Information Sciences 618 (2022) 53–71

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Salient detection via the fusion of background-based and multiscale frequency-domain features ☆

Sensen Song [a], Zhenhong Jia [a,*], Jie Yang [b], Nikola Kasabov [c,d]

[a] College of Information Science and Engineering, Key Laboratory of Signal Detection and Processing, Xinjiang Uygur Autonomous Region, Xinjiang University, Urumqi 830046, Xinjiang, China
[b] Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200400, China
[c] Knowledge Engineering and Discovery Research Institute, Auckland University of Technology, Auckland 1020, New Zealand
[d] George Moore Chair in Data Analytics, Ulster University, Maggy BT48 7JL, UK

ARTICLE INFO

Article history:
Received 29 March 2021
Received in revised form 21 October 2022
Accepted 22 October 2022
Available online 3 November 2022

Keywords:
Background feature
Frequency-domain feature
Fusion framework
Salient object detection
Three-dimensional Gaussian convolution kernel

ABSTRACT

Salient object detection is a fundamental problem in image processing and computer vision. Many saliency detection algorithms based on the background and frequency-domain are used to extract salient object clues. However, the former causes the real object to be submerged in the detected object areas, especially in complex or small object scenes. While the latter will lead to the loss of some object information when detecting large objects. To solve these problems and achieve better object detection results, we propose a fusion framework for salient object detection by fusing background and frequency-domain features. The background features of the image are extracted by an improved background model. This model represents the spatial layout of the image area with respect to the image boundaries. Meanwhile, we present a new frequency-domain processing method to obtain multiscale frequency-domain features and mark the saliency of the object at different scales. Within our framework, inspired by human visual attention, we use the idea of a self-attention mechanism to capture the intrinsic relation between background and multiscale frequency-domain features. In addition, this fusion framework provides a three-dimensional Gaussian convolution kernel, which expands two-dimensional local information to three dimensions for feature fusion, thus producing more accurate salient objects. Experiment results demonstrate that the proposed method consistently outperforms eleven state-of-the-art methods on five challenging and complicated datasets in terms of four evaluation metrics.

1. Introduction

Salient object detection is one of the most challenging problems in computer vision. Moreover, it has many applications, such as visual tracking [10,33], image retrieval [46], face recognition [13], and others. For this reason, researchers have invested heavily in improving the ability of object detection systems to obtain precise object information by constructing various effective mathematical models. The algorithms for salient object detection can be categorized into bottom-up models [8,4,41,39,2,36,47,27,29,3] and top-down approaches [14,15,40,38]. The former establishes a mathematical model using

☆ This work was supported by the National Science Foundation of China under Grant 62261053 and Grant U1803261 and the International Science and Technology Cooperation Project of the Ministry of Education of the People's Republic of China (No. 2016–2196).

* Corresponding author.

https://doi.org/10.1016/j.ins.2022.10.103
0020-0255/© 2022 Elsevier Inc. All rights reserved.

the intrinsic relationship of low-level clues, while the latter requires supervised learning with manually labeled ground truth.

Among the existing bottom-up salient object detection methods, background-based models [38,49,45,43] assume that most parts of the main objects are usually near the center of the image, and the image regions along with the boundary are more likely to be the background. The background region can be found by revealing the boundary connections, and the remaining image region is the object. Although the background-based approach improves the accuracy of saliency detection to a certain extent, if the real object is submerged in the detected object areas, the results are likely to be inaccurate, especially in images with complex backgrounds and small objects, as shown in Fig. 1. Therefore, the first problem we must solve is how to effectively extract the complete information about the object from the background features.

To solve the above problem, we need to quickly and accurately locate object areas in the image. Fortunately, we find that the frequency-domain processing algorithm [9,12,24,23] can predict the region of the image object by simulating human visual attention, which provides an idea for us to overcome the shortcomings of the background model. The method assumed that a natural image consists of several salient regions and repeated non-salient patterns. The spikes in the amplitude spectrum correspond to the patterns in the spatial domain. If the spikes in the amplitude spectrum are smoothed on an appropriate scale, the non-salient regions of the image can be suppressed while highlighting the salient regions. Although the frequency-domain method can predict the object's position, some salient information about the larger objects in the image is lost, as shown in Fig. 2. Therefore, we must solve the second problem of obtaining complete object information in the image.

Because the real object is most probably submerged in the background features, the object information lost in the frequency domain features can be compensated by the background features. Moreover, as mentioned above, frequency-domain features can provide location information of the object for background features. Therefore, we believe the object can be effectively detected by fusing background and frequency-domain features. Inspired by the attention mechanism in the brain [49,35,18,17], we propose a novel fusion framework for salient object detection to address the above two problems. Furthermore, we improve the background model by introducing regional similarity into the boundary connection to obtain more suitable background features. Meanwhile, to extract more accurate frequency-domain features, a new frequency-domain processing method is presented by integrating the Gaussian pyramid algorithm with the anisotropy of the filter kernel [21].

In this paper, the main contributions of the proposed method, when compared to previous methods, are:

1. We propose an improved background model to extract the background features of the object. The most significant improvement is the introduction of the background probability, which increases the regional similarity to highlight the object features.
2. A new image frequency-domain processing method is proposed to obtain frequency-domain features, which can be used to better predict the position of the object in the image.
3. A novel fusion framework is presented for salient object detection to successfully overcome the limitations of the selected components and significantly improve the accuracy of salient object detection. The idea of a self-attention mechanism for feature fusion, inspired by the human attention mechanism, is exploited to capture the intrinsic relationship between image features. We also extend the two-dimensional local information to three-dimensional by using a three-dimensional Gaussian convolution kernel.

2. Related works

This paper focuses on the bottom-up methods for salient object detection. In this section, we briefly review previous works from the two algorithms of interest in this paper: the background algorithm and frequency-domain method for salient object detection.

Among the existing bottom-up salient object detection methods, the background-based model provides a new idea for detecting the background to obtain the object. It is highly effective, so it has received widespread attention. For example,

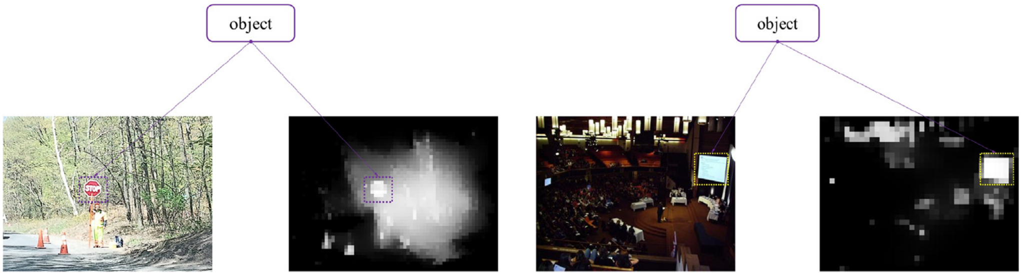

Fig. 1. Saliency maps of the background model in [49].

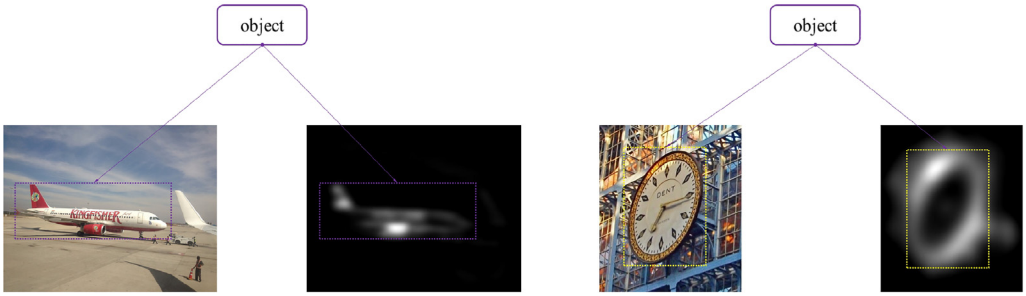

Fig. 2. Saliency maps of the frequency-domain model in [24].

in [37], Wei et al. proposed the concept of background priors (GS) to focus on the nature of the background rather than the foreground and used the region connected to the boundary as the background. This method usually takes super-pixels as calculation units, and the super-pixel connected to image boundaries are marked as background seed points. Then a specific propagation mechanism is used to estimate the saliency value of the unlabeled super-pixels. If the background region has been determined, the region that is not the background can be considered the salient object region. Wang et al. [49] believed that object regions are much less connected to image boundaries than background regions and proposed a robust background measure called boundary connectivity (wCtr), which quantifies how heavily a region is connected to the image boundary. In [45], Yuan Y et al. proposed a saliency regression correction method and a regularized random walk ranking (RCRR) model, which locates and removes foreground super-pixels near the boundary, thereby improving the accuracy and robustness of the saliency estimation based on the boundary priors. Y. Qin et al. [31] constructed a global color distinction and spatial distance matrix based on clustered boundary seeds and integrated them into a background-based map (BSCA). Then, a novel propagation method based on cellular automata was proposed to intuitively exploit the intrinsic relevance of similar regions. Perazzi et al. presented saliency filters (SF) [30], a method for saliency computation based on an image abstraction into structurally representative elements and contrast-based saliency measures, which can be consistently formulated as high dimensional Gaussian filters. In [43], saliency estimation was formulated as a ranking and retrieval problem, and the boundary patches were used as background queries (MR). Liu et al. [28] proposed a salient region detection model, namely, the foreground-center-background model, which was a novel and simple yet efficient method that combines foreground, center, and background saliency.

In addition, calculating the reconstructed residual is also a method of measuring the contrast between the object and background features for salient object detection. For example, Li et al. [22] utilized super-pixels connected to the boundary to construct a background dictionary and project each image block onto the background dictionary to obtain their dense and sparse expression coefficients (RP). A saliency map was obtained by calculating each image block's dense and sparse reconstruction errors. Such a method can also be extended to scenarios where background and foreground seed points are employed for saliency calculations. Tong et al. [34] calculated the contrast using CIEL*a*b* color features, RGB color features, local binary pattern (LBP) features, histogram of oriented gradient (HOG) features, and other underlying visual features of the super-pixel region to obtain a global salient clue map. Also, they constructed background and foreground dictionaries to generate a local salient clue map by calculating the reconstruction residuals. In [19], the authors presented the saliency map of an image as a linear combination of high-dimensional color spaces where the salient regions and backgrounds can be separated. The framework in [42] is guided by the gestalt-laws of perception(GLGO). Furthermore, the authors interpreted the gestalt-laws of homogeneity, similarity, proximity, and figure and ground in link with color and spatial contrast at the level of regions and objects to produce feature contrast maps.

In many cases, salient object detection in the frequency domain has been demonstrated to be simple, fast, and useful. Therefore, increasing attention has been paid to frequency-domain processing. Hou and Zhang proposed the residual spectrum (SR) method [12], which introduced saliency detection to the frequency domain for the first time. Guo et al. [9] postulated that the residual spectrum is not the saliency region of the image. Thus, they discarded all amplitude spectrum information and retained the phase spectrum information. After analyzing the SR model, Li et al. [24] considered that it could only highlight the object's edge but not the sizeable salient object. Therefore, they proposed a saliency model based on the hypercomplex Fourier transform (HFT) to perform low-pass filtering on the amplitude spectrum, smooth the spikes in the amplitude spectrum, suppress the frequent patterns (backgrounds), and highlight large and small salient objects. Although the saliency detection effect for a single object is good, the HFT model easily highlights the most salient object and weakens the remaining secondary objects when the image contains multiple salient objects of different scales. In [9], they presented a method called phase spectrum of quaternion Fourier transform to calculate spatiotemporal saliency maps (PQFT). The phase spectrum of the Fourier transform is the key to obtaining the salient regions' location. Hou, X. et al. [11] showed the image signature as a simple yet powerful descriptor of natural scenes (IS). This signature can be used to approximate the spatial location of a sparse foreground hidden in a sparse spectral background.

