, feature maps are fused at the same level, and the two most similar coarse saliency maps are selected for fusion. Accordingly, two feature saliency maps with large differences were selected for fusion. The first competition removes features with low detection accuracy for individual features, thereby ensuring the effectiveness of saliency maps at the same level. The second competition enriches the feature map boundary features and ensures the detection accuracy of the saliency map.

The remainder of this paper is organized as follows. In Section 2, we review related studies on saliency detection. In Section 3, the construction of various hierarchical coarse saliency maps is presented. Section 4 presents a saliency graph refinement of the competition mechanism. Section 5 presents the experimental data and analysis. Finally, Section 6 presents conclusions and discussion.

2. Related Work

Existing salient object detection methods can be divided into two categories: bottom-up and top-down. Bottom-up methods are driven by vision, and their performance relies on various low-level visual feature extractions such as color features, texture, and orientation. To date, many low-level features exist in the saliency detection work, and it has been confirmed that features such as color, salient object size, location, and image boundaries directly influence our visual attention mechanism and play an important role in saliency detection. Although top-down approaches are task-driven, supervised learning frameworks typically use top-down strategies, such as deep learning or convolutional neural networks.

Visual saliency detection is mostly based on bottom-up methods that use low-level visual features in images and videos. A previous study on saliency detection was presented by Itti et al. [13], in which a final saliency map of an input image was obtained by exploiting the center-surround differences in low-level image features based on visual features (including color, intensity, and orientation maps). Ma and Zhang [14] proposed a method to detect regions of visual attention in images based on local contrast analysis, which in turn is based on center-surrounding and chromatic aberration calculations. Wei et al. [15] proposed two priors, boundary and connectivity, and assumed that the image boundary was the main background. Combining this prior with the geodesic distance, an effective salient object detection model called the GS model was designed. Saliency detection was performed using the background priors and distances. Perazzi et al. [16] decomposed a given image into compact and perceptually homogeneous elements, evaluated the uniqueness and spatial distribution of these elements, inferred the contrast of these elements to derive saliency measures, and generated pixel-accurate saliency maps. Yang et al. [17] ranked the similarity of image elements to foreground or background cues through graph-based MR, representing an image as a closed-loop graph with superpixels as nodes based on the similarity of background and foreground queries, which is ranked based on the affinity matrix to effectively extract background regions and foreground salient objects. Wu et al. [7] used two perceptual cues, proximity and similarity, to generate background nodes through background probability measures and established a label propagation model to generate saliency maps. Wu et al. [18] proposed a boundary-guided graph structure to explore the correlations between superpixels and built an iterative propagation mechanism to refine the saliency map. Wang et al. [19] used the convex hull of interest points to estimate the locations of salient objects by employing random walks to propagate nodes and generate saliency maps. Jian et al. [20] used a discrete wavelet transform to extract directional features, calculated the centroids of salient objects, and combined the color contrast and background features to generate robust saliency maps.
As proposed by Al-Azawi et al. [21], local saliency detection used the structure of objects to determine saliency, whereas global saliency detection identifies saliency based on background-related contrast and uses irregularity as a two-stage saliency measure. Chakraborty and Mitra [22] used random walks on Markov chains to obtain saliency maps, and k-dense submaps enhanced the salient regions in the image.

Among these saliency detection methods, graph-based methods generate saliency graphs that measure the similarity between nodes based on different rules. In these classes of methods, the initial node selection is the key to the salient object detection process. Existing node selection methods introduce various types of prior knowledge to ensure the accuracy of the foreground or background node selection. These priors are significant for carefully selected samples such as ID photos. However, this is not the case for natural images alone, where the salient object locations are randomly distributed. This is a significant challenge for existing prior-based methods. In addition, superpixel node images generated using the SLIC algorithm have different shapes and numbers of nodes generated by images of different sizes. In other words, image layering enriches the input features of salient objects, thereby ensuring the robustness of the detection results, thus providing a basis for the subsequent feature selection.

As regards the advantages and disadvantages of saliency graph object detection methods based on node selection in graphical methods, we can conclude that prior knowledge determination plays an important role in node selection; however, the saliency graphs generated by different prior knowledge determinations of initial nodes also differ. To improve the accuracy of saliency graphs, some researchers have attempted to improve the accuracy by determining the initial node. If the initial node is initially misjudged, the accuracy of the saliency graph is significantly reduced, and even its feature method is invalid.

