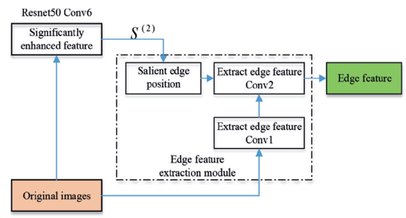
Image salient object detection algorithm based on adaptive multi-feature template



Algoritmo de detección de objetos destacados en imágenes basado en una plantilla adaptativa de múltiples características

Jinping Sun[1,2], Enjie Ding[1*], Bo Sun[3], Lei Chen[2] and Matthew Keith Kerns[4]

[1] School of Information and Control Engineering, China University of Mining and Technology, University Road 1#, Xuzhou 221008, China; IOT Perception Mine Research Center, China University of Mining and Technology, Xuzhou 221000, China, 15150066511, sjp614@163.com
[2] School of Information Engineering(School of Big Data), Xuzhou University of Technology, Fuchun Road 1#, Xuzhou 221008, China
[3] College of Information Science and Engineering, Shandong Agricultural University, Taian 271019, China
[4] G2K Equipment and Services LLC, 2209 Spruce Cir, Mckinney, TX, 75071, United States
[*] Corresponding author, email: sjp@xzit.edu.cn

DOI: http://dx.doi.org/10.6036/9844 | Recibido: 05/06/2020 • Inicio Evaluación: 07/06/2020 • Aceptado: 02/09/2020

RESUMEN

- La detección de objetos destacados en imágenes afectadas por la borrosidad de los bordes del área y la complejidad de la escena tiene varios problemas, como la extracción incompleta de los bordes y las representaciones destacadas borrosas. La fusión de múltiples características destacadas mejora el rendimiento de la detección, pero un algoritmo de fusión inapropiado puede reducir los resultados de la detección. Se propuso un algoritmo de detección de objetos destacados basado en una plantilla adaptativa de múltiples características para resolver la fusión ineficaz de varias características destacadas. En primer lugar, las características (salientes) destacadas de los bordes se obtuvieron utilizando Conv1 y Conv2 en el modelo Resnet50 combinando la información de los bordes locales con la información de localización global de alto nivel. En segundo lugar, aunque algunos atributos como la textura, el contraste de color, los rasgos espaciales y los rasgos de borde destacados se introdujeron en la plantilla multifunción adaptable, estos rasgos se extendieron a todas las capas de los autómatas celulares. La representación final de destacados se obtuvo calculando el histograma del objetivo, el fondo y toda el área de la imagen y generando automáticamente coeficientes de peso de diferentes características según la intersección del histograma. Los resultados muestran que el error absoluto medio (MAE) del algoritmo propuesto es sólo 0,044, mientras que el índice de evaluación global (F-score) llega a 0,899. Así pues, este algoritmo logra una mayor precisión y una mayor tasa de recuperación. La plantilla multifuncional adaptable resuelve eficazmente el problema de la fusión de múltiples características destacadas y puede obtener con precisión las áreas salientes de la imagen. Este estudio proporciona referencias para la segmentación de imágenes, la clasificación de imágenes, el seguimiento de objetos y otros campos de la visión por computadora.
- **Palabras clave:** Detección de objetos destacados, Borde saliente, Autómatas celulares, Plantilla adaptable de múltiples características.

ABSTRACT

Salient object detection affected by area edge blurring and scene complexity has various problems, such as incomplete edge extraction and blurry salient maps. Fusing of multiple salient features improves the detection performance, but an inappropriate fusion algorithm may reduce the results of detection. A salient object detection algorithm based on adaptive multi-feature template was proposed to solve the ineffective fusion of various salient features. First, salient edge features were obtained using Conv1 and Conv2 in the Resnet50 model by combining local edge information with high-level global location information. Second, while some attributes such as texture, color contrast, spatial features, and salient edge features were input into the adaptive multi-feature template, these features were spread to every layer of cellular automata. The final saliency map was obtained by calculating the histogram of the target, background, and entire area of the image and automatically generating weight coefficients of different features according to the intersection of the histogram. Results show that the average absolute error (MAE) of the proposed algorithm is only 0.044, while the comprehensive evaluation index (F-score) reaches 0.899. Thus, this algorithm achieves better accuracy and higher recall rate. The adaptive multi-feature template effectively solves the fusion problem of multiple salient features and can accurately obtain the salient areas of the image. This study provides references for image segmentation, image classification, object tracking, and other fields in computer vision.

Keywords: Salient object detection, Salient edge, Cellular automata, Adaptive multi-Feature template.

1. INTRODUCTION

Visual saliency detection has a natural advantage in processing pictures [1-2]. Visual saliency detection conforms to the habit of browsing pictures and the fast-paced lifestyle of contemporary people. The detection can identify the places on the picture that people are most concerned and use them for subsequent picture processing. Accordingly, the processed pictures will be more valuable and easier to be accepted by people. In early 1985, Koch and Ullman began the initial work in this direction at the laboratory. Itti [3] first proposed a salient object detection model based on the center-surround mechanism as inspired by the human visual attention mechanism. The saliency detection method using different underlying features is often effective only for a specific type of image and cannot be applied to multi-object images in complex scenes. In recent years, automatic learning based on deep learning [4-6] has begun to be applied in image saliency detection to obtain deep features. The detection results of the saliency model based on deep learning on opening datasets are far better that of the traditional algorithms. However, a complex and deep network system is adopted to improve the accuracy of model detection, and this consideration greatly reduces the detection efficiency of the model.

Natural images contain various complex backgrounds. Thus, salient objects are difficult to extract from complex backgrounds with a single low-level feature. Fusing of multiple salient features improves the detection performances. On the basis of these aspects, scholars have conducted works [7-10] on the detection method that combines multi-feature and prior information, but many problems still need to be addressed, such as incomplete edge extraction and blurry salient maps.

Therefore, a salient object detection algorithm based on adaptive multi-feature template is proposed. Salient edge features are extracted by studying the complementarity of high-level salient object information and low-level salient edges. Then, the attributes such as texture, color contrast, spatial features, and salient edge features are sent to the cellular automata network, and weight coefficients of different features are adaptively generated to obtain the final salient image after fusion.