The fusion framework proposed in this paper is different from the previous methods in the following aspects.

1. The idea of a self-attention mechanism for feature fusion is exploited to capture the intrinsic relationship between image features. It highlights the common areas of the fusion features while weakening other regions.
2. The three-dimensional Gaussian convolution fusion kernel is the most important innovation of the fusion framework, extending the two-dimensional smoothing filter to a three-dimensional one. It not only can smooth the object's features but also preserve the details of the object.

3. Proposed algorithms

In this section, we present the proposed method from three aspects: First, we improve the background model for background feature extraction; Second, we describe how a new frequency-domain processing approach obtains multiscale frequency-domain features; Third, we introduce the novel fusion framework for detecting the final object saliency map by fusing background and frequency-domain features. The structure of our algorithm is shown in Fig. 3.

3.1. Feature maps

We obtain complementary features through an improved background model and a new frequency-domain processing method, and these features are ready for subsequent processing. We present the algorithms for extracting these features and analyze the characteristics of these features as follows.

3.1.1. Background features

In [49], boundary connectivity is defined as the proportion of each region containing boundary to determine the superpixels belonging to the background, which can obtain background information. The background weight contrast is applied to estimate the salient object, and its equation[49] is defined as follows:

$$wCtr(p) = \sum_{i=1}^{N} d_{app}(p, p_i) w_{spa}(p, p_i) w_i^{bg} \tag{1}$$

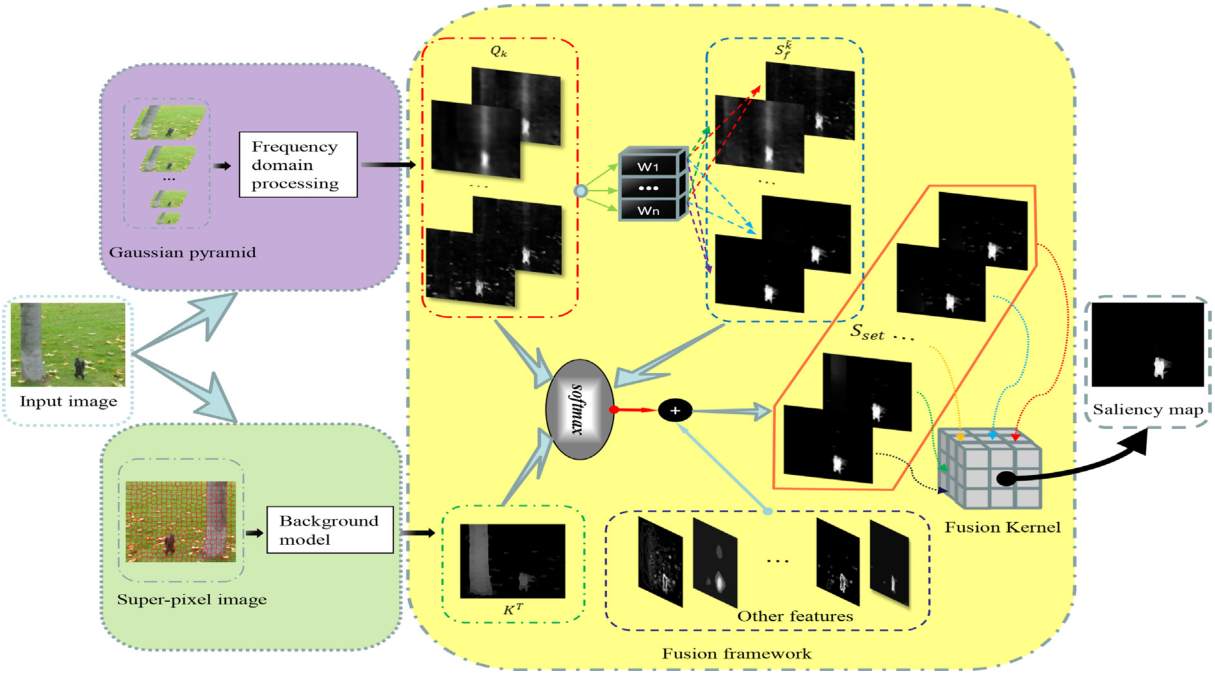

Fig. 3. The proposed fusion framework. First, the image is clustered into a set of nearly regular super-pixels. Second, the background features K^T of the image are extracted from an improved background model. Meanwhile, a new frequency-domain processing method is applied to obtain multiscale frequency-domain features Q_k. Finally, the fusion framework is proposed for salient object detection by integrating the background and frequency-domain features to obtain the object saliency map. $S_f^{\bar{k}}$ is the fusion of all frequency-domain features except for the kth feature. S_{set} is the fusion feature set obtained by Eq. (13). The details of the fusion kernel are shown in Fig. 6.

where N is the number of super-pixels. $w_{spa}(p,p_i) = exp(-\frac{d^2_{spa}(p,p_i)}{2\sigma^2_{spa}})$, $d_{app}(p,p_i)$ is the distance between the center super-pixel p and the ith super-pixel p_i, $d_{spa}(p,p_i)$ is the Euclidean distance between their average colors in the CIE-L*a*b* color space, and $\sigma_{spa} = 0.25$, w_i^{bg} is mapped from the boundary connectivity value of super-pixel p_i, and its definition is

$$w_i^{bg} = 1 - exp(-\frac{BndCon^2(p_i)}{2\sigma^2_{BndCon}}) \tag{2}$$

where σ_{BndCon} is set to 1, and the boundary connectivity is $BndCon(p) = \frac{Len_{bnd}(p)}{\sqrt{Area(p)}}$. $Len_{bnd}(p)$ represents the length of p along the boundary, and $Area(p)$ is the spanning region of each super-pixel p.

To improve the contrast of object features in non-background areas, we propose an improved background model for background feature extraction. The most significant improvement is the difference in the background probability. Inspired by non-local algorithms [16], we introduce a new background probability to highlight the object information and suppress unnecessary image regions. The new background probability is

$$w = e^{\lambda(-\frac{w_i^{bg} \cdot w_{(p,p_i)}}{w_i^{bg} + w_{(p,p_i)}})} \tag{3}$$

where w_i^{bg} is obtained by Eq. (2) and $w_{(p,p_i)} = \frac{\sum_{i\in\Omega} d_{spa}(p,p_i) f(p_i)}{\sum_{i\in\Omega} d_{spa}(p,p_i)}$, since it is a non-local operation, the intrinsic relationship between image regions is integrated into the background model. λ is the adjustment parameter, which is set to 0.6. $f(p_i)$ is the mean value in its neighborhood. The weight w is related not only to the strength of the boundary connection in the background model but also to the intrinsic relationship between image regions, which can highlight the contrast between the object and background features, as shown in Fig. 4. Therefore, the background model can be rewritten as

$$wCtr(p) = \sum_{k=1}^{N} d_{app}(p,p_i) w_{spa}(p,p_i) \cdot w \tag{4}$$

In this paper, the super-pixel image is generated by the simple linear iterative clustering (SLIC) algorithm [20]. The number of super-pixels N is generally set to 250 according to experience.

3.1.2. Frequency-domain features

Our work presents a new image frequency-domain processing method to obtain frequency-domain features. The most significant difference from the previous work is that we apply the Gaussian pyramid in the spatial domain of the image to obtain images with different scales. Furthermore, to preserve the irregular shape information of objects with different scale features, the anisotropic filter [21] kernel is employed to process the input images of different scales.

A Gaussian pyramid is essentially a multiscale representation of an image [25]. It mimics that the near image seen by the human eye is detailed (corresponding to the bottom of the pyramid), and the image seen in the distance is blurry (corresponding to the top of the pyramid). The mathematical expression is as follows:

$$I_{n+1} = D \cdot I_n \tag{5}$$

where $0 \leqslant n \leqslant N, N$ is the number of layers of the Gaussian pyramid, and $N = 5$ is based on experience, and it is discussed later in the Parameter settings subsection. D represents downsampling. I_n is the nth layer image. The multiscale images $f_n(x,y)$ are obtained by a Gaussian filter:

$$f_n(x,y) = I_n * g_\sigma \tag{6}$$

where $g_\sigma = \frac{1}{\sqrt{2\pi}\sigma} e^{(x^2+y^2)/\sigma^2}$ is a Gaussian function with standard deviation σ, and $*$ represents a convolution operation.

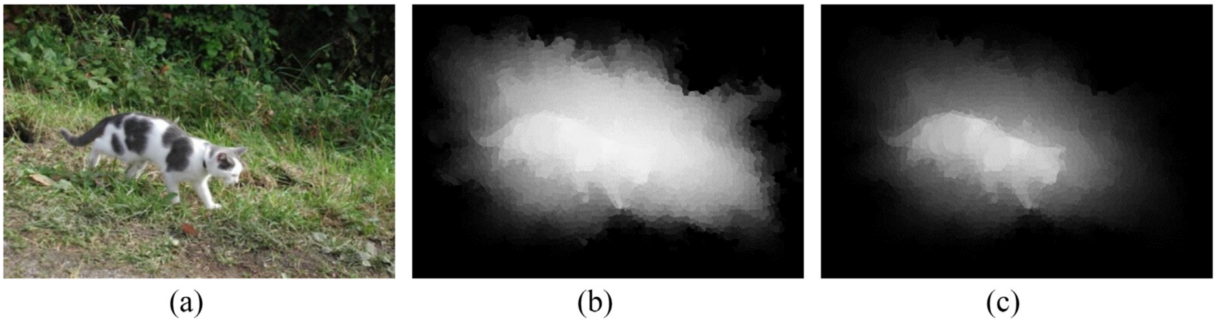

Fig. 4. (a). Input image, (b). Saliency map constructed using the weight w_i^{bg}, (c). Saliency map constructed using the weight w.

We apply the Fourier transform to obtain the frequency-domain representation of Eq. (6):

$$F_n(u,v) = \mathscr{F}[f_n(x,y)] = I_n(u,v)\cdot e^{-2\pi^2\sigma^2(u^2+v^2)} \tag{7}$$

where $\mathscr{F}[\cdot]$ denotes Fourier transform. The proof is given in Appendix **A. Fourier Transform**. According to Eq. (7), the Fourier transform of a Gaussian function is still a Gaussian function. To preserve the irregular shape information of the object, u and v have different coefficients. Therefore, we replace the Gaussian filter in the frequency-domain with anisotropic Gaussian filter, and Eq. (7) can be rewritten as

$$F_n(u,v) = I_n(u,v)\cdot e^{-2\pi^2\sigma_n^2\left(\frac{u^2}{\rho_1}+\frac{v^2}{\rho_2}\right)} \tag{8}$$

where σ is the standard deviation of the Gaussian function, and $\sigma_n = \frac{2^{n-1}}{(n-1)^2}\sigma^2$. ρ_1 and ρ_2 are the anisotropy factors that control the shape of the Gaussian kernel. Due to the different sizes of the salient objects in different images, their characteristics of frequency bands are also different. For this reason, an adaptive parameter strategy is put forward to calculating the anisotropy factors ρ_1 and ρ_2 according to the high-frequency and low-frequency components of the image. ρ_1 and ρ_2 are defined as

$$\rho_1 = \frac{\iint u(1-w(u,v))|F[u,v]|dudv}{\iint u\cdot w(u,v)(|F[u,v]| - \frac{d^2|F[u,v]|}{dv^2})dudv} \tag{9}$$

$$\rho_2 = \frac{\iint v\cdot w(u,v)|F[u,v]|dudv}{\iint v(1-w(u,v))(|F[u,v]| - \frac{d^2|F[u,v]|}{dv^2})dudv} \tag{10}$$

where $w(u,v)$ is the Gaussian low-frequency window. $\iint u(1-w(u,v))|F[u,v]|dudv$ represents high frequency cutoff coefficient of the image, and $\iint v\cdot w(u,v)|F[u,v]|dudv$ is the low frequency cutoff coefficient of the image. Because the low-frequency part of the image shows periodic characteristics due to the amplitude spectrum of the image containing many repeated shapes, and the scale information of the shape itself often appears before the first valley of the amplitude spectrum, so it is necessary to combine the second derivative $(|F[u,v]| - \frac{d^2|F[u,v]|}{dv^2})$ of the amplitude spectrum to calculate its low-frequency cutoff coefficient.