Consequently, we propose a saliency map refinement method using a competitive mechanism. First, multiple feature maps are generated in parallel, and then, the feature saliency maps are further screened to determine useful and eliminate useless feature maps. Here, the similarity of the saliency maps generated by a certain number of features was compared, and effective feature maps were selected. For convenience, this study adopted three different types of features and compared the similarity of the saliency maps generated by the three features.

In contrast to bottom-up saliency detection models, top-down methods are based on high-level priors on images and are thus always task-specific-driven training and learning methods. In [6], Yang et al. used joint learning of a conditional random field (CRF) and a visual dictionary for salient object detection, which comprised a hierarchical structure from top to bottom: CRF, sparse coding, and image patches to sparse coding. As an intermediate layer, CRF is learned in a feature-adaptive manner and also used as an output layer to learn a dictionary under structured supervision, and a top-down saliency detection model was proposed. Shan et al. [23] proposed a neural network for three-stage layering: first, fast R-CNN extracts feature for each superpixel; second, an attention mechanism was used to expand the receptive field from one superpixel to the surrounding and related; finally, the value of the saliency score was obtained using the global regression model to generate a saliency map for top-down saliency detection. Although deep learning-based salient object detection has achieved good results [24], these methods are extremely sensitive to data attributes, which may influence their adaptability to different situations.

3. Construction of Multiple Layered Coarse Saliency Maps

The selection of initial nodes differs, the detection results of saliency graphs also differ significantly, and a competitive and complementary relationship exists between them. Based on various factors, such as salient object size, color, location, and image boundary, a multilevel hierarchical saliency detection algorithm with different prior knowledge is designed to determine the initial nodes. The algorithm framework of this study is illustrated in Figure 1. First, we divide the input image into three levels: Z1, Z2, and Z3, of different sizes. Subsequently, we simultaneously processed Z1, Z2, and Z3 by building three different types of saliency graph models. Finally, we defined a competition mechanism to fuse the saliency map twice to obtain a refined saliency map. In Figure 1, ① adopts the SLIC algorithm to generate super-pixel images, ②, ③, and ④ represent the feature method based on center prior, the feature method based on color prior, and the feature method based on boundary connectivity prior, respectively. ⑤, ⑥, and ⑦ represent the feature analysis of their respective feature maps, ⑧ feature map of comparison, and ⑨ feature fusion.

3.1. Coarse Saliency Map Generation Based on Boundary Prior. Zhu et al. [10] proposed a method to measure the degree of connection between region Y and the image boundary; that is, the boundary connectivity when the salient object is located at the image boundary. It is expressed as the ratio of the perimeter of a region on the boundary to the square root of the area of the region.

$$B(Y) = \frac{|\{v_i | v_i \in Y, v_i \in C\}|}{\sqrt{|\{v_i | v_i \in Y\}|}} = \frac{\text{Len}_{\text{bnd}}(Y)}{\sqrt{\text{Area}(Y)}}$$

(1)

where \(v_i\) is a node in area \(Y\), \(C\) is a collection of image boundary nodes, \(\text{Len}_{\text{bnd}}(\cdot)\) represents the perimeter of a region at the boundary, and \(\text{Area}(\cdot)\) represents the area of a region.

To facilitate the calculation of area \(Y\), \(v_i\) is placed as a set of nodes with similar colors.

$$\text{Area}(Q^Y_i) = \sum_{n=1}^{\text{len}} Z^Y_{ni}$$

(2)

where \(Z^Y_{ni}\) represents the total number of pixels of node \(v_i\), and \(|Q^Y_i|\) represents the number of nodes in \(Q^Y_i\). Similarly, the image boundary length is defined as follows:
where $\delta(\cdot)$ is 1 at the boundary. Otherwise, it is 0.