2. STATE OF THE ART

Many studies have been done on saliency target detection. Zhang [11] used sparse cross-layer connection and multi-scale fusion structures to improve the accuracy of model detection, but the network model was not robust. Kang [12] developed a more effective detection algorithm that effectively enhances the object area and eliminates background clutter. However, the edges extracted by the algorithm were incomplete in complex scenarios. Azaza [13] adopted prior information, such as frequency, color, and location, to generate saliency maps and merged a new simple prior saliency detection method. However, the algorithm was not robust. The growing neural gas (GNG) algorithm proposed by Jan [14] calculated the salience of the object using traditional salient feature operations. However, the salience object was incorporated into the background area as a background when it was small. Andrea [15] used an active contour model to automatically extract object contours for detection as an improved edge map feature. However, the edges and textures were not fully fused to produce an effective feature, which caused low accuracy of the algorithm. Qin [16] utilized the cellular automata mechanism (CA) to repair salient values to obtain salient maps according to the update criteria. However, the update criteria cannot be adjusted adaptively, and this inability causes poor detection results in the actual application scenarios. Zhang [17] proposed a salient object detection algorithm that combines depth features with multiple kernel (MK) enhanced learning. The algorithm worked well but runs slowly. Chakraborty [18] used potentially dense sub-images in an image to find visually salient areas in the image, but image detection where salient objects appear at vertices was slightly inadequate. Cui [19] proposed a salience object detection method based on multiple features (MF) that combined multiple features with prior information, which had few deficiencies in dealing with non-salient areas in complex images. Ren [20] studied similar image-driven salience integration and Nikitha [21] combined spatial-temporal characteristics for video salience integration. Both methods achieved good results in specific datasets, but non-salient features cannot be effectively suppressed when detecting images in complex environments. Jenjira [22] proposed an adaptive multi-scale salience detection, and zhang [23] introduced the deep fusion(DF) salient object detection algorithm. However, the algorithm was not robust. Zhao [24] proposed an EGnet model with good results. However, a few of computing resources were wasted because the first layer of information is unutilized. Guo [25] studied the efficiency of salient object detection and xuan [26] used high-level priori semantics to detect salient objects. However, the highly complex model [27-28] affected the efficiency of the algorithm. Wang [29] designed a recursive full convolution network for salient object detection, which could be combined with saliency prior knowledge for more accurate reasoning. However, the edges were incomplete when the network detects prominent objects in blurred images. Tsai [30] designed a deep co-salience detection of super-imposed self-encoding fusion. The algorithm produced good test results on opening datasets but low execution efficiency. Jeong [31] proposed a co-explicit object detection method and prashant [32] designed a new compact end-to-end depth network for moving object detection. The extracted foreground information was doped with insignificant features. Alar [33] proposed a space-time salience estimation model. But the edges detected for dim images were incomplete. Koteswar [34] achieved efficient detection of video salient objects with salient activation trajectories. However, the edges of salient areas extracted by the algorithm are blurred. Li [35] proposed a multi-object salience detection method. This algorithm was more accurate for multi-salient object detection but generated low detection efficiency and detection results with background noise.

Although the abovementioned algorithms can significantly improve the model detection performance, they are affected by the blurring of area edges and the complexity of the scene. The results of their detection also have various problems, such as incomplete edge extraction and blurry salient map. Each feature has certain limitations, and feature fusion can compensate or correct the defect of a single feature to achieve a more salient goal in line with human vision. This study proposes an adaptive multi-feature template. This template propagates the edge locations of the sixth layer salient areas of Resnet50 [6,36] to the side paths of the second layer to suppress the insignificant edges, propagates the edge features from the first layer to the second layer, and obtains the salient edge features from the second layer. These features are sent to an adaptive multi-feature template along with three low-level features, namely, texture, color contrast, and spatial features. The template uses the synchronous updating method of multi-layer cellular automata (MCA). This method distributes the input features throughout each layer of the cellular automata. Then, it calculates the histogram of the object, background, and whole area of the image. Thereafter, the method automatically generates the weight coefficients of different features according to the intersection of histograms. Finally, this method optimizes the weighted fusion of texture, color contrast, spatial features, and salient edge features to obtain the final result that can satisfy the human eyes[37].

The remainder of this study is organized as follows. Section 3 describes the extraction methods of different salient features and constructs an adaptive multi-feature template fusion model. Section 4 verifies the effectiveness and robustness of the algorithm through four aspects: edge extraction algorithm, template evaluation, quantitative analysis, and visual effect. Section 5 summarizes the conclusions.

3. METHODOLOGY

The overall architecture is shown in Fig. 1. In this section, we begin by describing the texture feature extration in section 3.1, then introduce color contrast feature extraction in section 3.2, spatial features extraction in section 3.3, and edge feature extraction module in section 3.4, finally introduce the proposed one-to-one guidance module in section 3.5.

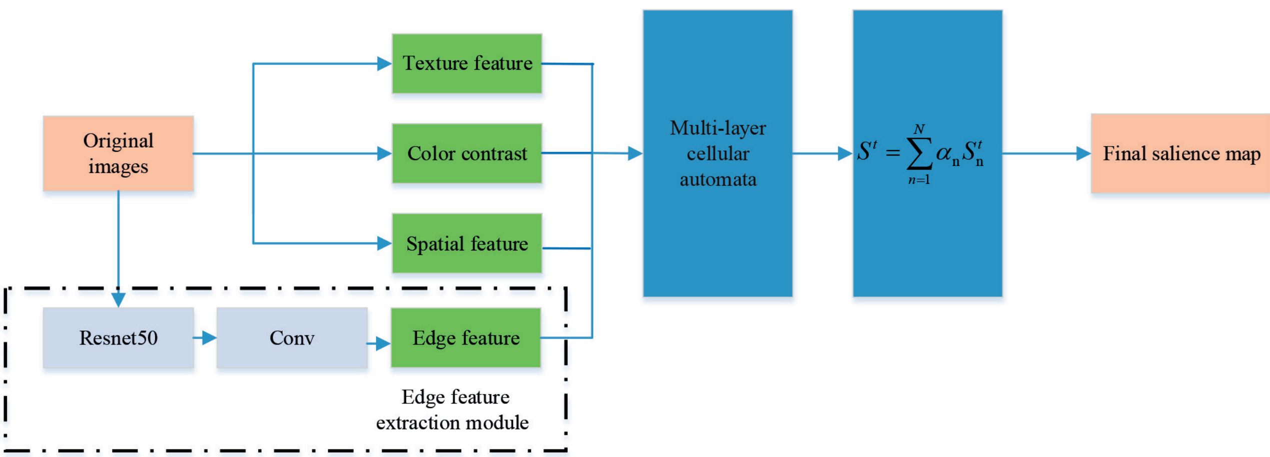


Fig. 1. Algorithm model

3.1. TEXTURE FEATURE EXTRACTION

The calculated texture features may be biased when disturbed by external factors, such as resolution changes or uneven illumination. The original image is pre-processed by a de-sharpening mask, which is equivalent to a low-pass filter, to overcome the aforementioned disadvantage. Then, the difference is computed point by point with the original image, multiplied by a correction factor, and summed with the original image to improve the high-frequency components in the image and enhance the image outline.