Finally, the frequency-domain features are obtained by

$$S_f^k = [\mathscr{F}^{-1}\{F_n(u,v)\}]^2 \tag{11}$$

where $\mathscr{F}^{-1}\{\cdot\}$ is the symbol of the inverse Fourier transform. k is the dimension of the output features. An example of the frequency-domain feature extraction procedure is shown in Fig. 5.

3.2. Fusion framework

The self-attention mechanism is a classic deep learning model in natural language processing [35] that can reduce the dependence on external information and is good at capturing the internal correlation of data or features. Therefore, the self-attention mechanism can also be applied in image processing to obtain object information by capturing the information of different image features.

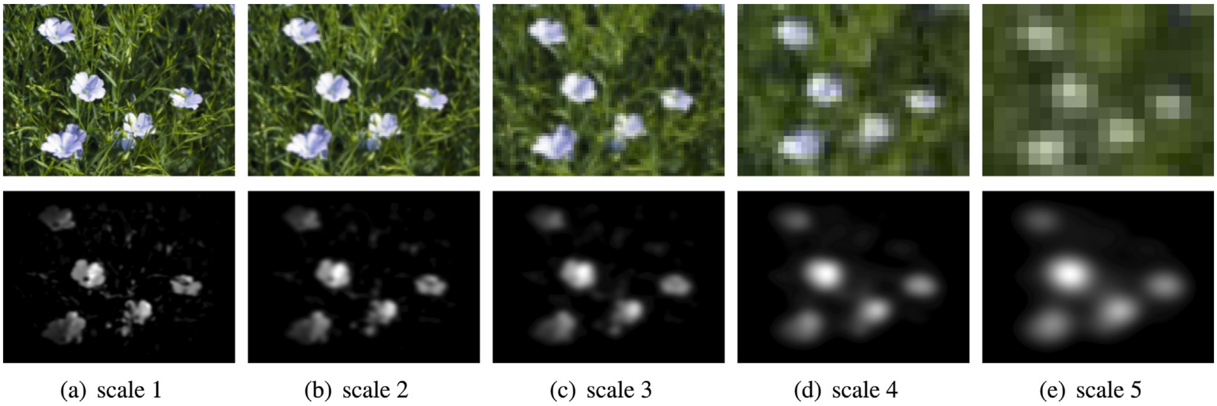
(a) scale 1 (b) scale 2 (c) scale 3 (d) scale 4 (e) scale 5

Fig. 5. Different scale images obtained by Gaussian pyramid algorithm and their frequency-domain saliency maps.

As discussed in [35], the self-attention mechanism provides an effective way to capture the global context information by the key (K) vector, the query (Q) vector, and the value (V) vector.

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{12}$$

where Q, K, and V are input related information, and d_k is the dimension of the vectors.

Inspired by the self-attention mechanism, a novel fusion framework is proposed for salient object detection by fusing frequency-domain and background features. In the first step, to better fuse the multiscale frequency-domain features with the background feature, we present a feature fusion formula to obtain the object features set,

$$S_{set} = \bigcup_{k=1}^{N} \left\{ softmax(\tfrac{Q_k K^T}{\sqrt{d_k}}) \cdot S_f^{\tilde{k}} + \lambda \cdot \|Q_k - S_f^{\tilde{k}}\|_2^2 + \sum_k \|w_k S_f^k - S_f^{\tilde{k}}\|_2^2 \right\} \tag{13}$$

where Q_k is the kth frequency-domain feature, and K^T is the background feature of Eq. (4). N is the same as N in Eq. (5). $S_f^{\tilde{k}}$ is the fusion of all frequency-domain features except for the kth feature. The $softmax(\frac{Q_k K^T}{\sqrt{d_k}}) \cdot S_f^{\tilde{k}}$ term contributes the main information of the object. The role of $\lambda \cdot \|Q_k - S_f^{\tilde{k}}\|_2^2$ is to obtain the detailed complementary information of the feature Q_k by adjusting λ, and $\lambda = 0.1$. $\sum_k \|w_k S_f^k - S_f^{\tilde{k}}\|_2^2$ prevents the loss of salient object information at different scales. $w_k = exp(-(mean(S_f^k) - mean(S_f^{\tilde{k}}))^2)$, which are the weights between frequency-domain features of different scales.

According to Eq. (13), we fuse the background and frequency-domain features into a set of multidimensional salient object features. However, this is not our final result. Therefore, the next step is to extract a more precise object from the salient object feature set. Pointwise convolution can integrate multidimensional information and reduce the dimension of data, which is suitable for object extraction. Nevertheless, because there are significant differences between features of different dimensions in the salient features set, fusing these features into one will result in high contrast among the different areas of the object, which requires us to apply a smoothing filter during the fusion of the features. A Gaussian filter is one of the best smoothing filters, but it is generally applied to two-dimensional images. To this end, we extend the two-dimensional processing to three-dimensional and propose a three-dimensional Gaussian fusion kernel for smoothing and filtering multidimensional features.

In the actual application of image convolution, we assumed that the three components of the three-dimensional convolution kernel are independent, and the variances of the three components are the same. In addition, the center position of the three-dimensional convolution kernel is taken as the origin. The proof is given in Appendix **B. Three-dimensional Gaussian fusion kernel**, and its spatial distribution is shown in Fig. 7. With the above properties, we obtain the three-dimensional Gaussian fusion kernel,

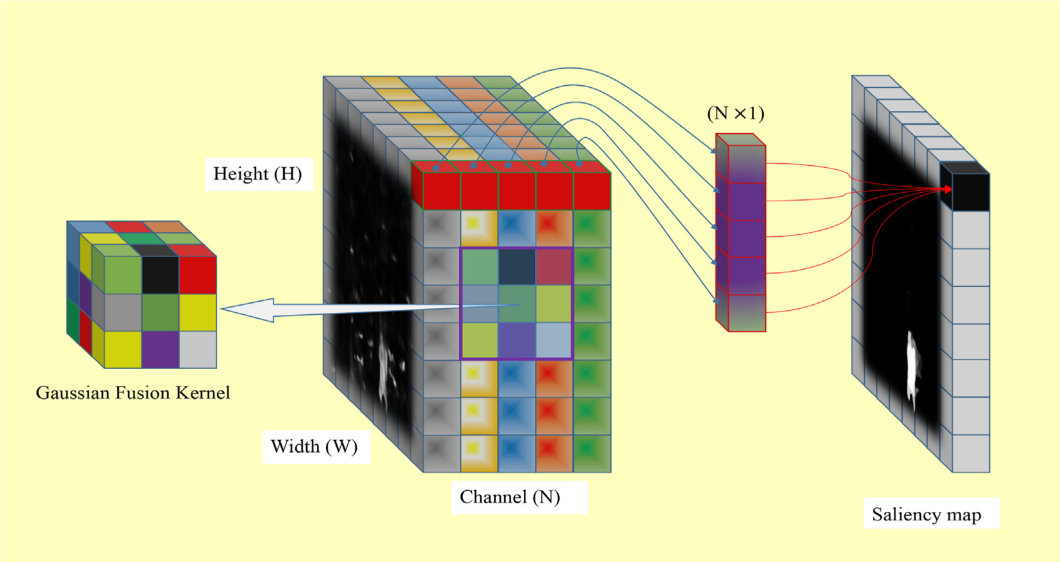

Fig. 6. Multidimensional Gaussian convolution fusion.

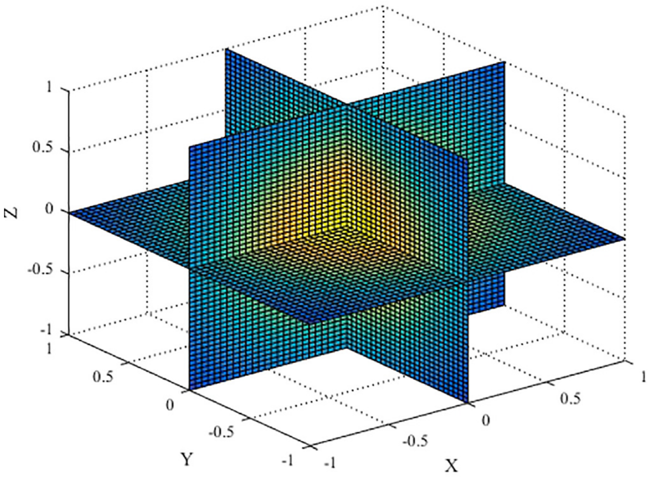

Fig. 7. Distribution of three-dimensional Gaussian fusion kernel.

$$G_f(h,w,n) = \frac{1}{(2\pi)^{\frac{3}{2}}\sigma^3} e^{-(h^2+w^2+n^2)/2\sigma^2} \tag{14}$$

where h, w, and n are the height, width, and dimension of the feature set, respectively. σ^2 is the variance of the Gaussian fusion kernel, and $\sigma = 1.2$.

The pointwise convolution kernel is $1 \times 1 \times N$ convolution. Its value is related to the weight of the input data. The weight is defined as

$$w_m = softmax(exp(-\frac{x_m - mean_N}{std_N})) \tag{15}$$

where x_m is the mth feature value in a channel or dimension. $mean_N$ is the mean of a group feature values across all dimensions. std_N is the variance of a group feature values across all dimensions. Moreover, the larger the deviation, the smaller its weight.

Therefore, the final output can be obtained by

$$S_{out} = \bigcup_{c=1}^{H \cdot W} \left\{ \sum_{m=1}^{N} w_m \cdot [S_{set} * G_f(h,w,n)]_m \right\}_c \tag{16}$$

where H and W represent the height and width of the image. N is the number of dimensions or channels of the feature set. Here the size of $G_f(h,w,n)$ is set to $3 \times 3 \times 3$. The detailed process of integrating multidimensional features is shown in Fig. 6.

In our fusion framework, there are two steps for extracting an object. The first step is the fusion of features, which highlights the common areas of the fusion features while weakening other regions. That is because, considering the characteristics of the fusion features, the common region of features is more likely to be the object. In this step, the object features of different scales are obtained. The second step is to extract a complete and precise object from the multidimensional features generated in the previous step. The pointwise convolution is essentially a weighted summation. The difference is that each group of data weights is different because the convolution kernel adaptively adjusts to the input data. In this step, the most critical portion is the three-dimensional Gaussian convolution fusion kernel, which can effectively smooth the high contrast between features of different scales. This fusion kernel, which extends the two-dimensional smoothing filter to three-dimensional, is the most important innovation of the fusion framework.

