WILEY | Hindawi

Retraction

Retracted: Intelligent Image Saliency Detection Method Based on Convolution Neural Network Combining Global and Local Information

Security and Communication Networks

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] S. Zou, W. Chen, and H. Chen, "Intelligent Image Saliency Detection Method Based on Convolution Neural Network Combining Global and Local Information," *Security and Communication Networks*, vol. 2022, Article ID 3647387, 9 pages, 2022.

WILEY | Hindawi

Research Article

Intelligent Image Saliency Detection Method Based on Convolution Neural Network Combining Global and Local Information

Songshang Zou,[1] Wenshu Chen,[2] and Hao Chen [1]

[1]*College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China*
[2]*College of Computer Science, University of Bristol, Bristol BS8 1QU, UK*

Correspondence should be addressed to Hao Chen; chenhao@hnu.edu.cn

Image saliency object detection can rapidly extract useful information from image scenes and further analyze it. At present, the traditional saliency target detection technology still has the edge of outstanding target that cannot be well preserved. Convolutional neural network (CNN) can extract highly general deep features from the images and effectively express the essential feature information of the images. This paper designs a model which applies CNN in deep saliency object detection tasks. It can efficiently optimize the edges of foreground objects and realize highly efficient image saliency detection through multilayer continuous feature extraction, refinement of layered boundary, and initial saliency feature fusion. The experimental result shows that the proposed method can achieve more robust saliency detection to adjust itself to complex background environment.

1. Introduction

The introduction of visual attention mechanism in computer image processing can not only improve the data filtering ability and operation speed of the computer, but also have important application value in the fields of target recognition, target tracking, image analysis, and understanding [1]. Image saliency refers to the local region which attracts human visual attention when observing the view of a certain area. This local region is usually called as saliency region. Saliency detection is mainly used to underscore the saliency regions in images or videos [2]. Along with the rapid development of computer vision and the emergence of Artificial Intelligence (AI), to use computer technology to simulate the attention mechanism of human eyes has become a newly emerging but challenging research hotspot [3]. At present, image saliency detection has been widely applied in the following domains: image segmentation, object detection, and video coding. The study of saliency object detection has become increasingly mature and its practical

applications have been expanded to more and more fields [4]. Saliency object detection is an essential step in image processing and it can be used in image retrieval, semantic segmentation, object recognition, object tracking, and many other fields. In object recognition, image saliency object detection has finished in advance the related preprocessing work, which is good for further recognition [5]. In object tracking tasks, image saliency algorithms can filter the object information of interest quickly from the scenes, make a model based on this saliency visual attention mechanism, and realize object tracking [6]. In summary, saliency object detection algorithm is not perfect in processing saliency information and more satisfactory methods need to be explored.

Deep learning has made breakthroughs in image processing, automatic speech recognition (ASR), and natural language processing and other fields, laying a solid foundation for the future development of AI. This paper has applied the method of deep learning in saliency object detection, reduced the usage of the method to describe

artificially designed specific features, used the adaptive feature learning and description ability of neural network, effectively extracted features, and trained tag data to introduce feasible priori knowledge in the detection process. The experimental result has shown that the feature extraction method based on CNN has stronger image information representation ability.

The specific contributions of this paper include the following:

(i) This paper has analyzed visual saliency detection algorithm from image information process level, object of saliency detection, contrast of image features, and other perspectives and referred to and understood the characteristics and shortcomings of different types of algorithms.

(ii) To improve the image performance of CNN, this paper has conducted theoretical analysis on CNN model and studied the factors which affect the performance of CNN.

(iii) For the issues of low accuracy and efficiency of existing saliency detection methods extracting saliency object regions, and based on the selectivity and initiative of human visual neural system, this paper has combined the color contrast features, color distribution features, and location information in the bottom layer of image, fused multichannel features, conducted multiscale analysis, calculated saliency features, and extracted saliency regions from the image.

(iv) This paper has used CNN to make up for the incomplete dimensions of artificially designed specific features and constructed the saliency detection framework for complex images. The testing experiment has proven the effectiveness of the algorithm in this paper.

The rest of this paper is organized as follows. Section 2 discusses related work, followed by data computing and neural network structure designed in Section 3. The proposed algorithm and optimal model are discussed in Section 4. Section 5 shows the simulation experimental results, and Section 6 concludes the paper with summary and future research directions.

2. Related Work

When faced with massive visual stimulation, the unique attention mechanism of human eyes can filter the contents of importance or interest and these content regions are saliency regions. Visual saliency studies that the important regions can be quickly extracted in media processing tasks, and limited computing resources can be distributed to the locations of concern. According to different information processing levels, visual saliency detection algorithms can be divided into 3 types: bottom-up data-driven saliency region detection algorithm, top-down task-driven saliency object detection algorithm, and the combined method of top-down method and bottom-up method [7]. According to different

saliency detection objects, the algorithms include space-based saliency detection, object-based saliency detection, and feature-based detection. According to the contrast of image features, they are local feature contrast detection method and global feature contrast detection method. The saliency detection method based on local feature contrast mainly generates high saliency value in the edge regions of scenarios, making the saliency object region fail to have consistent highlight value. For example, IT algorithm and Ma have proposed the MZ algorithm based on local contrast analysis [8]. On the other hand, the saliency detection method based on global features is to overall examine the contrast relationship in scene image and measure saliency degree through pixel color differences, and spatial distance. This kind of method applies to the separation of foreground from background image and the typical algorithms include LC algorithm and RC algorithm [9]. In the field of computer vision, CNN has been studied more and more widely and it works extraordinarily in solving practical complex issues. Based on large-scale image data, CNN uses high-level semantic information to fully dig the detailed information in the context and greatly improves the accuracy and speed of image detection and recognition [10].