$O(x, y)$ is defined as the pixel value after the anti-sharpening mask transformation, which can be expressed as follows:

$$O(x,y) = F_m(x,y) + \frac{i}{\sigma_m(x,y)} \times [m(x,y) - F_m(x,y)], \quad (1)$$

where $m(x,y) - F_m(x,y)$ is the high-frequency part of the local region and i presents he value of contrast gain.

The local variance is $\sigma_m^2(x,y) = \frac{1}{(2n+1)^2}\sum_{k=x-n}^{x+n}\sum_{l=y-n}^{y+n}[m(k,l) - F_m(x,y)]^2$.

The texture feature map of the image is obtained using the LBP operator, that is,

$$S_{ven} = \sum_{p=0}^{P-1} 2^p S(i_p - i_c). \quad (2)$$

In the equation, P represents the p-th pixel in the neighborhood, S represents the symbol function, i_p represents the gray value of the pixel in the neighborhood, i_c represents the gray value of the center pixel, and P is the total number of pixels.

3.2. COLOR CONTRAST FEATURE EXTRACTION

The high-contrast visual signal is an important clue for human vision to locate the object of interest. This metric feature is also important in the salient object detection model. On this basis, the image salience features based on color contrast are proposed. Specifically, the salience of a pixel is equal to the sum of the contrast of the pixel and the color of other pixels in the image. The salience S_{cl} of the pixel I_x is defined as

$$S_{cl} \propto \sum_{\forall I_y \in I} D(I_x, I_y) \bullet Q(p_x, p_y). \quad (3)$$

Among them, I_y represents the other pixel points in the image and $D(I_x, I_y)$ represents the color distance between two pixels. $Q(p_x, p_y)$ is the weight of the distance between two pixel positions. p_x and p_y are divided into positions of pixel points x and y, respectively. The distance from the pixel x and the effect on the color distance differ. The effect is great when the distance is close.

$$Q(p_x, p_y) = \frac{1}{Z}\exp\left(-\frac{\|p_x, p_y\|^2}{2\sigma_p^2}\right), \quad (4)$$

where $Z = \sum_{y=1}^{N} Q(p_x, p_y)$ is a normalization factor. s is the parameter to adjust the distance, which controls the characteristic scale of s_{cl}. When s is small, $Q(p_x, p_y)$ approaches 0 and s_{cl} becomes a local contrast operator. On the contrary, s_{cl} represents a global contrast operator.

Color contrast is related to not only its own color but also the overall color distribution of the image. If the color contrast of the whole image is large, then the contrast between the two colors may be low even if the difference between the two colors is large; if the color contrast of the whole image is small, the contrast between the two colors may be high even if the difference between the two colors is small. Formula (3) can be converted to

$$S_{cl}(I_x) \propto f_x D(C_x, C_y) \bullet Q(p_x, p_y). \quad (5)$$

$$D(C_x, C_y) \propto 1 - \exp\left(-\frac{d(C_x, C_y)}{2\sigma}\right) \quad (6)$$

where $d(C_x, C_y)$ is euclidean distances between two colors and σ are variances of image colors. Pixel (I_x, I_y) has a pixel value of (C_x, C_y), and f_x is the probability density of the pixel C_x value. After the contrast of the image is obtained, the range is normalized to derive a contrast signature between [0,1].

3.3. SPATIAL FEATURES EXTRACTION

In the human visual system, the attention to different spatial positions differs. More attention can be paid to the position closer to the center. The distance between the pixels in different positions in the image and the center of the image satisfies the Gaussian distribution. For any pixel I_x, its spatial eigenvalue $W^I(I_x)$ is calculated using equation (7).

$$\omega^I(I_x) = \exp\left(-\frac{d(I_{x,i}, c)}{\sigma^2}\right), \quad (7)$$

where $I_{x,i}$ is the center coordinate of the pixel I_x and c is the center area of the image. A small average distance of the pixel block from the center of the image corresponds to a large spatial eigenvalue. The salience value S_{ct} of pixel I_x is expressed in equation (8).

$$S_{ct}(I_x) = \omega^c(I_x) \times \omega^I(I_x) \tag{8}$$

3.4. EDGE FEATURE EXTRACTION

ResNet is a residual learning network, which can effectively solve the problem of increasing the error rate when the model is deepened. This study uses the ResNet50 network suggested by other deep learning based methods to describe the proposed algorithm. Only low-level local information is insufficient and high-level semantic information or location information is required to obtain salient edge features. However, the location information of the upper layer is gradually diluted when gradually returning from the top layer to the bottom layer. In addition, the highest layer has the largest receptive field and the most accurate location. In order to analyze the characteristics of different convolution layers of resnet50 network more clearly, the results of each layer are visualized, and the results are shown in Fig. 2(see section: supplementary material). The specific principle is shown in Fig. 3. More salient edge information is diluted when more semantic information is enriched at the higher level. Following EGnet [24], we runcate the last three fully connected layers and connect anotherside path to the last pooling layer in Resnet50.Therefore, a top-down location propagation should be designed. It propagates the edge locations of the sixth layer salient areas to the side paths of the first or second layer to suppress the insignificant edges. Because the Conv1 is too close to the input and the receptive field is too small, the side path $S^{(1)}$ is thrown away. Conv2 preserves better edge information, and we leverage the $S^{(2)}$ to extract the edge features.The edge features of the first layer still need to be extracted, and then is propagated to the second layer. Finally, the salient edge features are obtained from the second layer, as shown in Fig. 3.

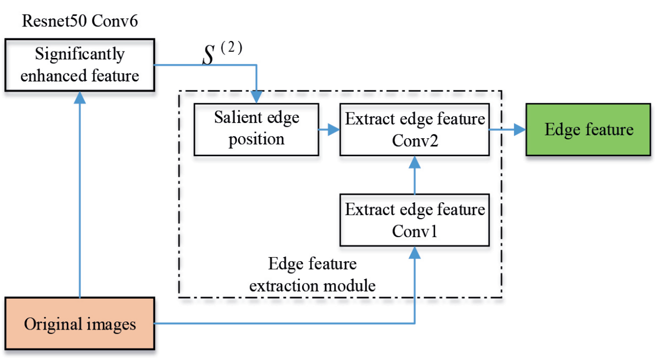


Fig. 3. Edge feature extraction module

The second layer of edge features $S_{edge}^{(2)}$ can be expressed as

$$S_{edge}^{(2)} = S_{edge}^{og\,(2)} + Up(ReLU(Trans(\hat{F}^{(6)};i)); S_{edge}^{og\,(2)}), \tag{9}$$

where $S_{edge}^{(2)}$ represents the original edge feature of the second layer. $Up(*, S_{edge}^{og\,(2)})$ is bilinear interpolation operation which aims to up-sample first parameter * to the same size as the second parameter $S_{edge}^{og\,(2)}$. Rectified linear unit($ReLU$) is a common activation function in artificial neural network. $Trans(\hat{F}^{(6)};i)$ convolves the significantly enhanced feature $\hat{F}^{(6)}$ of the sixth layer according to the parameter i. The purpose of this convolution is to change the number of channels of the feature.