4. Experiments

We evaluate the proposed salient object detection method on five image datasets. Qualitative and quantitative comparisons with some state-of-the-art methods are presented, and the results are discussed and analyzed.

4.1. Experiment fundamentals

4.1.1. Benchmark datasets

We use five popular datasets to validate our proposed algorithm: MSRA10K [5], ECSSD [32], PASCAL [6], DUT-OMRON [44], and SOCK [7]. There are 10000 images in MSRA10K and 1000 in ECSSD, all of which contain simple scenes with obvious objects. The PASCAL dataset is derived from the validation set of the PASCAL VOC 2010 segmentation challenge and contains 850 natural images with relatively complex backgrounds. DUT-OMRON and SOCK are composed of images with complex backgrounds, multiple salient objects, and small objects. Moreover, DUT-OMRON contains 5168 images, while there are 1800 images in SOCK obtained from SOC [7] by removing images with no-salient objects. These datasets contain images of various resolutions.

4.1.2. Comparative methods

The proposed method is compared to some of the current state-of-the-art methods, e.g., HFT [24], SF [30], GS [37], MR [43], wCtr [49], BSCA [31], HDCT [19], FCB [28], RCRR [45], GLGO [42], and RP [22].

4.1.3. Evaluation metrics

The precision-recall (P-R) curve, mean absolute error (MAE), area under the curve (AUC), and S-measure [7] are applied to demonstrate the effectiveness of the proposed algorithm. Precision is defined as the percentage of object pixels that are correctly assigned, and recall is defined as the ratio of correctly detected salient pixels to all ground truth pixels. The P-R curve focuses on the proportion of correctly allocated salient pixels but ignores the detection results of non-salient region pixels. The AUC is the area under the receiver operating characteristic curve and is the standard for discerning the advantages and disadvantages of binary models. The MAE is employed to assess the accuracy of salient object detection through the average pixelwise absolute difference between the object saliency map and its ground truth:

$$MAE = \frac{1}{M \times N} \sum_{k=0}^{M \times N} |S(k) - GT(k)| \tag{17}$$

where $S(k)$ is the object saliency map, and $GT(k)$ is the ground truth. M and N are the height and width of the object saliency map.

The above evaluation metrics are all pixelwise similarity comparisons that do not consider the similarity of the image structure. We should consider both the pixel-level comparison of images and the similarity evaluation of image structure, which can make the evaluation more comprehensive and fair. The S-measure applied in this paper is a structural similarity evaluation metric that describes the structural similarity between the saliency map and ground truth. It is defined as follows:

$$S = \alpha * S_o + (1 - \alpha) * S_r \tag{18}$$

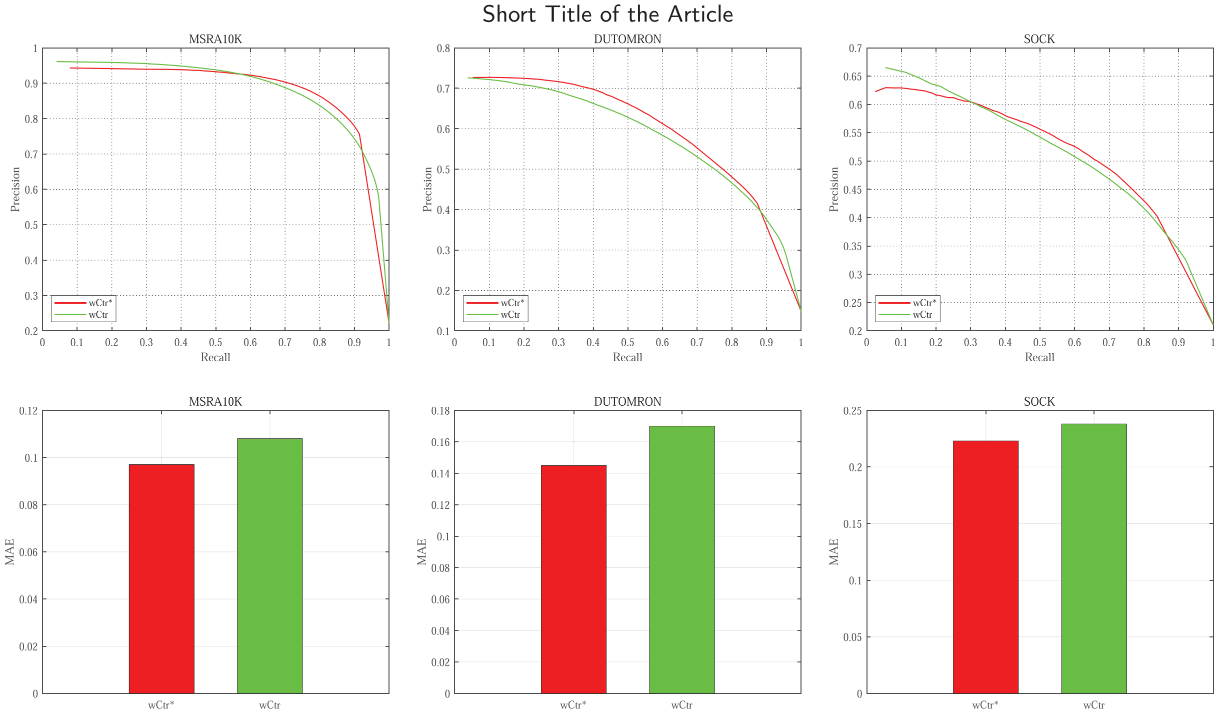

Fig. 8. Comparison of the improved background algorithm with the background model in terms of the P-R curves and MAE on three benchmark datasets. *wCtr* is the standard background method and *wCtr** is the proposed improved background model.

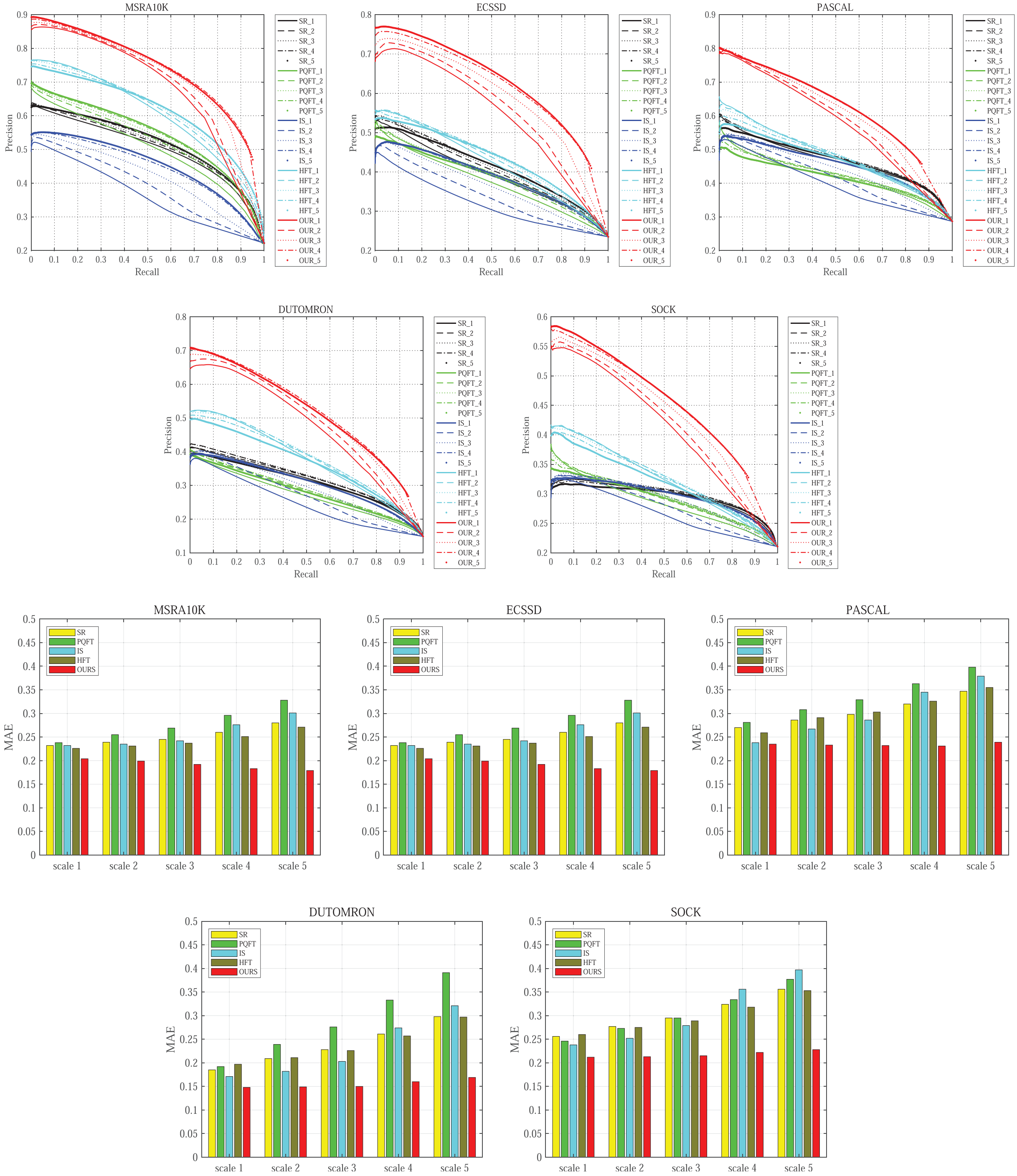

Fig. 9. A quantitative comparison of frequency-domain processing methods in terms of P-R curves and MAE shows that the proposed method is superior to all other methods on all benchmark datasets. The numerical subscripts on the P-R curves correspond to the frequency scales, indicating the performance of the P-R curves at different frequency-domain scales. For example, OUR_1 represents the P-R curve performance of the proposed algorithm at scale 1.

where $\alpha$is within $[0,1]$. S_o denotes the object-aware structural similarity measure, and S_r is the region-aware structural similarity measure.

4.1.4. Parameter settings

In this paper, the results of all the comparison algorithms come from the source code provided by the comparison papers. The parameters are set according to the comparison papers. Furthermore, the parameter settings in this paper are noted under the formulas. For the setting of the two parameters in Eq. (13), λ and k are empirical values we obtained through many experiments. For the parameter k, when k = 1 (scale 1), more detailed information can be obtained, and when k = 5 (scale 5), more redundant information can be retained. In addition, when k = 4 (scale 4) and k = 5 (scale 5), our algorithm performs similarly on all datasets, as shown in Fig. 9, which indicates that it is limited to obtaining new frequency-domain features information when $k > 5$. We can conclude that too many scales fusion will increase the computational complexity, while too few scales will affect detection accuracy. Therefore, through many experiments, we have confirmed that k =5 is the optimal value. The parameters are fixed for all experiments. All experiments are performed on a PC workstation with a 3.6 GHz CPU and 8 GB RAM using MATLAB 2019a.

4.2. Experimental results and analysis

We show the advancement and superiority of our proposed algorithm from the following aspects: 1. The salient features are extracted by the improved background model with less error than the comparison algorithm. 2. The new frequency-domain processing method predicts the object's position more accurately. 3. The proposed fusion framework combines the advantages of the above two features and performs salient object detection more effectively, especially in scenes with complex and small objects.

