IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 28 September 2023, accepted 14 December 2023, date of publication 26 December 2023,
date of current version 5 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3347583

RESEARCH ARTICLE

The Application of FLICSP Algorithm Based on Multi-Feature Fusion in Image Saliency Detection

ZHIXIAN LI, JIANWEI WU, AND GUOQIANG WANG
School of Computer and Information Engineering, Luoyang Institute of Science and Technology, Luoyang 471023, China
Corresponding author: Guoqiang Wang (litwgq@163.com)

This work was supported in part by the National Natural Science Foundation of Henan, Research on the Evaluation Method of Taijiquan Actions Based on Deep Learning under Grant 232300420157.

ABSTRACT To raise the performance of the current image saliency detection method and promote the extraction effect of salient regions in complex images, from the perspective of multi feature fusion, the image saliency detection method is studied. Corresponding saliency maps is obtained from color and texture channels respectively, and the final saliency map is obtained through linear fusion. On the basis of deep prior information, an improved deep convolutional neural network is used to extract relevant feature vectors such as image pixels, and the nearest neighbor classifier is trained to determine the ownership between pixels and image regions. Zero sample learning iterations are performed to improve the effectiveness of image saliency detection. On this basis, the combination of the above two methods is carried out to obtain the final image saliency detection method. Experimental design is conducted and Matlab software is selected, with a learning rate of 0.001 and 100 iterations. The results showed that compared to other algorithms, the improved algorithm had better detection performance. The detection error of this algorithm was smaller and the accuracy was higher. Its average absolute error value in the JUDD dataset was the smallest, 0.143, which was 0.34 less than the original algorithm. Its detection accuracy in the PASCAL dataset was the highest at 0.786. The missed detection rate and error detection rate of the improved algorithm were both higher than those of the original algorithm, with a missed detection rate of 4.7%, which was 2.5% lower than the original algorithm. Its error detection rate was 5.1%, while the original algorithm's missed detection rate was 3.2%. The improved algorithm resulted in better image quality, with a maximum peak signal-to-noise ratio of 39.45-dB, which was 6.82-dB higher than the original algorithm. The maximum similarity index value was 0.892. Research methods can effectively detect the saliency of complex images.

INDEX TERMS FLIC, VGG network model, texture features, color features, saliency testing.

I. INTRODUCTION

Along with the rapid growth of computer technology and various intelligent terminals, while greatly improving people's convenience in life, it also provides people with a large amount of high-quality external information such as images and videos. Computer vision is an important branch of computer technology, and it is a complex field. An important task of computer vision is to process the collected image and video information, and the processing effect can be similar to that of human processing [1], [2], [3]. This technology has broad applications, which involve high processing requirements, and some tasks require human assistance. Therefore, how to use intelligent devices such as computers to quickly and accurately analyze and process extensive images, so that they can independently complete various visual tasks like humans, is currently the most interesting issue in computer vision [4], [5], [6]. The clearer the picture, the more details it requires, and the more memory it requires. Therefore, computer programs must first read the complete picture, which will result

The associate editor coordinating the review of this manuscript and approving it for publication was Li Zhang.

in low efficiency. In real life, the images people see are usually very complex, including multiple targets and detailed information. However, people always selectively ignore some unimportant information and focus on certain information. However, computers do not possess such capabilities. Therefore, exploring methods for detecting image saliency is of great significance. In order to solve the problem of insufficient utilization of complex image features and improve the effectiveness of saliency detection, from the perspective of multi-feature fusion, research is conducted on image saliency detection (ISD) methods, and deep convolutional neural networks (CNNs) are applied to deeply mine image information to improve the effectiveness of ISD. The research is composed of four parts. The first part is a literature review, which introduces the research on ISD methods and depth CNN. The second part establishes an ISD algorithm with multi-feature fusion and optimizes the detection effect through deep CNNs. The third part analyzes the results and studies the detection effectiveness of the constructed saliency detection algorithm. The fourth part summarizes the research objectives, methods, and other contents, and proposes future research directions based on the insufficient research. Unlike previous studies, this article aims to address the issues of mutual interference between image fusion features and insufficient utilization of image features. It extracts deep uncertain features from images and mines useful information to obtain more accurate saliency maps.