The edge feature extracted from the first layer is denoted by $S_{edge}^{(1)}$, and the edge feature $S_{edge}^{(2)}$ extracted from the second layer is fused to obtain a salient edge feature S_{edge}. This feature can be expressed as

$$S_{edge} = S_{edge}^{(2)} + Up(ReLU(Trans(S_{edge}^{(1)};i)); S_{edge}^{(2)}). \tag{10}$$

3.5. ADAPTIVE MULTI-FEATURE TEMPLATE

Through the salient feature extraction algorithm from Section 3.1 to Section 3.4, the texture, color contrast, spatial features, and salient edge features are obtained, respectively. This study uses the MCA synchronous update method to perform fusion optimization on different salient metrics for further improving the results of salience. A final salience map S with better fusion effect is thus obtained.

$$S^t = \frac{1}{N}\sum_{n=1}^{N} S_n^t \tag{11}$$

In the equation, the N value is 4, which is the number of different scale salient maps. S_n^t iteratively updated by MCA mechanism represents the salient value of all pixels after time t, that is, $\underset{n=1\to4}{S_n^t} \propto [S_{ven}^t, S_{cl}^t, S_{ct}^t, S_{edge}^t]$.

In different scenarios, the contribution of the salient feature map differs. Thus, each feature needs to be weighted, and equation (11) can be changed to

$$S^t = \sum_{n=1}^{N} \alpha_n S_n^t = \alpha_1 S_1^t + \ldots + \alpha_n S_n^t, \tag{12}$$

where $\sum_{n=1}^{N} \alpha_n = 1$, $\alpha_1, \ldots, \alpha_n$ is the weight coefficient of different salient feature maps. When the background of the saliency map changes, the weight coefficient $\alpha_1, \ldots, \alpha_n$ should also be different and can be adaptively changed. The value of each feature coefficient α_n is determined by comparing the histograms of different features, such as texture, color contrast, and edge, in the salient area and the environmental background.

D is set as the target area, B as the background area, H as the whole area.The histograms of the three areas are d_{u_n}, b_{u_n} and h_{u_n}. u_n is the eigenvector of the n-th salient map. H_n is the contribution of the salient feature map n. Then,

$$H_n = \frac{\sum_{u_n=1}^{N} \min(d_{u_n}, h_{u_n}) - \sum_{u_n=1}^{N} \min(d_{u_n}, b_{u_n})}{\sum_{u_n=1}^{N} \min(d_{u_n}, h_{u_n})}. \tag{13}$$

In feature space n, the case in which the foreground intersects the histogram of the entire area is denoted by $\sum_{u_n=1}^{N} \min(d_{u_n}, h_{u_n})$; the case in which the foreground intersects the histogram of the background is denoted by $\sum_{u_n=1}^{N} \min(d_{u_n}, b_{u_n})$.In the n-th feature map, if the distinguishing ability between the foreground and background of the salient target is poor, then the value of H_n is small. Thus, α_n should be small. Conversely, if the distinguishing ability between the foreground and background of the salient target is high, then α_n should be large. The feature weights can be obtained by normalizing the feature contribution.

$$\begin{cases} \alpha_n = \dfrac{H_n}{H_1 + H_2 + \ldots + H_N}, n = 1, 2 \ldots, N \\ \sum_{n=1}^{N} \alpha_n = 1 \end{cases} \tag{14}$$

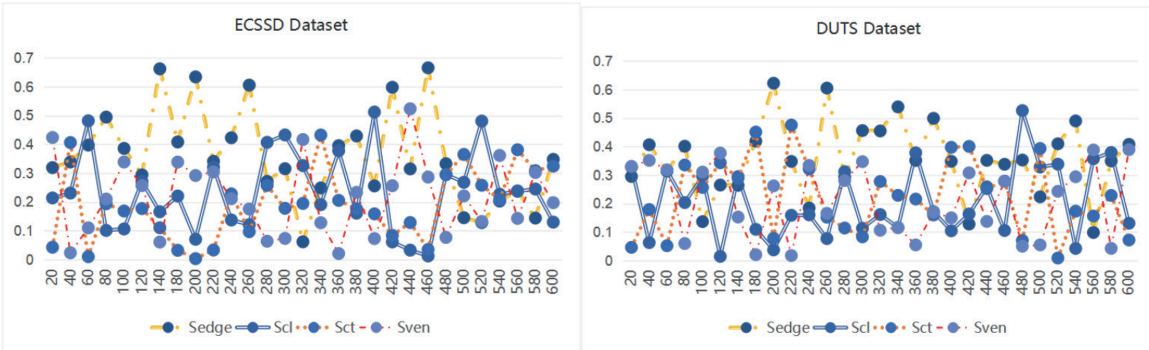


Fig. 4. Distribution of feature contribution

Firstly, the contribution degree of the four salient features of the picture (H_1, H_2, H_3, H_4) are calculated according to formula (13), and then the feature weight of each salient feature $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ is obtained by combining formula (14), and the final saliency map can be obtained by substituting them into formula (12).

4. RESULT ANALYSIS AND DISCUSSION

4.1. IMPLEMENTATION DETAILS AND EVALUATION METRIC

(1) The proposed model is implemented in MATLAB R2018b. The computer configuration is as follows: Intel Core i7-8550u CPU, 2.0GHz main frequency, 8GB memory, windows10 operating system. The hyperparameters are set as followed: learning rate = 1e-5 and loss weight for each side output is equal to 1.

(2) Many classic datasets, such as ECSSD, PASCAL-S, DUT-OMRON, SOD, and DUTS, are available in the direction of visual salience detection. ECSSD[38] contains 1000 meaningful semantic images with various complex scenes. DUTS[39], which includes 10553 images for training and 5019 images for testing, is the largest salient target dataset. Most images are challenging with various locations and scales. In this study, DUTS is selected as the model training dataset, and comparative experiments are performed on the ECSSD dataset.