4.2.1. Improvement of the background feature

We present a new background probability to improve the background model, which can achieve superior performance, especially in complex scenes, as shown in Fig. 8. On MSRA10K, DUT-OMRON, and SOCK, the MAE obtained by the proposed method is lower than that of the previous algorithm. Moreover, our P-R curves have advantages on MSRA10K and DUT-OMRON and are equivalent on SOCK. These comparison results demonstrate the effectiveness of the improved background

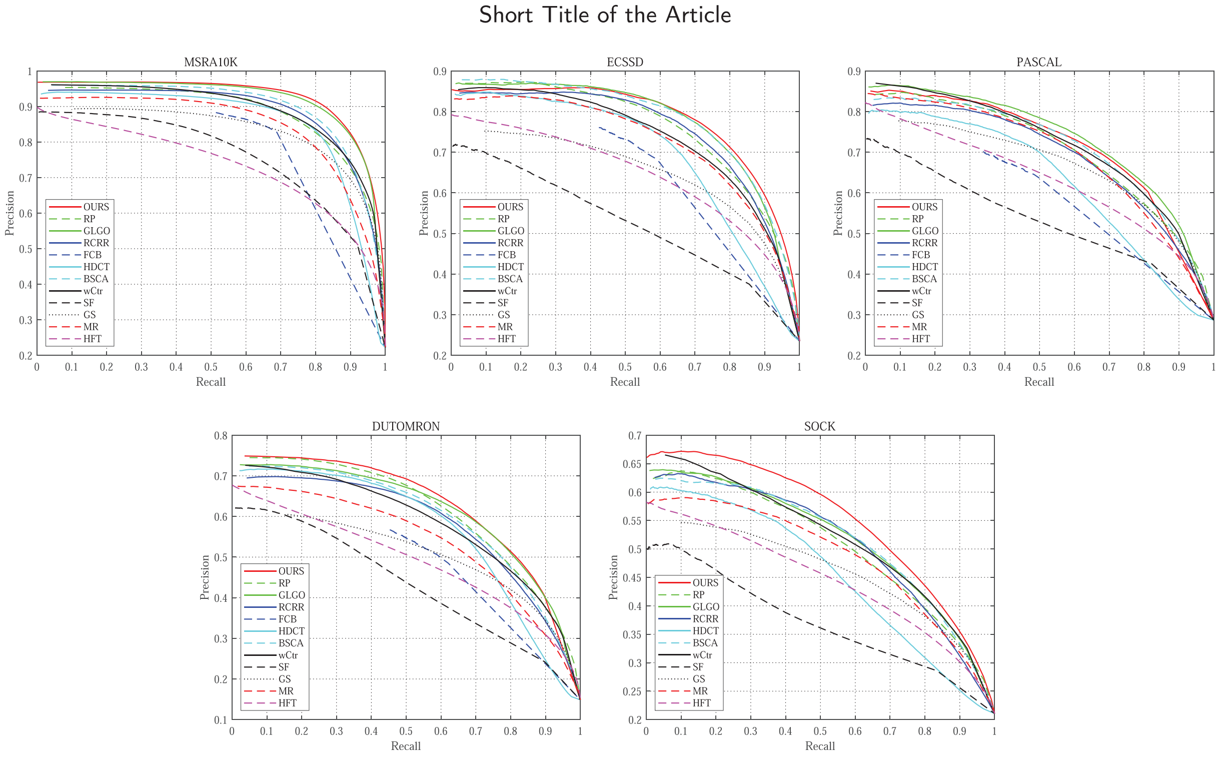

Fig. 10. Comparison of five datasets in terms of the P-R curves.

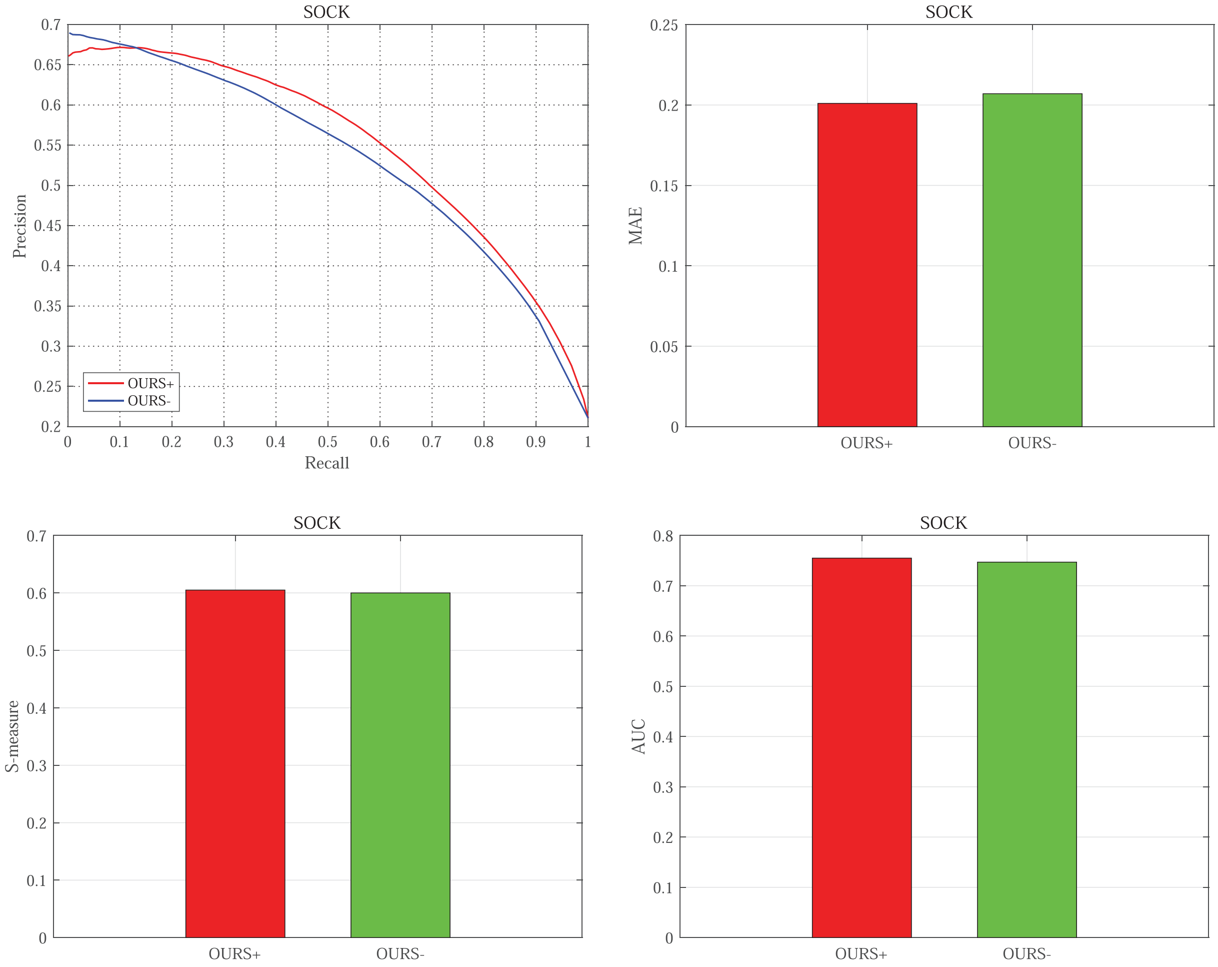

Fig. 11. Comparison of the anisotropic Gaussian filter and the filter without anisotropy. *OURS+* is the anisotropic Gaussian filter; *OURS-* is the Gaussian filter without anisotropy.

model, which can improve the accuracy of background features because background probability w is related to the strength of the boundary connection and the correlation between image regions.

4.2.2. The performance of our frequency-domain processing method

We propose a new image frequency-domain processing method to predict the position of the object in the image. To demonstrate the superiority of our algorithm, we compare the results with some classic algorithms, namely, the SR, PQFT, IS, and HFT models. Moreover, the inputs are images of different scales obtained by the Gaussian pyramid algorithm. A comparison of the results on the multiscale image is shown in Fig. 9.

Fig. 9 shows that the MAEs and P-R curves of our proposed algorithm at different scales are better than those of the comparison algorithms on all datasets. The better P-R curves and MAEs demonstrate that our frequency-domain processing method can significantly improve the accuracy of object location prediction.

Moreover, on the SOCK dataset with complex scenes, the anisotropic Gaussian filter performs better than the filter without anisotropy in terms of the MAE, S-measure, and AUC, as shown in Fig. 11. That proves that the anisotropic Gaussian filter can effectively preserve the irregular shape information of objects, improving object detection accuracy, especially in complex scenes.

4.2.3. Validation of our fusion framework for salient object detection

(1) **Comparison with some fusion strategies** Many fusion strategies exist, such as multilayer perceptron (MLP) [1], SCA (BSCA) [31], and GLGO [42]. SCA is a saliency detection algorithm that employs an asynchronous update mechanism to fuse multilayer cellular automata. GLGO is an integrated framework consisting of bottom-up and top-down attention mechanisms, which can integrate multiple features to produce an object saliency map. MLP is essentially the weighted fusion out-

put of each component. We adopt a Gaussian distribution to replace the training process to obtain the weights in this paper. Here, we apply the three fusion strategies of MLP, SCA, and GLGO to fuse background and frequency-domain features for comparison and conduct experiments on MSRA10K, DUT-OMRON, and SOCK. The results in Fig. 12 show that our fusion algorithm is better than the comparison algorithms in terms of the P-R curves and MAEs, demonstrating that our fusion algorithm can effectively fuse background and frequency-domain features.

(2) **Comparison with state-of-the-art methods** To fully evaluate the validity of our proposed algorithm, we conduct a series of experiments using five benchmark datasets involving various scenarios and four evaluation metrics, including eleven state-of-the-art approaches for comparison. The performance of the proposed algorithm is shown in Fig. 10 and Table 1. Some visual comparisons of the eleven state-of-the-art methods are shown in Fig. 13, Fig. 14, and Fig. 15. The results of our proposed algorithm show that it can overcome the shortcomings of background-based algorithms in scenes with complex backgrounds and small objects and solve the information loss problem in frequency-domain processing algorithms when detecting large objects.

We compare and analyze the results on each dataset to demonstrate the superiority of our algorithm. On MSRA10K, our method outperforms the other methods in terms of the P-R curves and MAEs. For the S-measure and AUC, its performance is still competitive. Our algorithm is not the best for each evaluation metric on ECSSD, but it still has advantages compared to other algorithms, as shown in Fig. 10 and Table 1. Similar to the results on MSRA10K and ECSSD, the performance of our algorithm on PASCAL has no apparent advantages. Although our MAE is superior to all the other algorithms, the P-R curve is not as good as that of GLGO, the S-measure is inferior to that of wCtr, and the AUC performance is unsatisfactory. However, it is worth noting that our approach is superior to the comparison algorithms for all evaluation metrics on the DUT-OMRON and SOCK datasets. On DUT-OMRON, the performance of our proposed algorithm has improved by more than 10% in terms of the MAE. In particular, on SOCK, the performance in terms of the MAE and P-R curves has improved significantly. The above comparative analysis demonstrates the effectiveness and superiority of our algorithm.