A similarity also exists in colors among these nodes. The salient object in the image has smaller spatial changes, whereas the background areas distributed over the entire image have higher spatial changes. In other words, tightness is an appropriate supplement to boundary connectivity, which can suppress areas where errors are prominent. The spatial compactness $D$ can be calculated using the intraclass distribution.

$$D(Q_i^c) = \sum_{i=1}^{[Q_i]} \frac{p_i - \mu_i}{\|p_i - \mu_i\|},$$

where $p_i$ is the position of node $v_i$ and $\mu_i$ is the center position of $Q_i^c$. At this moment,

$$\text{BndCon}(Q_i^c) = \frac{\text{Len}_{\text{bnd}}(Q_i^c)}{\sqrt{\text{Area}(Q_i^c)} \cdot D(Q_i^c)}.$$  

Further, the above formula is extended to the background probability, which is based on the boundary connectivity value of node $v_i$: the background probability $A_{i}^{bg} = [a_i^{bg}]_{n_{x} \times 1}$,

$$a_i^{bg} = 1 - \exp\left( -\frac{\text{BndCon}(Q_i^c)}{2\gamma_b^2} \right),$$

where $\gamma_b$ is the adjustment parameter used to control the intensity $A_{i}^{bg}$. Because the background probability evaluates the possibility that a node belongs to the background, possible background nodes can be distinguished by the background probability. Therefore, the background exclusion can be expressed as follows:

$$S_A(i) = \begin{cases} 1, a_i^{bg} \leq \theta_{bg} - \frac{1}{n} \sum_{j=1}^{n} a_j^{bg}, \\ 0, \text{else,} \end{cases}$$

where $\theta_{bg}$ is the parameter that controls the background probability threshold.

3.2. Coarse Saliency Map Generation Based on Color Prior. Kim et al. [25] proposed a linear combination of colors based on high-dimensional color space to create an image saliency map when the salient object position is independent of the node position of the salient object.

The histogram feature is an effective measure of saliency. The $i$-th superpixel histogram feature is measured using the chi-square distance between other superpixel histograms, and it is defined as follows:

$$U_{Hi} = \sum_{j=1}^{N} \sum_{i=1}^{b} \frac{(h_{ik} - h_{jk})^2}{(h_{ik} + h_{jk})^2},$$

where $b$ is the number of histogram bins and each histogram has eight bins.

The global contrast feature $U_{Gi}$ of the $i$-th superpixel is as follows:

$$U_{Gi} = \sum_{j=1}^{N} d(c_i, c_j),$$

where $d(c_i, c_j)$ represents the Euclidean distance between the color value of the $i$-th superpixel node $c_i$ and the color value of the $j$-th superpixel node $c_j$, and the eight color channels of RGB, CIELab, hue, and saturation are used to calculate the color contrast feature. The local color contrast feature $U_{Li}$ is defined as follows:

$$U_{Li} = \sum_{j=1}^{N} \beta_{ij} d(c_i, c_j),$$

where $\beta_{ij}$ is a normalized term. Set $\kappa^2 = 0.25$ according to [16] here.
For texture shape features, superpixel regions, histogram of gradients, and singular value features were used. The histogram of the gradient uses the gradient information of the pixels to quickly provide appearance features. The singular value feature [26] is based on the features of the feature image. It decomposes an image through the weighted summation of multiple feature images, where each weight is a singular value obtained through singular value decomposition. The feature image corresponding to the larger singular value determines the overall outline of the original image, whereas the other smaller singular values describe detailed information. Therefore, for blurred images, the larger singular values occupy a higher weight.

First, the initial graph is divided into $2 \times 2$, $3 \times 3$, $4 \times 4$ regions, and a threshold is applied to each region. Owing to changes in color saturation or illumination, the threshold setting is influenced to a certain extent. Here, Otsu’s multi-level adaptive threshold [27] is used to control the ratio between the foreground, background, and unknown regions. Seven-level thresholds were used for each subarea. After three different threshold saliency maps were fused, we obtained a 21-level saliency map with a local threshold such that the saliency map had better local contrast. Therefore, although the local area may not be the most prominent area globally in the entire image, it can also capture the most prominent area locally. Finally, the coarse saliency map is obtained using the global threshold as follows:

$$S_k(i) = \begin{cases} 1, & \text{if } S(i) \geq 18, \\ 0, & \text{if } S(i) \leq 6, \\ \text{unknown, else}. \end{cases} \quad (11)$$