(3) This study uses precision, recall, and comprehensive evaluation index (F-measure) to analyze objective evaluation indexes. F-measure is a harmonic mean of average precision and average recall, formulated as:

$$F_\beta = \frac{(1+\beta^2)Precision \times Recall}{\beta^2 \times Precision + Recall}, \qquad (15)$$

where β^2 is set 0.3 to weigh precision more than recall as suggested in [24]. Precision denotes the ratio of detected salient pixels in the predicted saliency map. Recall denotes the ratio of detected salient pixels in the ground-truth map [40].

(4) This study uses the mean absolute error (MAE) to further comprehensively evaluate each salience detection algorithm. The calculation formula is

$$MAE = \frac{1}{w \times h}\sum_{i=1}^{w}\sum_{j=1}^{h}\left|S(i,j) - F(i,j)\right|, \qquad (16)$$

where h and w refer to the height and width of the input frame image. $S(i, j)$ represents the saliency map value of point (i, j), and $F(i, j)$ represents the true map value of point (i, j)normalized to [0,1].

4.2. TEMPLATE EVALUATION

Different fusion algorithms are tested on the ECSSD and DUTS classical datasets to verify the effectiveness of the adaptive multi-feature template. The detection results of various fusion algorithms is shown in Table I(see section: supplementary material). Adaptive weighted fusion can automatically adjust the weights according to the characteristics of each image to achieve a better training effect. During the training process, the contribution of each feature coefficient in 600 images is randomly intercepted, as shown in Fig. 4. As observed, LBP texture features and spatial features have similar contribution with some zero contribution degrees in image detection. Color contrast features also show better performance, and the contribution of salient edge is the highest, which indicates that salient edge features are highly robust in various images.

4.3. QUANTITATIVE COMPARISON

The proposed algorithm is combined with the existing six algorithms of GNG [14], CA [16], MK [17], MF[19], DF [23], and EGnet [24] evaluate the performance of various algorithms objectively. Comparative experiments are conducted on ECSSD and DUTS datasets. Table II summarizes the MAE and F-score indicators of six classical algorithms that generate salience maps on the ECSSD and DUTS datasets. The proposed algorithm is superior to other algorithms and is the same as EGnet. However, only the second layer of the convolutional neural network is used to extract significant edges, the algorithm executes faster than EGNET. The processing time of each algorithm is compared, and the comparison results are shown in Fig. 5(see section: supplementary material). In order to be fair, experiments are carried out in the same environment for the model that can obtain the source code. For the model without the source code, the maximum processing efficiency given by the author is selected as the comparative data. This study takes the least time to process the same image, and has obvious advantages in detection efficiency.

Algorithm	ECSSD [38]		DUTS [39]	
	MAE ↓	F-score ↑	MAE ↓	F-score ↑
GNG	0.211	0.858	0.256	0.834
CA	0.082	0.842	0.135	0.832
MK	0.083	0.867	0.1	0.846
MF	0.060	0.893	0.067	0.842
DF	0.079	0.874	0.086	0.851
EGnet	0.044	0.941	0.055	0.854
This study	0.044	0.943	0.054	0.864

Table II. MAE and F-score of different algorithms on ECSSD and DUTS

Fig. 6 shows the comparisons of the precision, recall, and F-score of seven different salient detection algorithms on the EC-SSD and DUTS datasets. Among other compared algorithms, the proposed algorithm has the highest precision, recall and F-score, and the precision and recall are more balanced. The proposed algorithm also has better generalization ability. F-score on complex datasets for the proposed algorithm gives excellent results as well. This algorithm performs better when dealing with more complex datasets, such as DUTS, which contains a large number of images with multiple salient objects.

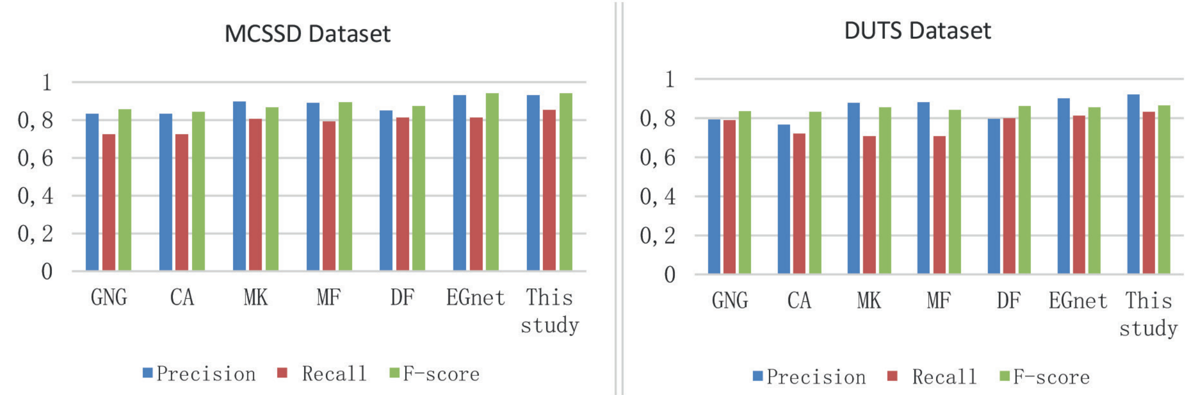


Fig. 6. Precision, recall, and F-score of different algorithms on ECSSD and DUTS

4.4. COMPARISON OF VISUAL EFFECTS

(1) Comparison of visual effects on datasets

The visual effects of the proposed algorithm and the existing six popular algorithms on the MCSSD and DUTS are compared, and the results of a slice of experimental salience tests are shown in Fig. 7. The figure shows that, although the MK, MF, and DF algorithms with good detection effects can completely highlight the salient regions, the texture information and contrast information of the three algorithms are less obvious and even blur. The brightness contrast of the GNG and MCA algorithms is low, and the edge detection of salient areas is incomplete and even presents voids. The proposed algorithm not only visually highlights the salience areas more comprehensively but also suppresses the non-salient areas. The detected salience maps of this algorithm are closer to the annotation map. Therefore, its performance is significantly better than the other comparison algorithms. Compared with the better performance of the deep model EGnet, the detection results of the proposed algorithm are more accurate for the edge processing of salient targets and are more similar to the manual annotation result GT. Results show that the proposed algorithm has a better performance comparing with other compared algorithms in detecting the most significant objects. Other contrast algorithms have some disadvantages, such as fuzzy edges, incomplete edge detection, and even holes.