Image saliency object detection tasks usually encounter some challenging natural scene image, e.g., a complex background, a low contrast, and a small object [11]. At this time, traditional saliency detection algorithm does not work very well in background suppression and it usually fails to lock the object, making the saliency object fail to be accurately and comprehensively detected [12]. Under this circumstance, it needs to further explore and dig the potential high-level semantic information and globally spatial information of the image. Now the image saliency object detection algorithm based on deep learning has made great progress in solving this kind of challenging and complex scenes [13].

In 2007, Dr. Zhang et al. [14] had proposed treating saliency detection as a two-value segmentation and used the saliency map in the form of grey-scale map as the result of saliency detection. Goekhan et al. [15] raised the frequency-tuned (FT) algorithm, in which image-based low-level features were used as the filter to process the image to obtain frequency domain information, enabling the algorithm to detect salient object through frequency domain information. Banu et al. [16] brought forward context-based saliency detection method, i.e., context-aware (CA) algorithm, which calculates the global and local color contrasts in a single scale and then expands this contrast to different scales to obtained multiscale image saliency information and finally extends the scope of saliency region according to high-level semantic information to get a more accurate result. Liang et al. [17] used color histogram information and regional color information to build saliency model and achieved a good effect. In [18], two different saliency detection algorithms were proposed, which are histogram-based contrast (HC) and region-based contract (RC). Traditional visual saliency methods have excellent performance in natural scenes which have a single object or a high contrast [2, 19], but they usually fail in challenging and complex scenes (e.g., complex

background, low contrast, and small objects) [20, 21]. By this time, these methods cannot conduct detection on all saliency objects in an accurate and comprehensive manner; meanwhile, they are weak in suppressing complex background. In recent years, some methods based on deep learning fully dig the global spatial information and high-level semantic information through CNN and they have made more highly efficient and accurate experiment results than traditional algorithms [22, 23]. By constructing multilevel network model, deep learning automatically conducts learning from end to end, extracts image features [24], extracts the key image features by initializing weights [25], focuses on the saliency region in each image to form attention mechanism, makes the process in which computers learn and process data closer to the process of human visual system in image understanding and analysis [26], extracts low-level features and high-level extract semantic features through multilevel network structure, and effectively reduces the dependency on artificially marked low-level features when realizing saliency object detection [27].

3. Data Computing and Design of Neural Network Structure

To achieve saliency object detection with deep convolutional neural network can effectively conquer the problem of incomplete specific characteristic dimension in artificial design. The foregoing network includes 2 convolutional layers, 2 pooling layers, and 3 fully connected layers. The data is input into the network from the input layer. After convolutional operation, pooling operation, and fully connected operation, CNN will use loss function to calculate the error between the output result and the real data, conduct gradient computing through back propagation on the weight coefficient and bias factors in every layer, get the optimal updated gradient value of the weight coefficient and bias factors, complete the modification of the weights (mainly including the weights in the feature extraction layer and those in the monolayer perceptron in the end) in the entire CNN, and update the whole CNN.

3.1. Construction of Training Data.
90% of data in the dataset are used to train the parameters of the neural network and the rest, 10%, will be used in the test. We will randomly choose n image regions Q_P with a size of $m \times m$ from the given images and their marked saliency images and identify their tag value according to the number of black and white pixels in the chosen region in the saliency images. If the number of black pixels is more than that of white pixels, the tag value is marked as 0; otherwise, it is marked as 1.

3.2. Database Preprocessing.
Normalize the filtered dataset and the processing method for Q_P region is as follows:

$$Q'_P = \frac{Q_P - \text{mean}(Q_P)}{\max(Q_P) - \min(Q_P)}. \tag{1}$$

mean($\cdot$) is the mean operator used to calculate the mean value of pixels; max($\cdot$) is the maximization operator to

calculate the maximum pixel value; and min($\cdot$) is the minimization operator to calculate the minimum pixel value.

3.3. Structural Design of Neural Network.
This design adopts an 8-layer network structure and the successive order of various layers is input layer-convolutional layer-pooling layer-pooling layer-convolutional layer-poling layer-fully connected layer 1-fully connected layer 2. After designing the overall structure, next comes the design for the functions of each layer, including the size, dimensions, and activation function for the convolutional kernel in the convolutional layer, the size, dimensions, and activation function for the filter in the pooling layer; and the number of neurons, activation, and classifier function in the fully connected layer.