II. RELATED WORKS

The growth of computer vision is conducive to promoting the development of object detection. And ISD is a hot topic in computer vision research. To achieve better segmentation results in welding defect segmentation, Lin Z et al. used local image enhancement methods to perform image enhancement, weld seam extraction, and other tasks, using the determination of local pixel non-uniformity factors to detect images. And based on the detection results, weld seam segmentation was performed using the maximum inter-class variance method. When identifying defects, the region growth algorithm was used. The results denoted that this method was beneficial for obtaining higher-quality defect images and improving segmentation precision [7]. In fuzzy and small vehicle detection, Xu K et al., to enhance the detection effect, built a relevant improved CNN, and enhanced the image through the relevant adaptive histogram and other methods. A multi-layer feature fusion target location network was used to locate the target in a complex background and other operations. Online hard sample mining algorithms and other methods were utilized to avoid negative sample imbalance. The findings indicated that compared to other models, the proposed method had better classification performance [8]. Zhou Y et al., in stereo vision, faced the problem of detecting significant areas in high-quality depth images. They used algorithms that directly detect significant objects for correlation detection, evaluated depth image confidence through depth reliability indicators, reflected main salient features using depth compactness, and enhanced depth contrast through rough background filtering. The outcomes indicated that the method used can perform significance detection well [9]. Wang G et al. conducted relevant anti-deception detection based on the characteristics of visual saliency to achieve better anti-deception detection effects in silicone mask attacks. Combined with the significance-weighted histogram, the corresponding local binary patterns operator was obtained to extract the relevant texture features. Under the action of the optical flow operator in the relevant direction, the facial motion features were extracted, and the relevant features were classified through the support vector machine. The outcomes denoted that the raised method is effective [10]. Singh et al. used a fully CNN to denoise the input image in the face of the problem of salient object detection in noisy images. Through end-to-end learning of relevant salient details, they performed salient object detection under the influence of convolutional layers. The results showed that the proposed model had good detection performance [11]. Govil et al. introduced convex hull-activated active contour technology in salient object detection to improve detection time and accuracy. After dataset validation, it was found that this method reduced computational time and had high detection accuracy [12].

Cui et al. combined the attention mechanism with extended convolution and relevant multi-scale features to obtain a relevant network model to achieve better ISD and reduce computational costs. This was used to selectively enhance saliency regions and integrate adjacent layer features to obtain more effective saliency results. The findings expressed that the raised network had a better detection effect [13]. Serte et al., faced with the problem of glaucoma fundus image detection, processed the fundus image by detecting the salient regions based on the image, removed its redundant parts, and constructed the relevant CNN model. The outcomes denoted that the obtained detection method could perform better detection [14]. In the visual saliency detection problem, Lin et al. reconstructed the CNN to input two-dimensional data and train on limited labeled data. The findings illustrated that the detection effect of this method was good [15]. To solve the problem of common saliency detection, Wei et al. obtained the group-depth collaborative saliency detection method on the basis of the full convolution network. Through it, they captured the group interaction information of group images, extracted the group features of their related significance detection, and constructed a unified depth learning scheme to obtain better common significance detection results and optimize the significance. The outcomes indicated that the proposed method had a good effect [16].

Sum up, in ISD, there are some detection algorithms, but the results obtained by many algorithms on images with complex content and texture are somewhat not the same with the true value, and there is the problem of image feature interference. In this regard, research has been conducted to separate image features, followed by multi-feature fusion, and the use of deep CNNs has a good mining effect to extract