### 3.3. Coarse Saliency Map Generation Based on Center Prior

Liu et al. [28] proposed a color name method based on the central prior when the salient object is located in the middle of the image. For input images of different sizes, the input image is first fixed to a certain pixel width to obtain the optimal parameters of the structural elements at a certain scale. Subsequently, the color name model is used to provide the im2c function and the mapping matrix and convert scaled image $I$ from the RGB space to the color space $C$. For RGB images with pixel values in the integer interval $[0, 255]$, the im2c function is defined as the RGB color values $(R_{x,y}, G_{x,y}, B_{x,y})$ at the image coordinates $(x, y)$ to calculate the mapping index $\text{Ind}_{x,y}$.

$$\text{Ind}_{x,y} = 1 + \left[ \frac{R_{x,y}}{8} \right] + 32 \times \left[ \frac{G_{x,y}}{8} \right] + 32^2 \times \left[ \frac{B_{x,y}}{8} \right], \quad (12)$$

where $\lfloor \cdot \rfloor$ represents rounding down, and the im2c function assumes the row corresponding to the index $\text{Ind}_{x,y}$ from the mapping matrix to obtain an 11-dimensional vector.

Through the above conversion, the RGB color value $(R_{x,y}, G_{x,y}, B_{x,y})$ of each pixel in $I$ is mapped to an 11-dimensional color name vector, and each element in the color name vector is a $[0, 1]$ floating-point number in the interval, indicating the probability that the pixel in $I$ belongs to each color name, the sum of 11 elements in the vector is equal to 1. Because only one probability map of the color name can be obtained each time the im2c function is called, the Color Name Space(CNS) method calls the im2c function 11 times to obtain 11 color name probability maps. For the convenience of expression, each color probability map becomes a color name channel. The entire color namespace $C$ is composed of $M$ color name channels, $C = \{C_1, C_2, \ldots, C_M\}$, where $M = 11$. To use a set of serialized threshold segmentation to obtain a Boolean graph, each color channel is normalized to an integer interval of $[0, 255]$. Although each channel is a probability map, it essentially reflects the intensity information of the color name; that is, the color name formed by these channels contains the perceptual color characteristics under the linguistic description.

The color name channel is used to generate a Boolean graph as follows: First, take a sampling interval $\delta$ in the interval $[0, 255]$ to obtain a set of serialized segmentation thresholds. Assuming that the total number of segmentation thresholds is $n$, the $j$-th threshold is denoted as $\phi_j$, and each color name channel $C_i \in C$ is segmented as follows using these $n$ thresholds:

$$H_i^j = \text{THRESH}(C_i, \phi_j), \quad (13)$$

where $H_i^j$ is the Boolean graph obtained by the threshold $\phi_j$ segmentation of the color name channel $C_i$, the subscript $i \in \{1, 2, \ldots, M\}$ represents the color name channel number, and the superscript $j \in \{1, 2, \ldots, m\}$ represents the sequence number of the segmentation threshold. The functions of the $\text{THRESH}(\cdot)$ function is the element value $H_i^j(x, y) = 1$ at coordinate $(x, y)$ when $C_i(x, y) \geq \phi_j$; otherwise, $H_i^j(x, y) = 0$. When the area of the salient object has a large color intensity, the serialized segmentation of the normal phase image can obtain the object area. Conversely, if the salient object has a low color intensity, only the normal phase image segmentation will make the object area the background. At this time, a reverse graph $\bar{H}^j$ is required.

$$H_i^j = \text{INVERT}(H_i^j). \quad (14)$$

At this time, $M \times 2 \times m$ Boolean graphs including the positive/negative of $M$ color name channels.

Three morphological processes are performed on the serialized Boolean graph generated above, including closing, hole-filling, and boundary object clear-border, to obtain a saliency map based on the surrounding clues.

After obtaining the saliency maps of all color name channels, these saliency maps are combined in a linear average manner to obtain a more stable saliency map.

$$S_i = \frac{1}{2m} \sum_{j=1}^{n} (S_i^j + \bar{S}_i^j), \quad (15)$$

$$S_C = \frac{1}{M} \sum_{i=1}^{m} S_i,$$

where $S_i^j$ is the positive saliency map of the $i$-th color name channel, $\bar{S}_i^j$ is the inverted saliency map of the $i$-th color.
4. Integration Mechanism of the Saliency Map

Useful coarse saliency maps were fused from the coarse saliency maps in the above three feature methods, and coarse saliency maps with low detection accuracy were eliminated. The initial nodes determined by different prior knowledge have different effects on salient object detection. The feature maps generated using this feature exhibited certain differences. To ensure the selection of effective feature saliency maps when merging at the same level, two feature saliency maps with the most significant similarity are selected for fusion. This also effectively avoids reducing the saliency of salient objects, owing to the low accuracy of a certain feature saliency map.