3.3.1. Convolutional Operation.
Convolutional operation mainly converts the image data into the data to be processed by computer and it is defined as follows:

$$Q_j^l = \text{sigm}\left(\sum_{i=1} I_i^{l-1} \otimes k_{ij}^l + b_j^l\right). \tag{2}$$

L refers to the number of total layers in the neural network; $l \in \{1, \ldots, L\}$, Q^{l-1} represents the input data of the 1st convolutional layer; Q^l is the output data of the 1st convolutional layer; k_{ij}^l is the convolutional kernel parameter; b_j^l is the bias item of neurons; and sigm($\cdot$) is sigmoid activation function, which means convolutional operation.

3.3.2. Pooling Operation.
Pooling layer filters the data according to the data features extracted in the convolutional layer through filter and without losing the key image information; it reduces the dimensions of data features and the complexity of network computing. The pooling operation is defined as follows:

$$Q_i^{l+1} = \text{pooling}\left(I_i^l, m\right) + b_j^{l+1}. \tag{3}$$

b_j^{l+1} is the bias item and pooling($\cdot$) means to execute one maximum pooling operation every other $m \times m$ region in the input image block with no overlapping between $m \times m$ regions.

3.3.3. Fully Connected Operation.
After convolutional layers and pooling layers, the data finally enter the fully connected layer for processing. In this layer, all features are connected and the output value is sent to the classifier.

3.4. Data Processing in Neural Network

Step 1. Select 90% original image data, tag saliency image, and choose 10% test data from the dataset. If the input image is a 28 ∗ 28 single channel image, the input shape is [batch]_ size, 28, 28, 1], and determine their tag value according to the number of black and white pixels.

Step 2. Normalize the data to make their dimensions consistent.

Step 3. The first convolution is composed of 32 5 ∗ 5 convolution kernels, whose shape is [5, 5, 1, 32], and its step size is [1, 1, 1, 1], followed by the max of 2 ∗ 2 in the first layer. The shape is [1, 2, 2, 1], and the step strides is [1, 2, 2, 1].

Step 4. The second convolution is composed of 64 5 ∗ 5 convolution kernels, whose shape is [5, 5, 32, 64], and its step size is [1, 1, 1, 1], followed by the max of 2 ∗ 2 in the first layer. The shape is [1, 2, 2, 1], and the step strides is [1, 2, 2, 1]

Step 5. Then, the size of the image after the above two convolutioning and the changes after two pooling are as follows:

[batch_ size, 28, 28, 1], (first layer convolution). [batch_ size, 28, 28, 32], (first layer pool). [batch_ size, 14, 14, 32], (second level convolution), [batch_ size, 14, 14, 64], (second layer pool). [batch_ size, 7, 7, 64].

Step 6. Input the data into the fully connected layer for processing, which contains 600 neurons, and output the result from the last layer, softmax classifier.

Step 7. Save the well-trained model and test it with the testing set.

4. Our Proposed Optimal Model and Algorithm

This paper has compared the method in this paper with the other 9 saliency algorithms (FES, Sigsaliency, GB, CA, MC, COV, SER, MZ, and SWD), conducted saliency test on CIFAR database and AFW database, plotted PR curve, ROC curve, and AUC indicator, and assessed the comprehensive performance of these algorithms.

4.1. Introduction of Algorithms Selected in This Experiment

(1) Sigsaliency is a saliency algorithm based on image signature and it studies whether the approximate foreground overlaps with the visual foreground. This algorithm can predict the best human fixation and it can be done within a shorter running time. As shown in the test, the saliency algorithm based on image signature has a result closer to human perception than the like algorithms.

(2) MC method formulates the saliency test by absorbing the Markov chain in the image model, which takes into consideration both the differences in appearance and spatial distribution of the outstanding object and the background. The method chooses the virtual boundary node as the absorption node in Markov Chain and calculates the absorption time from every transient node to the boundary absorption node. The absorption time of transient node measures its global similarities with all absorption nodes and when the absorption time is used as a measurement, it will separate consistently the outstanding object from the background. As the time

from transient node to absorption node depends on the weight of the path and its spatial distance, the background region in the image center may be very obvious. The experiment shows that this method has excellent robustness and efficiency.

(3) MZ method proposes the definition of saliency by considering the content that the visual system tries to optimize while directing attention and the model generated by this algorithm is a Bayesian framework. From bottom to top, saliency appears with visual features and the global saliency (combining the top-down information with the bottom-up information) shows up as the point-by-point information between features. Its performance is better than the existing algorithms. Unlike the existing saliency measurement, its measurement is obtained from the statistical information of natural images, while such statistical information is pre-acquired from the sets of natural images.

(4) SER method provides a novel unified framework for saliency detection. By adopting a bottom-up method, it can calculate the local regression kernel to measure the similarities between the pixel and its surrounding environment and then it calculates the visual saliency with the above similarity measurement. This framework will generate a saliency image, in which every pixel represents the statistical possibility of feature matrix saliency with its surrounding feature matrix given, use cosine similarity of matrix, and prove its performance in common human eye fixation data and some psychological patterns.

(5) CA method proposes the detection method of four principles based on psychological literature. This model has considered both the local features and global features of images at the same time and overcome the practice that salient region is a fixed region and the region only considers the foreground image and ignores the informative background image. This method can extract the outline of the salient region and prevent the distortion of important regions, which is good for subsequent processing; however, it needs to calculate the saliency of every pixel, equivalent to local region, which is a large amount of computation.