(3) **Superiority of the proposed method for detecting objects in scenes with small objects and complex backgrounds** In scenes with complex backgrounds and small objects, the frequency-domain processing algorithm proposed in this paper can effectively predict the salient region. As shown in Fig. 9, both the P-R curves and MAEs demonstrate that our frequency-domain processing method can significantly improve the accuracy of object location prediction. Moreover, the improved background model can achieve superior performance, as shown in Fig. 8. The performance of our proposed algorithm on the four evaluation metrics is shown in Fig.10 and Table 1, which shows that the proposed algorithm performs the best

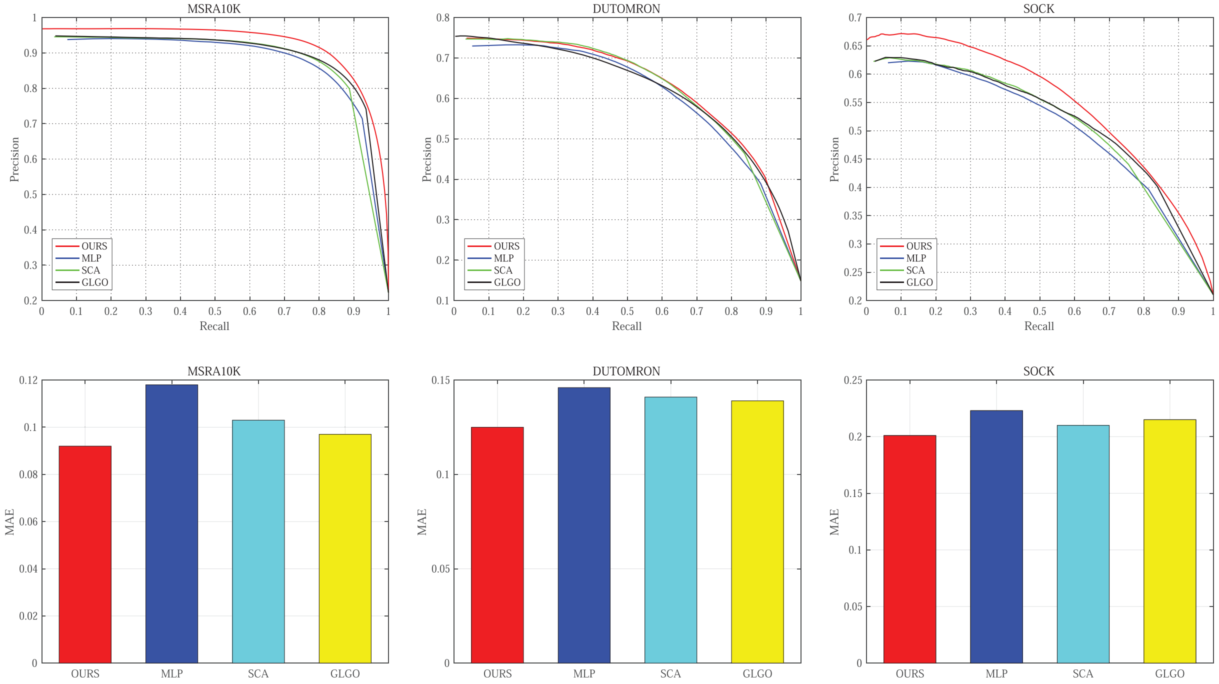

Fig. 12. Comparison among the proposed and known fusion strategies in terms of the P-R curves and MAE illustrating the advantages of the proposed method over the other methods on the benchmark experimental datasets.

Table 1
Comparison on five datasets in terms of the MAE, S-measure and AUC. The best three results are highlighted with red, green and blue fonts, respectively.

Datasets	Evaluation	OURS	RP	GLGO	RCRR	FCB	HDCT	BSCA	wCtr	SF	GS	MR	HFT
MSRA10K	MAE	**0.092**	0.121	**0.096**	0.122	0.121	0.143	0.125	**0.108**	0.179	0.155	0.156	0.200
	S-measure	**0.827**	0.781	**0.838**	0.791	0.719	0.797	0.813	**0.824**	0.553	0.772	0.771	0.634
	AUC	0.843	0.825	**0.847**	0.827	0.734	**0.846**	0.844	0.841	0.699	**0.848**	0.838	0.797
ECSSD	MAE	**0.162**	0.171	**0.158**	0.184	0.173	0.198	0.182	**0.170**	0.210	0.223	0.233	0.240
	S-measure	**0.719**	0.686	**0.724**	0.696	0.620	0.674	**0.725**	0.713	0.452	0.657	0.690	0.591
	AUC	**0.805**	0.785	0.802	0.795	0.684	0.785	**0.815**	0.798	0.603	0.793	**0.810**	0.769
PASCAL	MAE	**0.189**	0.204	**0.196**	0.225	0.213	0.229	0.222	**0.201**	0.237	0.241	0.243	0.252
	S-measure	**0.643**	0.583	**0.637**	0.600	0.515	0.568	0.633	**0.651**	0.388	0.618	0.626	0.521
	AUC	0.747	0.730	0.749	0.731	0.636	0.727	**0.754**	0.753	0.580	**0.759**	**0.760**	0.727
DUT-OMRON	MAE	**0.125**	**0.139**	**0.143**	0.182	0.149	0.164	0.191	0.170	0.151	0.210	0.218	0.197
	S-measure	**0.691**	**0.672**	**0.689**	0.660	0.605	0.656	0.652	0.656	0.513	0.620	0.614	0.583
	AUC	**0.822**	0.802	0.815	0.780	0.697	**0.815**	0.808	0.814	0.640	**0.816**	0.804	0.795
SOCK	MAE	**0.201**	**0.213**	0.217	0.242	-	0.234	0.245	0.238	**0.214**	0.259	0.276	0.251
	S-measure	**0.605**	0.564	**0.596**	0.582	-	0.554	**0.588**	0.587	0.434	0.564	0.569	0.528
	AUC	**0.755**	0.718	0.746	0.722	-	0.718	0.747	**0.754**	0.575	0.750	**0.755**	0.727

on the DUT-OMRON and SOCK datasets with small objects and complex backgrounds. Some visual examples of different saliency object detection methods on images containing small objects and complex backgrounds are shown in Fig. 14 and Fig. 15. These demonstrate the superiority of the proposed algorithm for processing scenes with these characteristics.

4.2.4. Analysis and discussion of our proposed algorithm

Background-based algorithms consider that most image objects are near the center of the image, so the image's border is assumed to be the background. For the region connected to the boundary, the boundary connection's strength is utilized as a condition for judging whether it belongs to the background. The main idea is to detect the object by determining the background. Although good detection results can be achieved, the object is often submerged in redundant information. To this end, we employ a frequency-domain detection algorithm to predict the object's location and extract the object information from the redundant information. The crux is to extract valuable object information from the background and frequency-domain features. Consequently, we propose a fusion framework that includes two main steps. The first step is the fusion of features, which highlights the common areas of the fusion features while weakening other regions. The second step is to extract a complete and precise object from the multidimensional features generated in the previous step. The most critical point of the second step is the novel three-dimensional Gaussian convolution fusion kernel, which can effectively smooth the high contrast of different scales' features. The above experimental results and analysis show that some shortcomings of background-based algorithms and frequency-domain methods can be overcome. Moreover, the proposed algorithm in this paper can effectively improve the accuracy of salient object detection. However, the detection results are not ideal due to the background and frequency-domain cues limitation. Therefore, the salient object detection algorithm based on a bottom-up strategy still needs to be explored and studied.

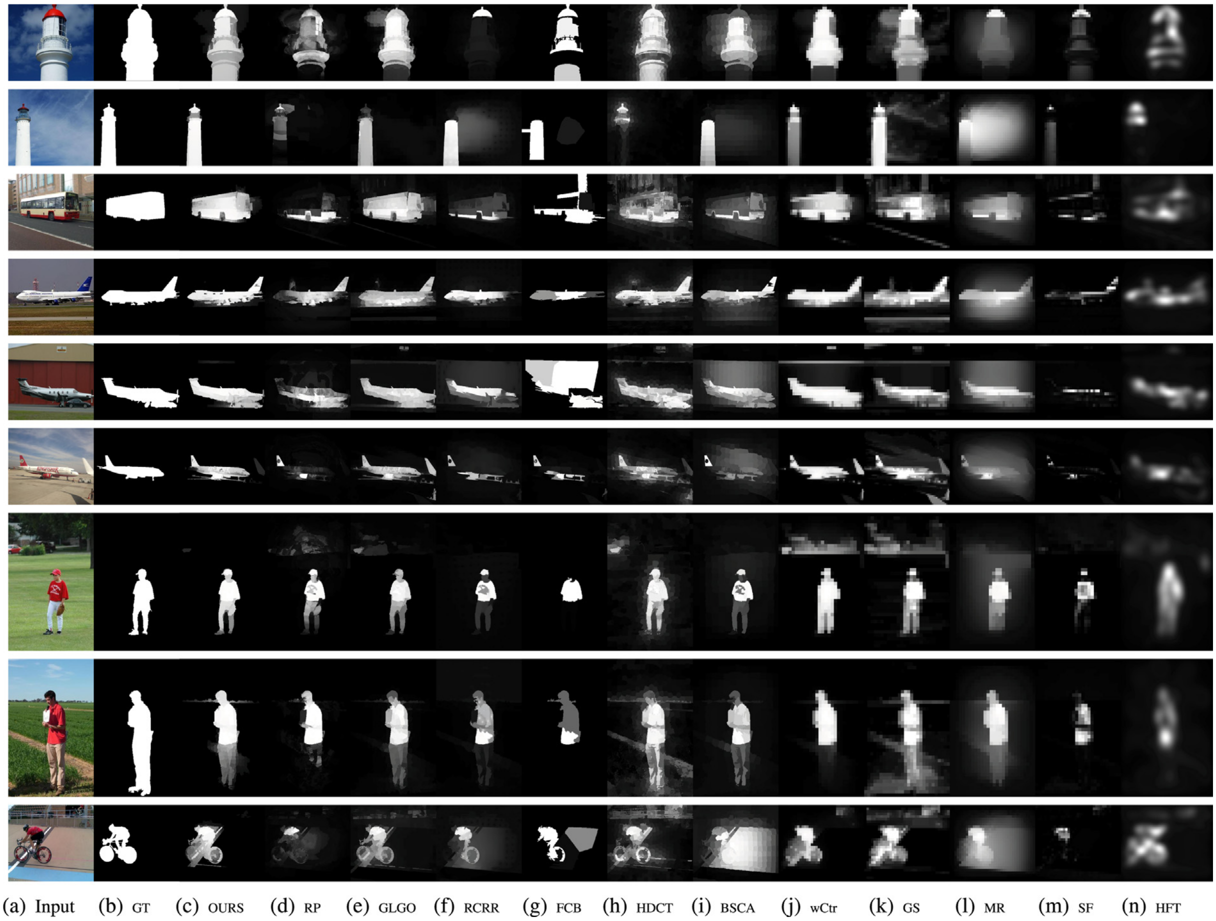

Fig. 13. Visual examples of different salient object detection methods on images with large objects.

5. Conclusions and direction for future research

This paper presents a novel bottom-up algorithm to fuse background and frequency-domain features for salient object detection. We solve the shortcomings of the background model and the frequency domain processing algorithm for salient object detection from feature extraction and the fusion framework. Firstly, for feature extraction, an improved background model is introduced to highlight the object features, which can effectively reduce the redundant information in the object features extracted by a classic background model by adding regional similarity. Meanwhile, A new image frequency-domain processing method is proposed to obtain frequency-domain features, which can better predict the position of the object in the image by replacing the Gaussian filter in the frequency-domain with an anisotropic Gaussian filter. Secondly, the fusion framework can successfully overcome the limitations of the selected features and significantly improve the accuracy of salient object detection. The key is applying a self-attention mechanism for feature fusion, inspired by the human attention mechanism, which is exploited to capture the intrinsic relationship between image features. Moreover, we extend the two-dimensional local information to three-dimensional by using a three-dimensional Gaussian convolution kernel to fuse the local information and spatial information, which can effectively smooth and filter abnormal features. Finally, the experimental results we obtained show the effectiveness of the proposed algorithm for salient object detection.