Therefore, a novel integration mechanism was defined. Three feature saliency maps between the same level (Z1, Z2, and Z3) were selected to fuse the most similar feature saliency maps of the two saliency maps, which effectively avoids the saliency caused by a certain method. The detection accuracy is not high or fails, thereby introducing new errors, and the competing mechanism is shown below.

\[
\min_{i \in 1,2,3} \left\{ \left| S_{Z1}^{Zi} - S_{Z2}^{Zi} \right|^2 \mid \right. \left| S_{Z1}^{Zi} - S_{Z3}^{Zi} \right|^2 \mid \left. \left| S_{Z2}^{Zi} - S_{Z3}^{Zi} \right|^2 \right\},
\]

where \( S_{Z1}^{Zi} \) is the \( Z_i \)-th level \( S_A \) feature saliency map; similarly, \( S_{Z2}^{Zi} \) and \( S_{Z3}^{Zi} \) are the \( Z_i \)-th level \( S_B \) and \( S_C \) feature saliency maps, respectively. The two selected saliency maps are then fused, and the fusion method is as follows:

\[
S_{Zi}^{Z} = \frac{1}{E_1} \left( S_{Z1}^{Zi} \times S_{Z2}^{Zi} + w_z \right),
\]

where \( S_{Z1}^{Zi} \) and \( S_{Z2}^{Zi} \) are the saliency maps selected by the first handsome, \( w_z \) are the weights calculated using \( S_{Z1}^{Zi} \) and \( S_{Z2}^{Zi} \), and \( w_z \) are the normalization terms.

Thus, the feature saliency map was further refined, the saliency maps of the three levels were adjusted to the original image size, and the different levels (\( Z_1, Z_2, \) and \( Z_3 \) ) were merged again to ensure the effectiveness of the feature saliency map, to maximize the enrichment of the boundary information of the feature saliency map, improve the boundary accuracy of the salient object, and select the fusion of the two different levels of the feature saliency map with the largest difference. The direct fusion of the three-layer feature map to reduce the saliency of the boundary and the fusion of two similar feature maps that lose the boundary information of the salient object are both avoided. The fusion method is as follows:

\[
\max \left\{ \left| S_{Z1} - S_{Z2} \right|^2, \left| S_{Z2} - S_{Z3} \right|^2, \left| S_{Z3} - S_{Z1} \right|^2 \right\},
\]

where \( S_{Z1}, S_{Z2}, \) and \( S_{Z3} \) are the feature saliency maps of the \( Z1, Z2, \) and \( Z3 \) levels adjusted to a uniform scale, respectively. After the second screening, the two feature saliency maps are fused as follows:

\[
S_i = \frac{1}{E_i} \left[ \alpha (S_{Z1} \times S_{Z2}) + W_f \right],
\]

where \( S_1 \) and \( S_2 \) are the two saliency maps screened for the second time, \( \alpha \) denote the set parameters, \( W_f \) denote the weights calculated from \( S_1 \) and \( S_2 \), and \( E_i \) denote the normalization terms.

Here, based on the first fusion, two saliency maps with the largest differences are selected such that the boundaries of the salient object are more abundant. The integration of the two feature saliency maps did not simply fuse all the feature saliency maps. For the first time, two similar feature maps at the same level were selected for fusion, which effectively avoided the failure of the single-feature method, which would reduce the salient object after feature fusion, hereby ensuring the validity of the feature saliency map. The second feature map fusion is to select feature saliency maps with significant differences in feature maps at different levels on the first basis, ensuring the accuracy of the saliency object boundary.

The proposed integration mechanism is based on two aspects:

1. When fusing feature maps of the same level (\( Z1, Z2, \) or \( Z3 \) ), to avoid a feature method from failing, select the feature map fusion of the two most similar feature maps, and the two saliency maps with saliency values close to 1-pixel regions, thereby ensuring their saliency.

2. For the fusion of feature saliency maps of different levels, based on the first fusion, attempt to ensure the richness of the boundary features of the salient objects and select the two feature saliency maps with the most significant differences between them; to improve the richness of the boundary features of the salient objects, the feature redundancy is suppressed to a certain extent, and the boundary accuracies of the salient objects are increased.