(6) SWD method is a visual saliency detection method based on spatial weighted dissimilarity and it measures saliency by combining the differences, spatial distance, and center deviation between image blocks and extracts the principal component by sampling patches from the current image. As suggested in the test, this method is better the current methods when predicting image saliency.

(7) GB method manifests that the effective color appearance model (including the principal selection of parameters and inherent spatial collection mechanism) in human vision can be promoted to get the saliency model better than the existing one. This
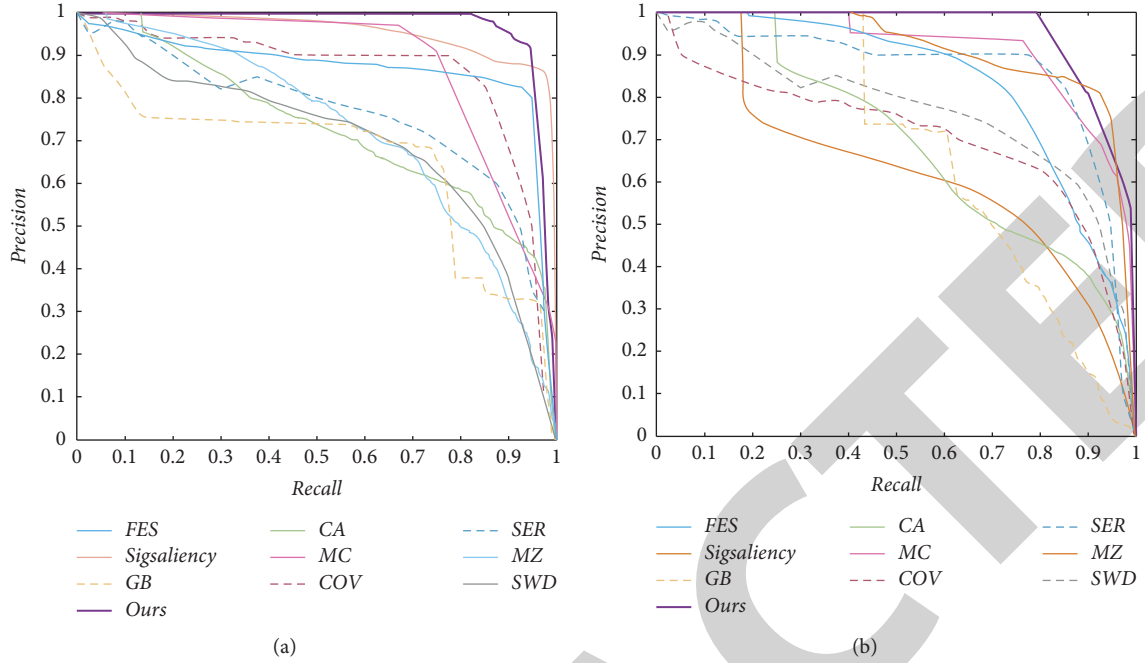
Figure 1: Comparison of PR curves. (a) CIFAR. (b) AFW.

method has optimized the proportional weighting function so as to better copy the data in color appearance, and determined the size of proper suppression window by training Gaussian mixture model to make the selection of various parameters more reasonable.

(8) FES method uses the center-surround method in the saliency detection, which is to estimate the saliency of local feature contrast under the Bayesian framework, especially by using sparse sampling and kernel density to estimate the necessary distribution. Besides, in essence, this method has implicitly considered the issue of center deviation. Compared with the existing method, this method is faster in speed and it can run in real time.

(9) COV method by detecting the visually salient elements in complex natural scenes calculates the saliency model from bottom to top and parallelly checks a multitude of feature channels (e.g., color and direction) to calculate a separate feature map for every channel. Besides, it uses the covariance matrix of simple image features as the meta-feature of saliency estimation, provides the nonlinear integral of different features by modeling their correlation, gets the final saliency image through combination, and finally effectively improves the algorithm performance.

4.2. Evaluation Indexes. In the experiment, the following 5 indexes are used to quantitively compare the performance of different methods: precision-recall (PR) curve, receiver operating characteristic (ROC) curve, area under ROC curve (AUC), mean absolute error (MAE), and weighted F-measure (weighted F_β).

S represents the final saliency map obtained by this model and G refers to the real image. First, normalize the saliency maps to [0, 1]. Then perform two-value processing on the normalized saliency maps with the threshold within the scope of [0, 1] and get the corresponding two-value map M. Therefore, precision and recall can be calculated in pairs to form the PR curve. The calculation of precision and recall is as shown in the following formula:

$$
\text{precision} = \frac{|M \cap G|}{|M|},
$$
$$
\text{recall} = \frac{|M \cap G|}{|G|}. \tag{4}
$$

$|*|$ refers to the non-zero pixel value in the two-value map. Meanwhile, the paired false positive rate (FPR) and the true positive rate (TPR) can be obtained so as to form the ROC curve. FPR and TPR are calculated with the following formula:

$$
\text{TPR} = \frac{|M \cap G|}{|G|},
$$
$$
\text{FPR} = \frac{|M \cap \overline{G}|}{|\overline{G}|}. \tag{5}
$$