In future work, we intend to explore how to describe and extract features in images containing different scenes effectively, study the relationships between different features to discover and understand the rules between them, and improve the accuracy of object feature extraction in different scenes. Our other focus will be to use the processing principles for more biologically plausible information in the frequency-domain and modeling attention. We will explore the possibility of using spike representation so that the brighter the pixels are, the earlier spikes they generate to be processed asynchronously in a spiking neural network. It will enable using the massively parallel and low power consumption neuromorphic platforms for real-time, online applications.

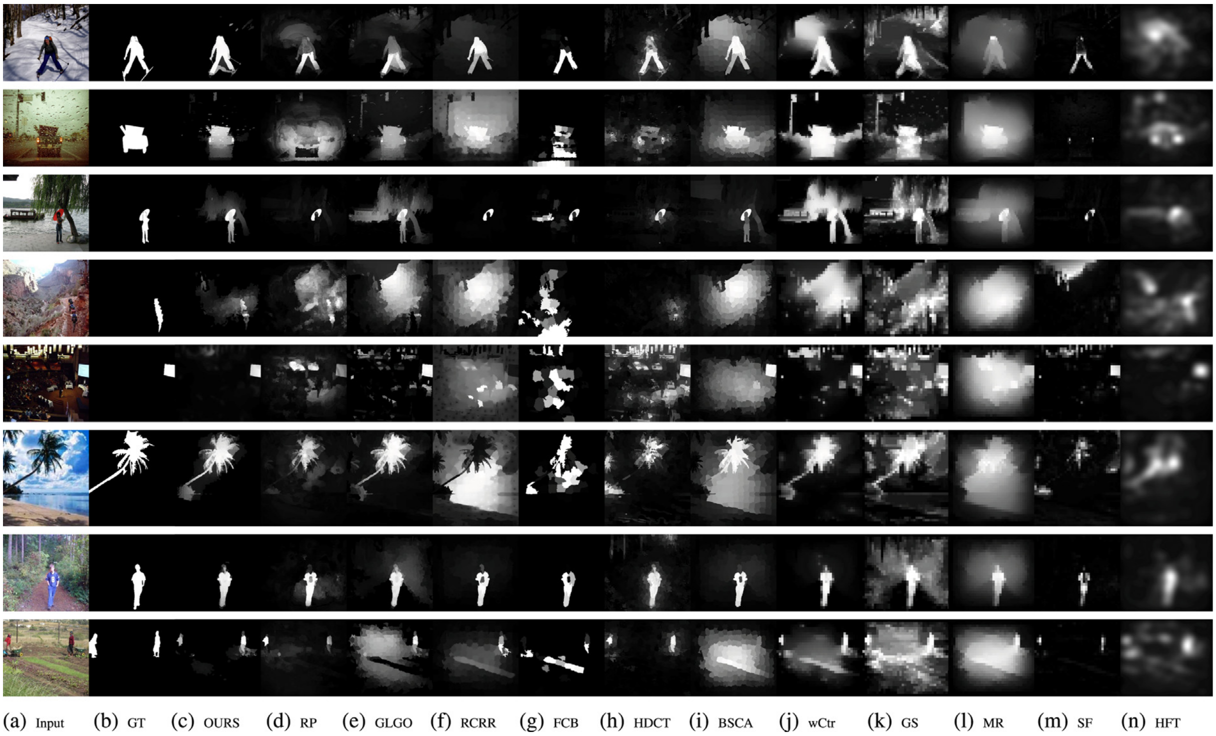

Fig. 14. Visual examples of different salient object detection methods on images with complex backgrounds.

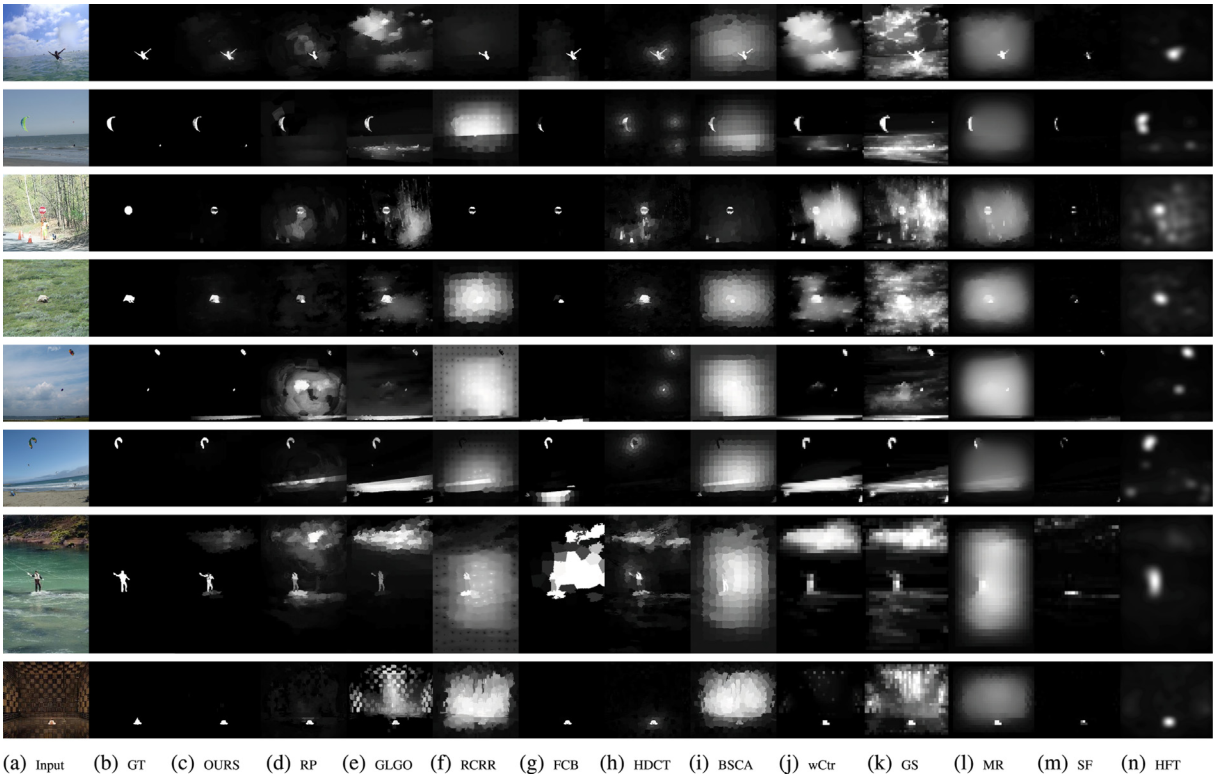

Fig. 15. Visual examples of different salient object detection methods on images with small objects.

CRediT authorship contribution statement

Sensen Song: Conceptualization, Methodology, Software. **Zhenhong Jia:** Validation, Data curation, Writing – original draft, Supervision. **Jie Yang:** Writing – review & editing. **Nikola Kasabov:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the National Science Foundation of China (No. U1803261 and 62261053) and the 0International Science and Technology Cooperation Project of the Ministry of Education of the People's Republic of China (No. 2016–2196).

Appendix A. Fourier transform

The Fourier transform of $f_n(x,y)$ in the Eq. (6) is given by

$$\begin{aligned} F_n(u,v) &= \mathscr{F}[f_n(x,y)] \\ &= \mathscr{F}[I_n \ * g_\sigma] \\ &= I_n(u,v)\cdot\mathscr{F}[\frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2}] \\ &= I_n(u,v)\cdot\int\int_{-\infty}^{+\infty}\frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2}\cdot e^{-i2\pi(ux+vy)}dxdy \\ &= I_n(u,v)\cdot\frac{1}{2\pi\sigma^2}\int_{-\infty}^{+\infty}e^{-x^2/2\sigma^2}\cdot e^{-i2\pi ux}dx\cdot\int_{-\infty}^{+\infty}e^{-y^2/2\sigma^2}\cdot e^{-i2\pi vy}dy \end{aligned}$$

where $\mathscr{F}[\cdot]$ is the symbol of Fourier transform.

Then we have

$$\begin{aligned} &\int_{-\infty}^{+\infty}e^{-x^2/2\sigma^2}\cdot e^{-iwux}dx \\ &= \int_{-\infty}^{+\infty}e^{-(x/\sqrt{2}\sigma+i\sqrt{2}\pi u\sigma x)^2}e^{-2\pi^2\sigma^2u^2}dx \\ &\overset{\theta=\frac{x}{\sqrt{2}\sigma}+i\sqrt{2}\pi\sigma u}{\Rightarrow} e^{-2\pi^2\sigma^2u^2}\cdot\sqrt{2}\sigma\int_{-\infty}^{+\infty}e^{-\theta^2}d\theta \\ &= e^{-2\pi^2\sigma^2u^2}\cdot\sqrt{2\pi}\sigma \end{aligned} \tag{19}$$

Finally, we can obtain

$$F_n(u,v) = I_n(u,v)\cdot e^{-2\pi^2\sigma^2(u^2+v^2)} \tag{20}$$

Appendix B. Three-dimensional Gaussian fusion kernel

In order to make the content of this article self-contained, we give an appendix on the three-dimensional Gaussian fusion kernel as follows. A random variable X is said to be normally distributed with mean μ and variance σ^2 if its probability density function (pdf) is

$$G(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} \tag{21}$$

In [26], a random vector $\bar{x} = [x_1, x_2, \ldots x_n]^T \in R_n$ is Gaussian if its pdf is

$$G(\bar{x}|\bar{\mu},\sigma^2) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}e^{-\frac{1}{2}(\bar{x}-\bar{\mu})^T\Sigma^{-1}(\bar{x}-\bar{\mu})} \tag{22}$$

where $\bar{\mu}$ is the mean vector of the random vector $\bar{x}$, and Σ represents the covariance matrix. It is assumed that $\bar{x}$ is a mutually independent vector of dimension n and that Σ^{-1} exists.