5. Experimental Results and Analysis

In this section, a series of experiments are performed on the proposed method. To provide a fair experimental evaluation, we first introduce several evaluation indicators and benchmark datasets. Further, we use the proposed framework to optimize other advanced methods to illustrate the robustness of the proposed framework. Finally, we compared the proposed algorithm with seven other saliency detection algorithms using three benchmark datasets.

5.1. Evaluation Metrics and Datasets. In this study, we verified the performance of the proposed algorithm based on four evaluation indicators: Precision-Recall (PR) curve, Receiver Operating Characteristic (ROC) curve, F-measure curve, F-measure score, and Mean Absolute Error (MAE). The PR curve is widely used in most salient object detection methods. The calculation method is as follows: Given a saliency map, we segment the map with a threshold from 0 to 255 and then compare each result with the real situation to
Figure 2: Continued.
Figure 2: Continued.
generate a precision-recall rate curve. As a supplement, we also introduced the F-measure score, which is the comprehensive performance of the precision and recall values and is calculated as follows:

$$F_{\beta} = \frac{(1 + \beta^2) \text{Precision} \times \text{Recall}}{\beta^2 \text{Precision} \times \text{Recall}},$$  \hspace{1cm} (20)

where $\beta^2$ is set to 0.3, according to [29]. The combinations of the PR curve and F-measure score are common in existing studies.

The MAE is a measure of the average difference between the estimated saliency map $S(x, y)$ and ground truth $G(x, y)$.

$$\text{MAE} = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} |S(x, y) - G(x, y)|,$$  \hspace{1cm} (21)

where $W$ and $H$ are the width and height of the given image, respectively.

In this study, all the methods were compared on three benchmark datasets. The ECSSD dataset [30] contains 1000 images and is composed of many complex scenes, such as...
animals, trees, and people. The Image Pair dataset [31] consists of 210 challenging images. Finally, the ASD dataset [32] contains 300 images.

5.2. Comparison with Other Methods. To verify the robustness and effectiveness of the proposed framework, we compared the proposed method with other seven most advanced methods, including Robust Background Detection (RBD) [10], Manifold Ranking (MR) [16], Saliency Filters (SF) [15], Deformed Graph Label (DGL) [33], Geodesic Saliency (GS) [14], CNS [28], and High-Dimensional Color Transform (HDCT) [24].

The quantitative results of all methods are shown in Figure 2. The proposed method in the ECSSD, ASD, and Image pair is a dataset that contains a large number of images, but the performance is better than the other methods. The proposed method achieves the highest F-measure and MAE in all datasets, indicating its superior performance.

Table 1: F-measure of different algorithms.

|       | SF   | RBD  | MR   | DGL  | GS   | CNS  | HDCT | Ours |
|-------|------|------|------|------|------|------|------|------|
| ASD   | 0.808| 0.878| 0.897| 0.869| 0.830| 0.843| 0.864| 0.917|
| ECSSD | 0.492| 0.682| 0.689| 0.684| 0.608| 0.666| 0.692| 0.710|
| Image pair | 0.700| 0.761| 0.483| 0.822| 0.706| 0.790| 0.823| 0.846|

Table 2: MAE of different algorithms.

|       | SF   | RBD  | MR   | DGL  | GS   | CNS  | HDCT | Ours |
|-------|------|------|------|------|------|------|------|------|
| ASD   | 0.131| 0.072| 0.077| 0.085| 0.108| 0.070| 0.084| 0.052|
| ECSSD | 0.219| 0.166| 0.189| 0.181| 0.206| 0.163| 0.170| 0.138|
| Image pair | 0.169| 0.144| 0.347| 0.124| 0.154| 0.118| 0.129| 0.089|

Figure 3: Visual comparison of saliency maps.
complex scenarios. For the information-rich ECSSD dataset, the proposed method achieved the best performance. The results of the PR curve of all methods are shown in Figure 2. As observed, the proposed method achieved the best overall performance on all datasets, particularly on the ECSSD dataset. Among the comparison methods, RBD, MR, SF, DGL, GS, CNS, and HDCT also achieved outstanding results; however, they are still inferior to our method. The three benchmark datasets contained a large number of complex scenarios, and the experimental results demonstrated the superiority of our method in complex scenarios.