$\overline{G}$ is the negation of the real image. AUC is obtained by calculating the area under ROC area and it extracts two-dimensional information to a single numerical value. This index did not take into consideration the true negative example. Instead, MAE is a more comprehensive evaluation index and it is defined as the calculation of the mean absolute error between the normalized saliency map S and the real image G, as shown in the following formula:
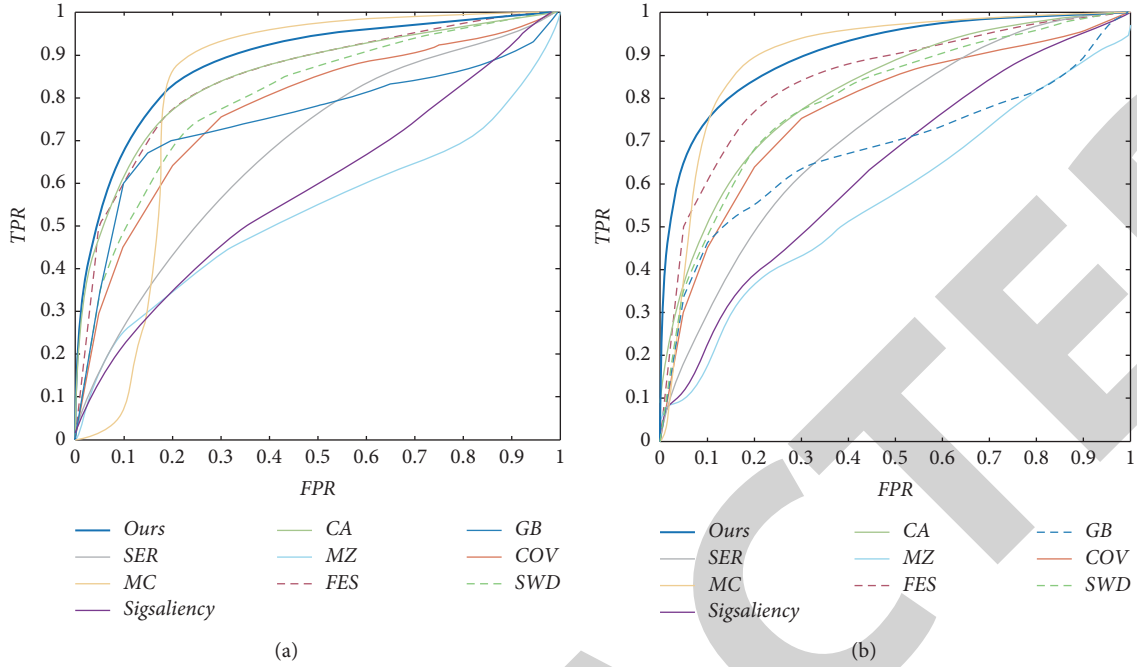
Figure 2: Comparison of ROC curves. (a) CIFAR. (b) AFW.

Table 1: Assessment of performance indexes.

| Algorithm | Data index | | | | | | | |
| | CIFAR | | | | AFW | | | |
	Recall	Precision	FPR	AUC	Recall	Precision	FPR	AUC
Ours	0.9844	0.9000	0.6000	0.8802	0.9801	0.9000	0.6000	0.9063
Sigsaliency	0.5432	0.2000	0.1000	0.5848	0.4691	0.4000	0.2000	0.6346
MC	0.9502	0.8000	0.7000	0.8180	0.9472	0.8000	0.7500	0.8986
MZ	0.9842	0.2000	0.1000	0.5306	0.9607	0.2000	0.1000	0.5661
SER	0.9928	0.8000	0.2000	0.6785	0.9747	0.8000	0.2000	0.7194
CA	0.9560	0.9000	0.7000	0.8476	0.9598	0.9000	0.6000	0.8149
SWD	0.9492	0.1000	0.2000	0.8141	0.9521	0.1000	0.2000	0.7961
GB	0.9067	0.1000	0.5000	0.6204	0.9306	0.1000	0.5000	0.6753
FES	0.9489	0.2000	0.1000	0.8096	0.9243	0.2000	0.1000	0.8087
COV	0.9278	0.7000	0.5000	0.8152	0.9021	0.7000	0.5000	0.8421

$$\text{MAE} = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} |S(x, y) - G(x, y)|. \qquad (6)$$

W and H represent the width and height of the image, respectively. As the defects of curve interpolation and all errors are deemed of equal importance, the above evaluation standards may not be reliable enough. Therefore, the experiment uses weighted F-measure to overcome this defect. The weighted F-measure is calculated according to the following formula:

$$F_\beta^\omega = \left(1 + \beta^2\right) \frac{\text{Precision}^\omega \times \text{Recall}^\omega}{\beta^2 \times \text{Precision}^\omega + \text{Recall}^\omega}. \qquad (7)$$

ω represents the weight function and β is set as 1 in the experiment.