(2) Comparison of visual effects in practical application scenarios

The proposed algorithm is deployed in the monitoring system in use and captures images under different camera scenes to bet-

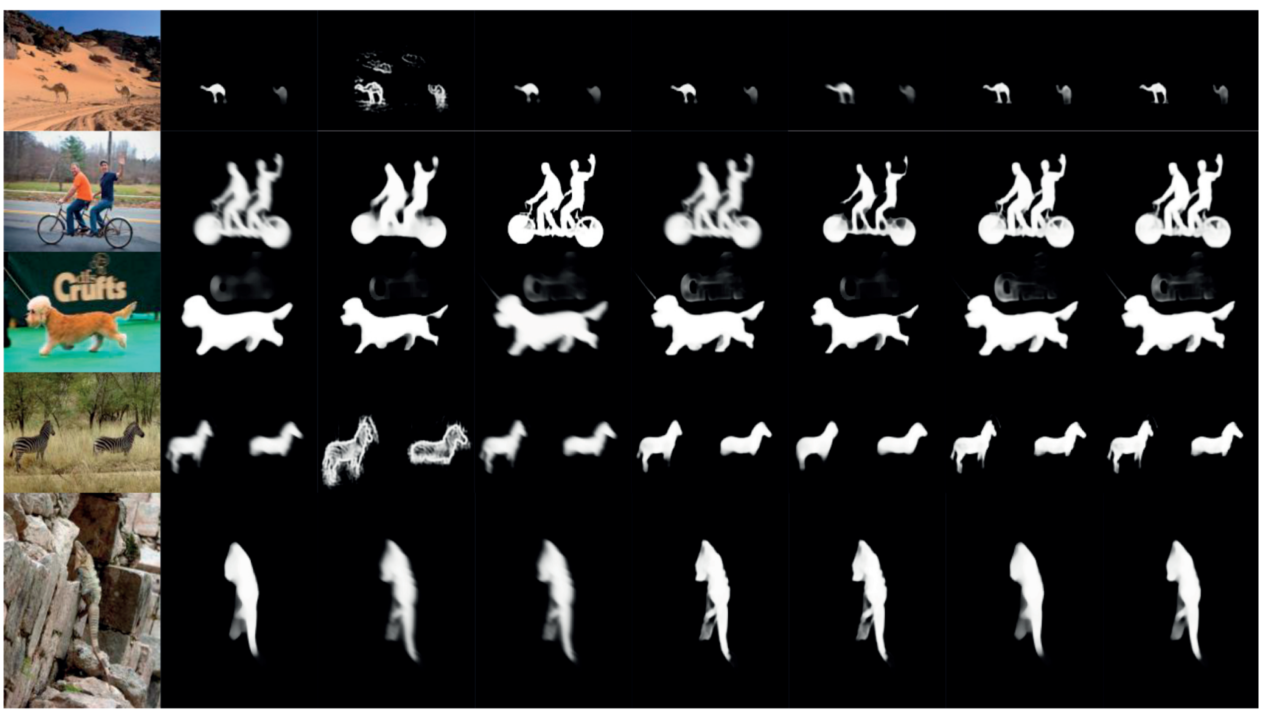


Image GNG CA MK MF DF EGnet This study

Fig. 7. Visual effect comparison of different algorithms

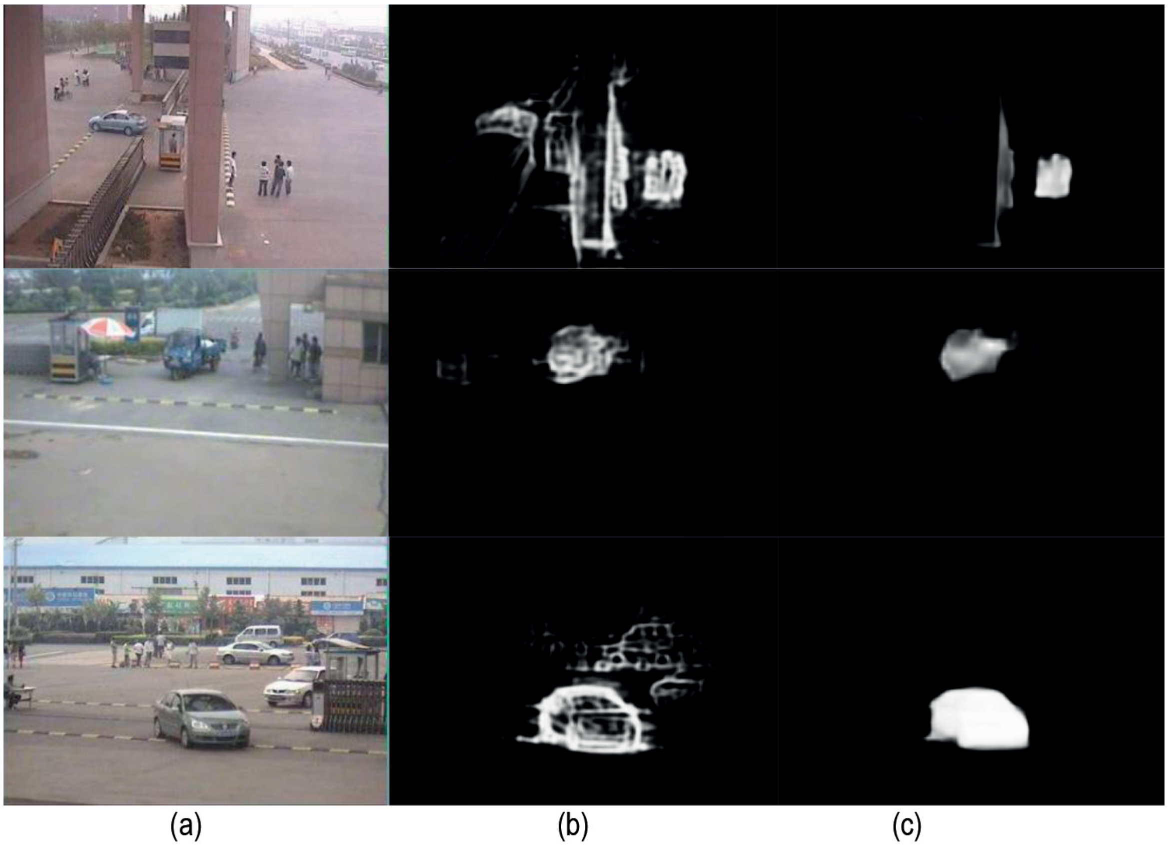


Fig. 8. Salient target detection results in practical application scenarios.(a) Original image. (b) Salient edge. (c)Results

ter test its robustness. Fig. 8 indicates the image captured by the surveillance cameras on the south and north gates of a university campus. Compared with the images in the dataset, the background is complex, the visibility is relatively low, and the noise interference is relatively large. Fig. 8(b) shows the extracted salient edge features. As observed, the outline is relatively clear, and the non-salient edge features in quite a few scenes are marked. Fig. 8(c) is obtained after the adaptive multi-feature template. The map indicates that the background area can be well excluded, and a salient target with consistent highlight can be obtained.