From top to bottom: ASD dataset, ECSSD dataset, and Image Pair dataset.

Table 1 shows the F-measure score; that is, the comprehensive performance of the precision and recall values. The proposed algorithm is superior to the above seven methods.

Table 2 compares the MAE, and the proposed algorithm achieved the optimal value. In summary, the proposed algorithm was validated on three benchmark datasets containing a large number of complex scenarios, and the results demonstrate the superiority of our method in complex scenarios.

5.3. Visual Contrast. A visual comparison of the results is presented in Figure 3. It can be observed that the proposed algorithm can simultaneously highlight salient regions and suppress irrelevant background regions, such as the second and fifth rows. For salient objects with rich texture features, the competition mechanism of coarse and fine saliency maps can overcome the failure limit of a certain feature map, such as the sixth row. The proposed method can correctly label salient regions and completely suppress background regions, particularly shadow regions. This is because the proposed graph model not only guarantees that the foreground is not influenced by the surrounding background regions but also enhances the contrast between the foreground and background. The visual comparison results further demonstrate the superiority of the proposed method.

6. Conclusions

In this study, we proposed a novel saliency detection framework. In the first stage, we divide the image into three different levels. On the one hand, the shape and number of nodes generated by images of different scales differ, which enriches the features of salient objects. On the other hand, it compensated for the first feature fusion. The feature saliency map may lose some details. Subsequently, different types of coarse saliency maps were constructed using three different types of prior knowledge features. In the second stage, we fully utilized the competition and complementarity between the feature saliency maps by introducing a novel competition mechanism. In the first fusion, two similarities in the layer were selected to ensure the validity of the selected features. The feature saliency map fusion with the largest degree also ensured the saliency of the salient objects. In the second fusion, the feature maps of different levels were first adjusted to a unified scale, and then two feature saliency maps with a significant difference were selected for fusion, which reduced feature redundancy and ensured the salience of information on the boundaries of salient objects. The proposed framework overcame the limitations of previous graph-based methods in complex scenarios. We compared the proposed method with seven other state-of-the-art methods, including graph- and non-graph-based methods. Furthermore, considering the importance of existing saliency detection methods, the proposed detection method can be applied to other non-graph-based methods in future studies, such as machine learning-based methods, because the competition mechanism can better exploit these feature methods.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 62075241).

References

[1] X. Zhou, Y. Li, B. He, and T. Bai, “GM-PHD-based multitarget visual tracking using entropy distribution and game theory,” IEEE Transactions on Industrial Informatics, vol. 10, no. 2, pp. 1064–1076, 2014.
[2] C. Guo and L. Zhang, “A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression,” IEEE Transactions on Image Processing, vol. 19, no. 1, pp. 185–198, 2010.
[3] R. Cong, J. Lei, H. Fu, Q. Huang, X. Cao, and N. Ling, “HSCS: hierarchical sparsity based co-saliency detection for rgbd images,” IEEE Transactions on Multimedia, vol. 21, no. 7, pp. 1660–1671, 2019.
[4] F. Fanman Meng, H. Hongliang Li, G. Guanghui Liu, and K. N. King Ngi Ngan, “Object co-segmentation based on shortest path algorithm and saliency model,” IEEE Transactions on Multimedia, vol. 14, no. 5, pp. 1429–1441, 2012.
[5] A. Furnari, G. M. Farinella, and S. Battiato, “An experimental analysis of saliency detection with respect to three saliency levels,” Lecture Notes in Computer Science, vol. 8927, pp. 806–821, 2014.
[6] J. Yang and M.-H. Yang, “Top-down visual saliency via joint CRF and dictionary learning,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 3, pp. 576–588, 2017.
[7] X. Wu, Z. Jin, J. Zhou, and X. Ma, “Saliency propagation with perceptual cues and background-excluded seeds,” Journal of Visual Communication and Image Representation, vol. 54, pp. 51–62, 2018.
[8] W. Wei and H. Chen, “Salient object detection based on weighted hypergraph and random walk,” Mathematical
Problems in Engineering, vol. 2020, pp. 1–14, Article ID 2073140, 2020.

[9] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang, “Salient detection via absorbing Markov chain,” in Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 1665–1672, Sydney, NSW, Australia, December 2013.