5. Experimental Test and Performance Analysis

This paper has compared the proposed algorithm with the other 9 saliency algorithms (FES, Sigsaliency, GB, CA, MC, COV, SER, MZ, and SWD), conducted saliency test on CIFAR database and AFW database, plotted PR curve, ROC curve, and AUC indicator, and assessed the comprehensive performance of these algorithms.

Conduct image saliency test by using the algorithm in this paper and the comparison of PR curves is shown in Figure 1.

In PR curve, the better performance an algorithm has, the closer its PR curve to the upper right corner. For CIFAR data, the PR curve of the algorithm in this paper is closer to the top right corner than those of FES, Sigsaliency, GB, CA, MC, COV, SER, MZ, and SWD. In AFW data, the PR curve
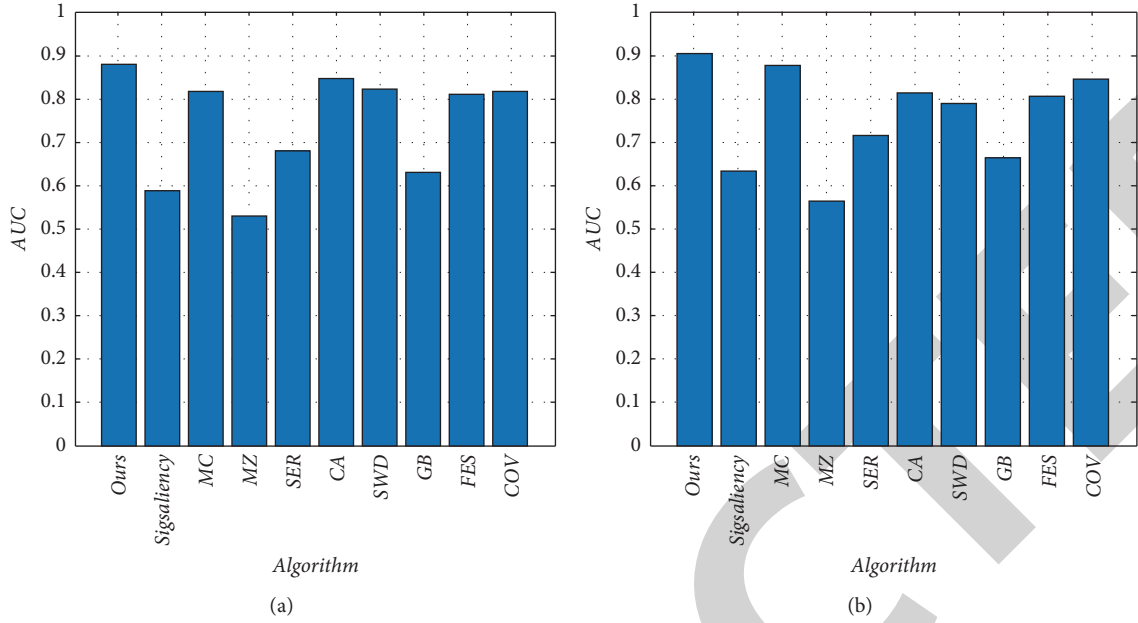
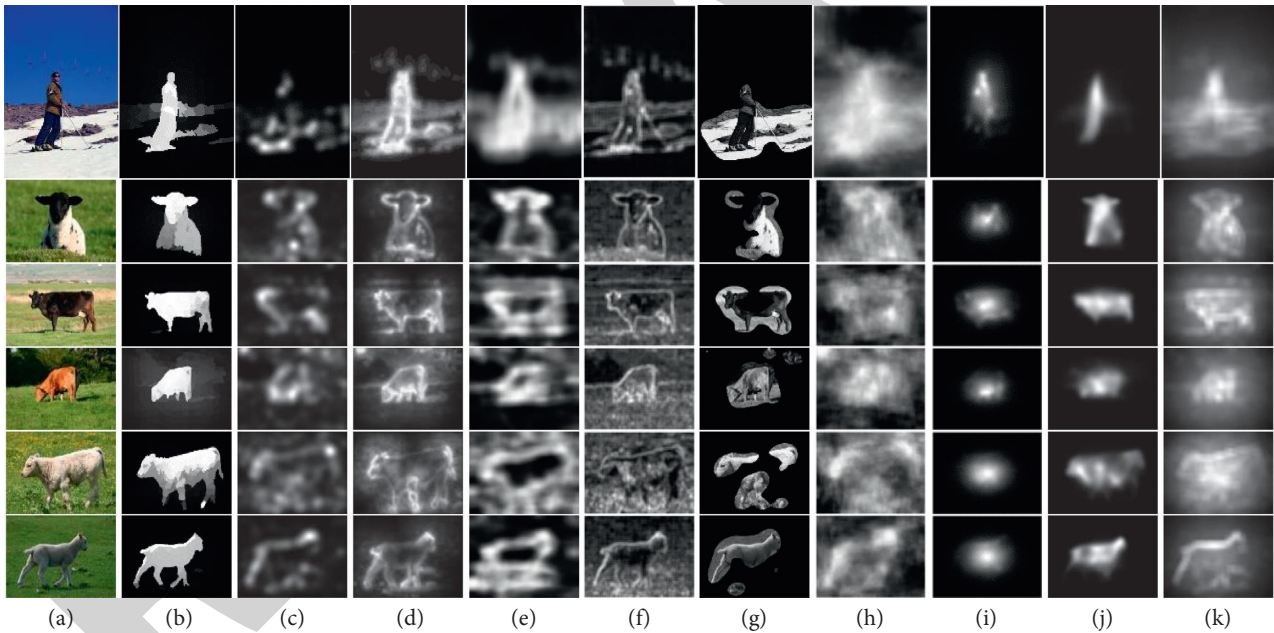Figure 3: Comparison of AUC value. (a) CIFAR. (b) AFW.