5. CONCLUSION

Multiple saliency features could not be effectively merged. Thus, an MCA synchronous update method, which started from extracting four low-level salience maps, such as texture, color contrast, spatial features, and salient edge features, was used to merge the designed adaptive multi-feature template for overcoming the abovementioned problem. The following conclusions could be drawn:

(1) The edge location of the salient region of the sixth layer through Resnet50 propagates to the second layer where salient edge is produced. The edge feature can eliminate too much background texture details and obtain better continuity of edge information.

(2) The salient edge feature has a greater contribution than the three other low-level salience maps in the adaptive multi-feature template, and the feature is more robust in practical application scenarios.

(3) The adaptive multi-feature template can adaptively optimize the texture, color contrast, spatial features, and salient edge features. Thus, the proposed template can obtain a salience map that meets the human visual requirements.

The proposed salience object detection algorithm has a certain reference to the fields of image segmentation, image classification, and object tracking in computer vision. The proposed algorithm lacks massive training in practical application scenarios. Thus, this algorithm has deficiencies in the detection of salient objects in real complex environments. In following study, dynamic convolution will be presented, a new design that increases model complexity without increasing the network depth or width to improve the efficiency of salient object detection.

REFERENCES

[1] Hajdek K, Bradic P, Hajdek M, et al. "Color Perception of the Observer with the Manifestation of the Chromatic Effect of Crispening". TEHNICKI GLASNIK-Technical Journal. September 2018. Vol. 12-3. p.159-165. DOI: https://doi.org/10.31803/tg-20180613152851

[2] Koch C, Ullman S. "Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry". Human Neurobiology. February 1985. Vol. 4-4. p.115-141. DOI: http://dx.doi.org/10.1007/978-94-009-3833-5_5

[3] Itti L, Koch C, Niebur E. "A Model of Saliency Based Visual Attention for Rapid Scene Analysis". IEEE Transactions on Pattern Analysis and Machine Intelligence. November 1998. Vol. 20-11. p.1254-1259. DOI: http://dx.doi.org/10.1109/34.730558

[4] Wu Z M, Wang J, Hu L, et al. "Co-saliency Detection Based on Convolutional Neural Network and Global Optimization". Journal of Electronics & Information Technology. December 2018. Vol. 40-12. p.2896-2904.

[5] Wang L Z, Wang L J, Lu H J, et al. "Salient Object Detection with Recurrent Fully Convolutional Networks". IEEE Transactions on Pattern Analysis and Machine Intelligence. July 2019. Vol. 41-7. p.1734-1746. DOI: http://dx.doi.org/10.1109/TPAMI.2018.2846598

[6] Li G B, Yu Y Z. "Contrast-Oriented Deep Neural Networks for Salient Object Detection". IEEE Transactions on Neural Networks and Learning Systems. December 2018. Vol. 29-12. p.6038-6051. DOI: http://dx.doi.org/10.1109/TNNLS.2018.2817540

[7] Chen S H, Tan X L, Wang B, et al. "Reverse Attention-Based Residual Network for Salient Object Detection". IEEE Transactions on Image Processing. January 2020. Vol. 29. p.3763-3776. DOI: http://dx.doi.org/10.1109/TIP.2020.2965989

[8] Du L, Li L, Wei D, et al. "Saliency-Guided Single Shot Multibox Detector for Target Detection in SAR Images". IEEE Transactions on Geoscience and Remote Sensing. December 2019. Vol. 58-5. p.3366-3376. DOI: http://dx.doi.org/10.1109/TGRS.2019.2953936

[9] Han J W, Cheng G, Li Z P, et al. "A Unified Metric Learning-based Framework for Co-saliency Detection". IEEE Transactions on Circuits and Systems for Video Technology. May 2017. Vol. 28-10. p.2473-2483. DOI: http://dx.doi.org/10.1109/TCSVT.2017.2706264

[10] Yun L, Jung C, Wang X, et al. "Part-Level Convolutional Neural Networks for Pedestrian Detection Using Saliency and Boundary Box Alignment". IEEE Access. February 2019. Vol. 7. p.23027-23037. DOI: http://dx.doi.org/10.1109/ACCESS.2019.2899105

[11] Zhang W M, Yao Z F, Gao Y K, et al. "A Deep Convolutional Network for Saliency Object Detection with Balanced Accuracy and High Efficiency". Journal of Electronics & Information Technology. May 2020. Vol. 42-5. p.1201-1208.

[12] Kang D, Park S, Paik J. "SdBAN: Salient Object Detection Using Bilateral Attention Network With Dice Coefficient Loss". IEEE Access. June 2020. Vol. 8. p.104357-104370. DOI: http://dx.doi.org/10.1109/ACCESS.2020.2999627

[13] Azaza A, Kabbai L, Abdellaoui M, et al. "Salient Regions Detection Method Inspired from Human Visual System Anatomy". In: 2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Monastir, Tunisia: IEEE, 2016, p.155-160. DOI: http://dx.doi.org/10.1109/ATSIP.2016.7523087

[14] Tünnermann J, Born C, Mertsching B. "Saliency From Growing Neural Gas: Learning Pre-Attentional Structures for a Flexible Attention System". IEEE Transactions on Image Processing. May 2019. Vol. 28-11. p.5296-5307. DOI: http://dx.doi.org/10.1109/TIP.2019.2913549

[15] Andrea M. "Direction Selective Contour Detection for Salient Objects". IEEE Transactions on Circuits and Systems for Video Technology. February 2018. Vol. 29-2. p.375-389. DOI: http://dx.doi.org/10.1109/TCSVT.2018.2804438

[16] Qin Y, Lu H C, Xu Y Q and Wang H. "Saliency Detection via Cellular Automata". In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA: IEEE, 2015, p.110-119. DOI: http://doi.org/10.1109/CVPR.2015.7298606

[17] Zhang Q, Li Y, Li W J, et al. "Salient Object Detection via Deep Features and Multiple Kernel Boosting Learning". Journal of Image and Graphics. 2019. Vol. 24-7. p.1096-1105.

[18] Chakraborty S, Mitra P. "A Dense Subgraph based Algorithm for Compact Salient Image Region Detection". Computer Vision and Image Understanding. April 2016. Vol. 145. p.1-14. DOI: https://dx.doi.org/10.1016/j.cviu.2015.12.005

[19] Cui L Q, Chen J J, Ren Q Y, et al. "Saliency Object Detection Based on Multiple Features and Prior Information". Journal of Image and Graphics. February 2020. Vol. 25-2. p.321-332.