[10] W. Zhu, L. Jiang, Y. Wei, and J. Sun, “Saliency optimization from robust background detection,” in Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2814–2821, Columbus, OH, USA, June 2014.

[11] C. Li, Y. Yuan, W. Cai, Y. Xia, and D. Feng, “Robust Saliency Detection via Regularized Random Walks Ranking,” IEEE CVPR, in Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2710–2717, IEEE CVPR, Boston, USA, June 2015.

[12] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süssstrunk, “SLIC superpixels compared to state-of-the-art superpixel methods,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 11, pp. 2274–2282, 2012.

[13] 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.

[14] Y. F. Ma and H. J. Zhang, “Contrast-based image attention analysis by using fuzzy growing,” in Proceedings of the Eleventh ACM International Conference on Multimedia, pp. 374–381, Berkeley, CA, USA, November 2003.

[15] Y. Wei, F. Wen, W. Zhu, and J. Sun, “Geodesic saliency using background priors,” in Proceedings of the 2012 Computer Vision - ECCV 2012, vol. 7574, pp. 29–42, Florence, Italy, October 2012.

[16] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, “Saliency filters: contrast based filtering for salient region detection,” in Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 733–740, Providence, RI, USA, June 2012.

[17] C. Yang, L. Zhang, H. Lu, X. Ruan, and M. Yang, “Saliency detection via graph-based manifold ranking,” in Proceedings of the 2013 IEEE conference on computer vision and pattern recognition (CVPR), pp. 3166–3173, Portland, OR, USA, June 2013.

[18] 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, Article ID 103048, 2021.

[19] H. Wang, C. Zhu, J. Shen, Z. Zhang, and X. Shi, “Salient object detection by robust foreground and background seed selection,” Computers & Electrical Engineering, vol. 90, Article ID 106993, 2021.

[20] M. Jian, J. Wang, H. Yu et al., “Visual saliency detection by integrating spatial prior of object with background cues,” Expert Systems with Applications, vol. 168, Article ID 114219, 2021.

[21] M. Al-Azawi, Y. Yang, and H. Istance, “Two-stage salient object identification and segmentation based on irregularity,” Multimedia Tools and Applications, vol. 80, no. 24, pp. 33275–33295, 2021.

[22] S. Chakraborty and P. Mitra, “A dense subgraph based algorithm for compact salient image region detection,” Computer Vision and Image Understanding, vol. 145, pp. 1–14, 2016.

[23] D. Shan, X. Zhang, L. Wang, T. Cao, and C. Zhang, “Saliency detection with a three-stage hierarchical network,” IEEE Intelligent Systems, vol. 36, no. 3, pp. 72–79, 2021.

[24] S. Yan, X. Song, and G. Liu, “Deeper and mixed supervision for salient object detection in automated surface inspection,” Mathematical Problems in Engineering, vol. 2020, pp. 1–12, Article ID 3751053, 2020.

[25] 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, 2016.

[26] H. Andrews and C. Patterson, “Singular value decompositions and digital image processing,” IEEE Transactions on Acoustics, Speech, & Signal Processing, vol. 24, no. 1, pp. 26–53, 1976.

[27] N. Otsu, “A threshold selection method from gray-level histograms,” IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62–66, 1979.

[28] J. Lou, H. Wang, L. Chen et al., “Exploiting color name space for salient object detection,” Multimedia Tools and Applications, vol. 79, no. 15-16, pp. 10873–10897, 2020.

[29] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, “Frequency-tuned salient region detection,” in Proceedings of the IEEE Computer Vision and Pattern Recognition, pp. 1597–1604, Miami, FL, USA, June 2009.

[30] Q. Yan, L. Xu, J. Shi, and J. Jia, “Hierarchical saliency detection,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 1155–1162, Portland, OR, USA, June 2013.

[31] H. Li and K. N. Ngan, “A Co-saliency model of image pairs,” IEEE Transactions on Image Processing, vol. 20, no. 12, pp. 3365–3375, 2011.

[32] T. Liu, J. Sun, N. Zheng, X. Tang, and H. Shum, “Learning to detect A salient object,” in Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8, Minneapolis, MN, USA, June 2007.

[33] X. Wu, X. Ma, J. Zhang, A. Wang, and Z. Jin, “Salient object detection via deformed smoothness constraint,” in Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 2815–2819, Athens, Greece, October 2018.