Figure 4: Comparison of saliency results (MSRA dataset). (a) Original. (b) Ours. (c) CA. (d) MC. (e) SER. (f) MZ. (g) Sigsaliency. (h) GB. (i) COV. (j) FES. (k) SWD.

of the algorithm in this paper is closer to the upper right corner than those of FES, GB, CA, MC, COV, SER, MZ, and SWD. The experiment result shows that the proposed algorithm has excellent performance and its effect gets the upper hand than most of other algorithms.

Make image saliency test by using the above-mentioned algorithms and the comparison of ROC curves is shown in Figure 2.

When it comes to ROC curve, the better performance an algorithm has, the closer its ROC curve to the upper left corner. In both CIFAR and AFW data, the ROC of the algorithm in this paper is closer to the top left corner than those of FES, Sigsaliency, GB, CA, MC, COV, SER, MZ, and SWD. The experiment result proves that the proposed algorithm (the curve marked bold) has extraordinary performance and it is better than the rest with regard to effect.

For CIFAR and AFW data, the comparisons of recall, precision, FPR, and AUC indicators between the algorithm in this paper and the other 9 algorithms are shown in Table 1 and the comparison of AUC value is shown in Figure 3.

It is clearly evident from Table 1 that our proposed algorithm has performed quite well in both CIFAR and AFW data in various performance indexes. As for precision and recall, it is much better than 8 other methods (FES, Sigsaliency, GB, CA, MC, COV, MZ, and SWD) in CIFAR data from the perspective of the numerical value of various indexes and it is almost the same as that of SER. And for AFW data, it is better than all other 9 algorithms in its performance.

As for FPR, in CIFAR data, the algorithm in this paper is better than FES, Sigsaliency, GB, COV, SER, MZ, and SWD; and in AFW data, it is better than FES, Sigsaliency, GB, CA, COV, SER, MZ, and SWD. With regard to AUC, this algorithm has better performance in both CIFAR and AFW data; in fact, it has the best performance. Compared with other algorithm except CA, it has also improved by over 3%.

Through the above experimental results, it can be observed that the change trend of the design curve is consistent with that of our proposed algorithm. But in general, our proposed algorithm is the global optimal algorithm among all the algorithms. Figure 4 is the comparison of saliency results in MSRA dataset.

As shown in Figure 4, our proposed algorithm can extract the foreground object more completely no matter whether the background is complex or not and it has well eliminated the redundant background information. Multi-layered continuous feature extraction is mainly based on CNN and it can get the initialized parameters of the pre-training in the training process and extract different layers to continuously extract and code the semantic information of different layers. Then through layered boundary refinement process, it converts low-resolution feature map into high-resolution feature map, preserves the edge information of the object in every layer, and continuously optimizes it to get clearer saliency object edges. The saliency information dug in each layer is different. In order to more comprehensively integrate different saliency elements, it fuses the initial saliency features and uses loss function for network training. It can be learnt that this algorithm can smoothen the saliency regions and optimize the saliency boundary, effectively fuse features at different layers, highlight the saliency regions, and make the saliency boundary more accurate.

6. Conclusion and Future Work

Human vision system can quickly filter the most valuable visual information from the real scenes. Image saliency detection is to extract the regions of great importance from the image by using computers to simulate human visual system so as to simplify the complexity of subsequent utilization. This paper has come up with a CNN-based visual saliency detection model. At first, it maps all convolutional features of CNN into multiple internal scales. Then, it adaptively integrates various scale features and predicts the saliency maps for these maps; meanwhile, it fuses these maps to get the fused saliency map and underlines the saliency regions with various scale features. Finally, it uses fully connected conditions to randomly process the network to get the saliency map. The experimental result shows that the proposed method can effectively enhance object region, suppress background region, enable the algorithm to have excellent robustness while improving the object recognition rate, and realize effective end-to-end saliency object extraction.

Data Availability

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

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] A. Kompella and V. R. Kulkarni, "Weakly supervised multi-scale recurrent convolutional neural network for Co-saliency detection and Co-segmentation," *Neural Computing & Applications*, vol. 32, no. 21, Article ID 16571, 2020.

[2] K. Huang and S. Gao, "Image saliency detection via multi-scale iterative CNN," *The Visual Computer*, vol. 36, no. 7, pp. 1355–1367, 2020.

[3] A. Lopez-Alanis, R. A. Lizarraga-Morales, M. A. Contreras-Cruz, V. Ayala-Ramirez, R. E. Sanchez-Yanez, and F. Trujillo-Romero, "Rule-based aggregation driven by similar images for visual saliency detection," *Applied Intelligence*, vol. 50, no. 6, pp. 1745–1762, 2020.