[20] Ren J R, Liu Z, Zhou X F, et al. "Saliency Integration Driven by Similar Images". Journal of Visual Communication and Image Representation. January 2018. Vol. 50. p.227-236. DOI: https://dx.doi.org/10.1016/j.jvcir.2017.12.002

[21] Nikitha R, Vedhapriyavadhana R, Anubala V P. "Video Saliency Detection using Weight Based Spatio-Temporal Features". In: 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India: IEEE, 2018, p.343-347. DOI: http://doi.org/10.1109/ICSSIT.2018.8748400

[22] Jaemsiri J, Titijaroonroj T, Rungrattanaubol J. "Modified Scale-Space Analysis in Frequency Domain Based on Adaptive Multiscale Gaussian Filter for Saliency Detection". In: 16th International Joint Conference on Computer Science and Software Engineering (JCSSE), Chonburi, Thailand: IEEE, 2019, p.218-223. DOI: http://doi.org/10.1109/JCSSE.2019.8864211

[23] Zhang D M, Jin G Q, Dai F, et al. "Salient Object Detecion Based on Deep Fusion of Hand-Crafted Features". Chinese Journal of Computers, September 2019. Vol. 42-9. p.2076-2085. DOI: http://dx.chinadoi.cn/10.11897/SP.J.1016.2019.02076

[24] Zhao J X, Liu J J, Fan D P, et al. "EGNet: Edge Guidance Network for Salient Object Detection contrast". In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South): IEEE, 2019, p.8779-8788. DOI: http://dx.doi.org/10.1109/ICCV.2019.00887

[25]Guo F, Wang W G, She J B, et al. "Video Saliency Detection Using Object Proposals". IEEE Transactions on Cybernetics. November 2018. Vol. 48-11. p.3159-3170. DOI: http://dx.doi.org/10.1109/TCYB.2017.2761361

[26] Xuan D D, Wang J, Wang Z. "Salient Target Detection Based on High-level Priori Semantics". Journal of Chongqing University of Posts and Telecommunications(Natural Science Edition). April 2020. Vol. 32-2. p.304-312. DOI: http://dx.chinadoi.cn/10.3979/j.issn.1673-825X.2020.02.018

[27] Avsar E, Salcin K. "Detection and Classification of Brain Tumours from MRI Images using Faster R-CNN". TEHNICKI GLASNIK-Technical Journal. December 2019. Vol. 13-4. p.337-342. DOI: https://doi.org/10.31803/tg-20190712095507

[28] Akram A, Wang N N, Gao X B, et al. "A Comparative Study on Face Sketch Synthesis". IEEE Access. July 2018. Vol. 6. p.37084-37093. DOI: http://dx.doi.org/10.1109/ACCESS.2018.2852709

[29] Wang L Z, Wang L J, Lu H C, et al. "Salient Object Detection with Recurrent Fully Convolutional Networks". IEEE Transactions on Pattern analsis and Machine Intelligence. June 2018. Vol. 41-7. p.1734-1746. DOI: http://dx.doi.org/10.1109/TPAMI.2018.2846598

[30] Tsai C C, Hsu K J, Lin Y Y, et al. "Deep Co-Saliency Detection via Stacked Autoencoder-Enabled Fusion and Self-Trained CNNs". IEEE Transactions on Multimedia. August 2019. Vol. 22-4. p.1016-1031. DOI: http://dx.doi.org/10.1109/TMM.2019.2936803

[31] Jeong D J, Hwang I, Cho N. "Co-Salient Object Detection Based on Deep Saliency Networks and Seed Propagation over an Integrated Graph". IEEE Transactions on Image Processing. July 2018. Vol. 27-12. p.5866-5879. DOI: http://dx.doi.org/10.1109/TIP.2018.2859752

[32] Prashant W, Murala S. "MSFgNet: A Novel Compact End-to-End Deep Network for Moving Object Detection". IEEE Transactions on Intelligent Transportation Systems. November 2018. Vol. 20-11. p.4066-4077. DOI: http://dx.doi.org/10.1109/TITS.2018.2880096

[33] Çağlar A, Possegger H, Mauthner T, et al. "Spatiotemporal Saliency Estimation by Spectral Foreground Detection". IEEE Transactions on Multimedia. June 2017. Vol. 20-1. p.82-95. DOI: http://dx.doi.org/10.1109/TMM.2017.2713982

[34] Jerripothula K R, Cai J F, Yuan J S. "Efficient Video Object Co-Localization With Co-Saliency Activated Tracklets". IEEE Transactions on Circuits and Systems for Video Technology. March 2019. Vol. 29-3. p.744-755. DOI: http://dx.doi.org/10.1109/TCSVT.2018.2805811

[35] Li D M, Li J, Liang D C, et al. "Multiple Salient Objects Detection Using Multi-scale Prior and Deep Features". Acta Automatica Sinica. November 2019. Vol. 45-11. p.2058-2070. DOI: http://dx.chinadoi.cn/10.16383/j.aas.c170154

[36] Akram A, Wang N N, Gao X B, et al. "Integrating GAN with CNN for Face Sketch Synthesis". In: 4th International Conference on Computer and Communications (ICCC), Chengdu, China: IEEE, 2018. p.1483-1487. DOI: http://dx.doi.org/10.1109/CompComm.2018.8780648

[37] Li Y, Cao X, Li J. "A new cyber security risk evaluation method for oil and gas SCADA based on factor state space". Chaos Solitons & Fractals.2016. Vol.89,p.203-209.DOI: https://doi.org/10.1016/j.chaos.2015.10.030

[38] Yan Q, Xu L, Shi J P, et al. "Hierarchical Saliency Detection". In: 2013 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Portland, USA: IEEE, 2013, p.1155-1162. DOI: http://dx.doi.org/10.1109/CVPR.2013.153

[39] Wang L J, Lu H C, Wang Y F, et al. "Learning to Detect Salient Objects with Image-level Supervision". In: 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Honolulu, USA: IEEE, 2017, p.136-145. DOI: http://dx.doi.org/10.1109/CVPR.2017.404

[40] Yang L , Cao X , Geng X . "A novel intelligent assessment method for SCADA information security risk based on causality analysis". Cluster Computing, 2017.Vol.22. p. 5491–5503.DOI: https://doi.org/10.1007/s10586-017-1315-4.

ACKNOWLEDGEMENTS

This study was supported by the National Key Research and Development Plan of China (No.2017YFC0804400, No.2017YFC0804401), Ministry of Housing and Urban-Rural Development Science and Technology Planning Project (2016-R2-060), Jiangsu technology project of Housing and Urban-Rural Development (No.2018ZD265), Xuzhou Science and Technology Plan Projects (KC18011).

SUPPLEMENTARY MATERIAL

http://www.revistadyna.com/documentos/pdfs/_adic/9844-1.pdf