For three-dimensional Gaussian distribution, its variables $\bar{x} = [x_1 \quad x_2 \quad x_3]^T$, its mean vector $\bar{\mu} = [\mu_1 \quad \mu_2 \quad \mu_3]^T$, and its covariances $\bar{\sigma} = [\sigma_1 \quad \sigma_2 \quad \sigma_3]^T$. Moreover, its covariance matrix is

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

Since σ_1, σ_2, and σ_3 are independent of each other, then $\sigma_{12} = \sigma_{13} = \sigma_{21} = \sigma_{23} = \sigma_{31} = \sigma_{32} = 0$. Therefore, Σ can be written as

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}$$

and its inverse is

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_3^2} & 0 & 0 \\ 0 & \frac{1}{\sigma_2^2} & 0 \\ 0 & 0 & \frac{1}{\sigma_1^2} \end{bmatrix}$$

and $|\Sigma| = \sigma_1^2\sigma_2^2\sigma_3^2$, then put them into Eq. (21) to obtain the pdf $G(x_1, x_2, x_3)$ of the three-dimensional Gaussian distribution.

$$\begin{aligned} G(\bar{x}) &= \frac{1}{(2\pi)^{3/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu})^T \Sigma^{-1} (\bar{x}-\bar{\mu})} \\ &= \frac{1}{(2\pi)^{\frac{3}{2}}\sigma_1\sigma_2\sigma_3} exp(-\frac{1}{2}[x_1-\mu_1 \quad x_2-\mu_2 \quad x_3-\mu_3] \cdot \begin{bmatrix} \frac{1}{\sigma_3^2} & 0 & 0 \\ 0 & \frac{1}{\sigma_2^2} & 0 \\ 0 & 0 & \frac{1}{\sigma_1^2} \end{bmatrix} \cdot \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \\ x_3-\mu_3 \end{bmatrix}) \\ &= \frac{1}{(2\pi)^{\frac{3}{2}}\sigma_1\sigma_2\sigma_3} exp(-\frac{1}{2}[x_1-\mu_1 \quad x_2-\mu_2 \quad x_3-\mu_3] \cdot \begin{bmatrix} \frac{x_1-\mu_1}{\sigma_3^2} \\ \frac{x_2-\mu_2}{\sigma_2^2} \\ \frac{x_3-\mu_3}{\sigma_1^2} \end{bmatrix}) \\ &= \frac{1}{(2\pi)^{\frac{3}{2}}\sigma_1\sigma_2\sigma_3} e^{-\frac{1}{2}\left[\frac{(x_1-\mu_1)^2}{\sigma_3^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} + \frac{(x_3-\mu_3)^2}{\sigma_1^2}\right]} \end{aligned}$$

Let $\bar{\mu} = 0$, and $\sigma = \sigma_1 = \sigma_2 = \sigma_3$, then we have

$$G(x_1, x_2, x_3) = \frac{1}{(2\pi)^{\frac{3}{2}}\sigma^3} e^{-(x_1^2+x_2^2+x_3^2)/2\sigma^2} \tag{23}$$

References

[1] G. Calcagno, A. Staiano, G. Fortunato, V. Brescia-Morra, E. Salvatore, R. Liguori, S. Capone, A. Filla, G. Longo, L. Sacchetti, A multilayer perceptron neural network-based approach for the identification of responsiveness to interferon therapy in multiple sclerosis patients, Inf. Sci. 180 (2010) 4153–4163.

[2] K.Y. Chang, T.L. Liu, H.T. Chen, S.H. Lai, Fusing generic objectness and visual saliency for salient object detection, in: 2011 International Conference on Computer Vision, IEEE, 2011, pp. 914–921.

[3] Z. Chen, R. Wang, Z. Zhang, H. Wang, L. Xu, Background–foreground interaction for moving object detection in dynamic scenes, Inf. Sci. 483 (2019) 65–81.

[4] M.M. Cheng, N.J. Mitra, X. Huang, P.H. Torr, S.M. Hu, Global contrast based salient region detection, IEEE Trans. Pattern Anal. Mach. Intell. 37 (2014) 569–582.

[5] M.M. Cheng, N.J. Mitra, X. Huang, P.H.S. Torr, S.M. Hu, Global contrast based salient region detection, IEEE Trans. Pattern Anal. Mach. Intell. 37 (2015) 569–582.

[6] M. Everingham, L. Van Gool, C.K. Williams, J. Winn, A. Zisserman, The pascal visual object classes (voc) challenge, Int. J. Comput. Vis. 88 (2010) 303–338.

[7] D.P. Fan, M.M. Cheng, J.J. Liu, S.H. Gao, Q. Hou, A. Borji, Salient objects in clutter: Bringing salient object detection to the foreground, in: Proceedings of the European conference on computer vision (ECCV), 2018, pp. 186–202.

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

[9] C. Guo, Q. Ma, L. Zhang, Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2008, pp. 1–8.

[10] X. He, C.Y.C. Chen, Learning object-uncertainty policy for visual tracking, Inf. Sci. 582 (2022) 60–72.

[11] X. Hou, J. Harel, C. Koch, Image signature: Highlighting sparse salient regions, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2011) 194–201.

[12] X. Hou, L. Zhang, Saliency detection: A spectral residual approach, in: 2007 IEEE Conference on computer vision and pattern recognition, IEEE, 2007, pp. 1–8.

[13] G. Jeevan, G.C. Zacharias, M.S. Nair, J. Rajan, An empirical study of the impact of masks on face recognition, Pattern Recogn. 122 (2022) 108308.

[14] Y. Ji, H. Zhang, F. Gao, H. Sun, H. Wei, N. Wang, B. Yang, Lgcnet: A local-to-global context-aware feature augmentation network for salient object detection, Inf. Sci. 584 (2022) 399–416.

[15] Y. Ji, H. Zhang, Z. Zhang, M. Liu, Cnn-based encoder-decoder networks for salient object detection: A comprehensive review and recent advances, Inf. Sci. 546 (2021) 835–857.

[16] V. Karnati, M. Uliyar, S. Dey, Fast non-local algorithm for image denoising, in: 2009 16th IEEE International Conference on Image Processing (ICIP), 2009, pp. 3873–3876.

[17] N. Kasabov, N. Scott, E. Tu, S. Marks, N. Sengupta, E. Capecci, M. Othman, M. Doborjeh, N. Murli, R. Hartono, et al, Design methodology and selected applications of evolving spatio-temporal data machines in the neucube neuromorphic framework, Neural Networks 78 (2016) 1–14.

[18] N.K. Kasabov, Time-space, spiking neural networks and brain-inspired artificial intelligence, Springer, 2019.

[19] Kim, J., Han, D., Tai, Y.W., Kim, J., 2014. Salient region detection via high-dimensional color transform, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 883–890.

[20] K.S. Kim, D. Zhang, M.C. Kang, S.J. Ko, Improved simple linear iterative clustering superpixels, in: 2013 IEEE International Symposium on Consumer Electronics (ISCE), IEEE, 2013, pp. 259–260.

[21] B. Kurt, V.V. Nabiyev, K. Turhan, Medical images enhancement by using anisotropic filter and clahe, in: 2012 International Symposium on Innovations in Intelligent Systems and Applications, IEEE, 2012, pp. 1–4.

[22] C. Li, Z. Chen, Q.J. Wu, C. Liu, Saliency object detection: integrating reconstruction and prior, Mach. Vis. Appl. 30 (2019) 397–406.

[23] J. Li, L.Y. Duan, X. Chen, T. Huang, Y. Tian, Finding the secret of image saliency in the frequency domain, IEEE Trans. Pattern Anal. Mach. Intell. 37 (2015) 2428–2440.

[24] J. Li, M.D. Levine, X. An, X. Xu, H. He, Visual saliency based on scale-space analysis in the frequency domain, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2012) 996–1010.

[25] S. Li, Q. Hao, X. Kang, J.A. Benediktsson, Gaussian pyramid based multiscale feature fusion for hyperspectral image classification, IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 11 (2018) 3312–3324.

[26] M.A. Lifshits, Multi-Dimensional Gaussian Distributions, Springer, Netherlands, 1995.

[27] X. Lin, Z.J. Wang, X. Tan, M.E. Fang, N.N. Xiong, L. Ma, Mcch: A novel convex hull prior based solution for saliency detection, Inf. Sci. 485 (2019) 521–539.

[28] G.H. Liu, J.Y. Yang, Exploiting color volume and color difference for salient region detection, IEEE Trans. Image Process. 28 (2018) 6–16.

[29] K. Oh, M. Lee, Y. Lee, S. Kim, Salient object detection using recursive regional feature clustering, Inf. Sci. 387 (2017) 1–18.

[30] F. Perazzi, P. Krähenbühl, Y. Pritch, A. Hornung, Saliency filters: Contrast based filtering for salient region detection, in: 2012 IEEE conference on computer vision and pattern recognition, IEEE, 2012, pp. 733–740.

[31] Y. Qin, H. Lu, Y. Xu, H. Wang, Saliency detection via cellular automata, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 110–119.

[32] J. Shi, Q. Yan, L. Xu, J. Jia, Hierarchical image saliency detection on extended cssd, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2016) 717–729.

[33] A.W. Smeulders, D.M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, M. Shah, Visual tracking: An experimental survey, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2013) 1442–1468.

[34] N. Tong, H. Lu, Y. Zhang, X. Ruan, Salient object detection via global and local cues, Pattern Recogn. 48 (2015) 3258–3267.

[35] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017) 5998–6008.

[36] H. Wei, C. Yang, Q. Yu, Efficient graph-based search for object detection, Inf. Sci. 385–386 (2017) 395–414.

[37] Y. Wei, F. Wen, W. Zhu, J. Sun, Geodesic saliency using background priors, European conference on computer vision, Springer (2012) 29–42.

[38] Y.H. Wu, Y. Liu, L. Zhang, M.M. Cheng, B. Ren, Edn: Salient object detection via extremely-downsampled network, IEEE Trans. Image Process. 31 (2022) 3125–3136.

[39] Y. Xie, H. Lu, M.H. Yang, Bayesian saliency via low and mid level cues, IEEE Trans. Image Process. 22 (2012) 1689–1698.

[40] X. Xu, J. Chen, H. Zhang, G. Han, Sa-dpnet: Structure-aware dual pyramid network for salient object detection, Pattern Recogn. 127 (2022) 108624.

[41] Q. Yan, L. Xu, J. Shi, J. Jia, Hierarchical saliency detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 1155–1162.

[42] Y. Yan, J. Ren, G. Sun, H. Zhao, J. Han, X. Li, S. Marshall, J. Zhan, Unsupervised image saliency detection with gestalt-laws guided optimization and visual attention based refinement, Pattern Recogn. 79 (2018) 65–78.

[43] C. Yang, L. Zhang, H. Lu, X. Ruan, M.H. Yang, Saliency detection via graph-based manifold ranking, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 3166–3173.

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

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

[46] C. Zhang, F. Nie, Z. Wang, R. Wang, X. Li, Fast local representation learning via adaptive anchor graph for image retrieval, Inf. Sci. 578 (2021) 870–886.

[47] Y. Zhang, Z. Mao, J. Li, Q. Tian, Salient region detection for complex background images using integrated features, Inf. Sci. 281 (2014) 586–600. Multimedia Modeling.

[48] Q. Zhou, X. Liu, Q. Wang, Interpretable duplicate question detection models based on attention mechanism, Inf. Sci. 543 (2021) 259–272.

[49] W. Zhu, S. Liang, Y. Wei, J. Sun, Saliency optimization from robust background detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 2814–2821.

Sensen Song received the B.S. degrees from Xinjiang Normal University, Urumqi, China, in 2014, and the M.S. degrees from Department of Information Science and Engineering, Xinjiang University, Urumqi, China. He is currently pursuing the Ph.D. degree with the Department of Information Science and Engineering, Xinjiang University. His research interests are in the area of image Segmentation and image processing.

Zhenhong Jia received the B.S. degree from Beijing Normal University, Beijing, China, in 1985, and the M.S. and Ph.D. degrees from Shanghai Jiao Tong University, Shanghai, China, in 1987 and 1995, respectively. He is currently a Professor with the Autonomous University Key Laboratory of Signal and Information Processing Laboratory, Xinjiang University, China. His research interests include digital image processing, optical information detection, and machine learning.

Jie Yang received the B.S. degrees from Shanghai Jiao Tong University, Shanghai, China, in 1982 and 1985, respectively, and the Ph.D. degree from the Department of Computer Science, Hamburg University, Hamburg, Germany, in 1994. He is currently a Professor with the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University. His major research interests are object detection and recognition, data fusion and data mining, and medical image processing.

Nikola Kasabov (M'93-SM'98-F'10) received MS. degree in computing and electrical engineering and his Ph.D. degree in mathematical sciences from the Technical University of Sofia, Bulgaria, in 1971 and 1975, respectively. He is the Founding Director of KEDRI and Professor of knowledge engineering with the School of Computing and Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand. His major research interests include information science, computational intelligence, neural networks, bioinformatics, neuroinformatics, where he has published more than 650 works.