[4] J. Ren, Z. Liu, X. Zhou, C. Bai, and G. Sun, "Co-saliency detection via integration of multi-layer convolutional features and inter-image propagation," *Neurocomputing*, vol. 371, pp. 137–146, 2020.

[5] W. Shu, K. Cai, and N. Xiong, "Research on strong agile response task scheduling optimization enhancement with optimal resource usage in green cloud computing," *Future Generation Computer Systems*, vol. 124, pp. 12–20, 2021.

[6] H. Tang, B. Wang, and L. Han, "Image saliency detection based on object compactness and regional homogeneity strategy," *Journal of Electronics and Information Technology*, vol. 41, no. 10, pp. 2532–2540, 2019.

[7] R. Cong, J. Lei, H. Fu, Q. Huang, X. Cao, and N. Ling, "HSCS: hierarchical sparsity based Co-saliency detection for RGBD images," *IEEE Transactions on Multimedia*, vol. 21, no. 7, pp. 1660–1671, 2019.

[8] Y. Rakesh and K. S. R. Krishna, "Saliency detection in stereoscopic images using adaptive Gaussian Kernel and Gabor filter," *Soft Computing*, vol. 23, no. 8, pp. 2485–2498, 2019.

[9] C.-C. Tsai, W. Li, K.-J. Hsu, X. Qian, and Y.-Y. Lin, "Image Co-saliency detection and Co-segmentation via progressive joint optimization," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 56–71, 2019.

[10] D. P. Bavirisetti and D. Ravindra, "Multi-focus image fusion using multi-scale image decomposition and saliency detection," *Ain Shams Engineering Journal*, vol. 9, no. 4, pp. 1103–1117, 2018.

[11] Z. Gao, N. Ayoub, D. Chen, B. Chen, and Z. Lu, "SAMM: surroundedness and absorption Markov model based visual saliency detection in images," *IEEE Access*, vol. 6, Article ID 71422, 2018.

[12] B.-J. Han and J.-Y. Sim, "Saliency detection for panoramic landscape images of outdoor scenes," *Journal of Visual Communication and Image Representation*, vol. 49, pp. 27–37, 2017.

[13] H. Avik, C. Subhasis, and V. Rajbabu, "An image texture insensitive method for saliency detection," *Journal of Visual Communication and Image Representation*, vol. 43, pp. 212–226, 2017.

[14] L. Zhang, J. Chen, and B. Qiu, "Region-of-Interest coding based on saliency detection and directional wavelet for remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 1, pp. 23–27, 2017.

[15] Y. Goekhan, S. Debashis, and K. Mohan, "Evaluating salient object detection in natural images with multiple objects having multi-level saliency," *IET Image Processing*, vol. 14, no. 10, pp. 2249–2262, 2020.

[16] A. S. Banu, P. Vasuki, and S. M. M. Roomi, "Target detection in SAR images using bayesian saliency and morphological attribute profiles," *Computer Communications*, vol. 160, pp. 738–748, 2020.

[17] H. Liang, J. Zou, K. Zuo, and M. J. Khan, "An improved genetic algorithm optimization fuzzy controller applied to the wellhead back pressure control system," *Mechanical Systems and Signal Processing*, vol. 142, Article ID 106708, 2020.

[18] T. Zhou, W. Wu, L. Peng et al., "Evaluation of urban bus service reliability on variable time horizons using a hybrid deep learning method," *Reliability Engineering & System Safety*, vol. 217, Article ID 108090, 2020.

[19] H. Liang, J. Zou, Z. Li, M. J. Khan, and Y. Lu, "Dynamic evaluation of drilling leakage risk based on fuzzy theory and PSO-SVR algorithm," *Future Generation Computer Systems*, vol. 95, pp. 454–466, 2019.

[20] W. Shu, K. Cai, and N. N. Xiong, "A short-term traffic flow prediction model based on an improved gate recurrent unit neural network," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–12, 2021, 10.1109/TITS.2021.3094659.

[21] L. Dong, M. N. Satpute, W. Wu, and D.-Z. Du, "Two-phase multidocument summarization through content-attention-based subtopic detection," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 6, pp. 1379–1392, 2021.

[22] W. Wei, B. Zhou, D. Połap, and M. Woźniak, "A regional adaptive variational PDE model for computed tomography image reconstruction," *Pattern Recognition*, vol. 92, pp. 64–81, 2019.

[23] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015.

[24] M. Cheng, G. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 409–416, Colorado Springs, CO, USA, June 2011.

[25] R. Zhao, W. Ouyang, H. Li, and X. Wang, "Saliency detection by multi-context deep learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1265–1274, Boston, MA, USA, June 2015.

[26] X. Wu, H. Wang, W. Tan, D. Wei, and M. Shi, "Dynamic allocation strategy of VM resources with fuzzy transfer learning method," *Peer-to-Peer Networking and Applications*, vol. 13, no. 6, pp. 2201–2213, 2020.

[27] X. Li, L. Zhao, L. Wei et al., "DeepSaliency: multi-task deep neural network model for salient object detection," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3919–3930, 2016.