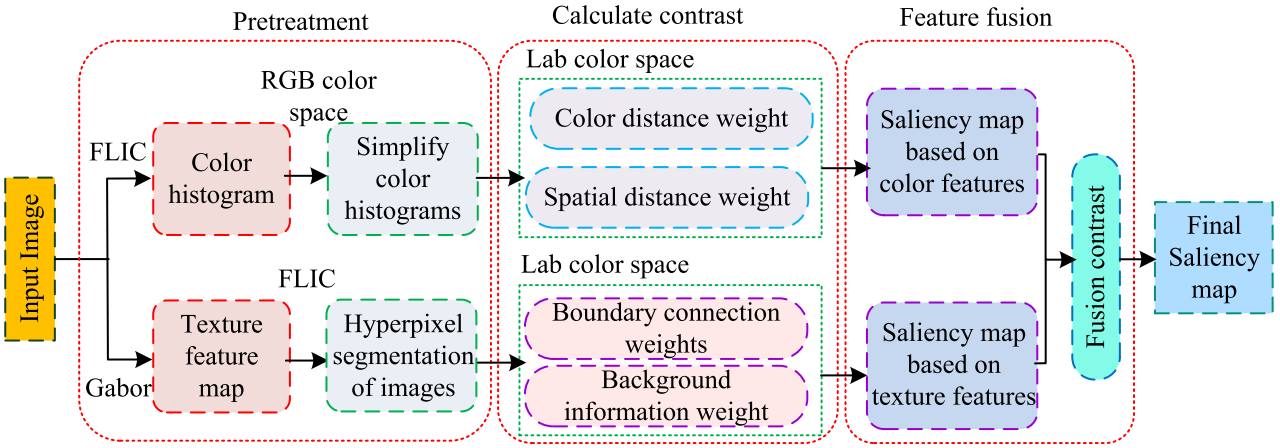

FIGURE 1. Process of correlation saliency detection algorithm.

deep uncertain characteristics, to improve the saliency detection effect. In terms of innovation and contribution, compared to previous research, the article considers the extraction of deep uncertain features in images, fully mining and utilizing image information, and providing new ideas for complex ISD problems.

III. ISD BASED ON FLICSP ALGORITHM

To raise the effectiveness of saliency detection in images with complex textures and cluttered backgrounds, a fast linear iterative clustering with active search (FLICS) based ISD algorithm is proposed, which integrates color and texture features. For the purpose of easy writing, it is set as the FLICS algorithm. After pre-processing the images, color feature-based and texture feature-based contrast calculations are respectively performed on the images. Feature fusion is performed to obtain the fused feature saliency map. By combining this algorithm with a significance detection model based on deep prior information assistance, an ISD algorithm based on multi-feature fusion and deep prior information using fast linear iterative clustering with active search (FLICSP) was obtained, to improve the accuracy of saliency maps.

A. APPLICATION OF FILCS ALGORITHM AND FUSED FEATURES IN ISD

In image processing technology, ISD is one of the most important techniques. This technology is applied in various tasks. The bottom-up detection algorithm is a type of ISD technology, which has a fast detection speed. However, this algorithm has certain drawbacks, and its performance in detecting images with complex textures and cluttered backgrounds is not ideal. Based on this type of algorithm, research is conducted to separate image color and texture features and fully apply image spatial information. Propose the FLICS algorithm, and its related process is shown in Figure 1.

In Figure 1, the process of FLICS algorithm can be mainly separated into three parts, namely image pre-processing, calculating contrast, and feature fusion. In image pre-processing, the FLICS algorithm in the super-pixel segmentation algorithm is selected to segment the texture features of the image. In natural images, neighboring pixels have similar properties. Compared to the Simple Linear Iterative Clustering (SLIC) algorithm, the FLICS algorithm utilizes active search strategies to fully utilize this prior information and has pixel label decision-making, which can quickly process each super-pixel block. By using the image round-trip traversal strategy, fully considering all information around the seed points, the segmentation effect is improved. Among them, the FLICS traversal image method is shown in Figure 2.

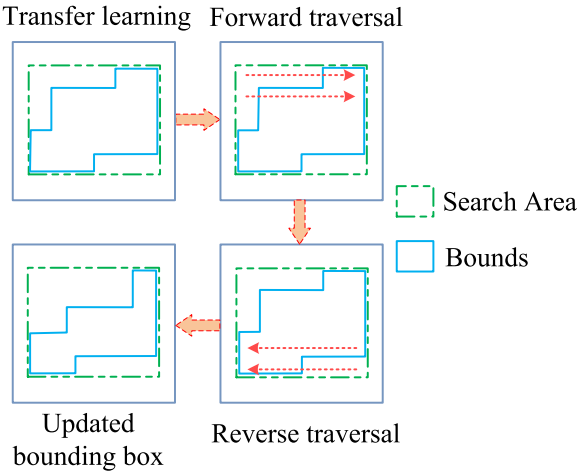

FIGURE 2. Traversalimage method.

In Figure 2, during forward traversal, the label of each pixel is determined by the information of the pixels around the top of the super pixel. During the reverse traversal process, information about the pixels around the bottom of the super-pixel will be obtained. Through this traversal method, all information around the seed point can be obtained. The FLICS algorithm utilizes joint allocation and update steps to improve execution efficiency and facilitate better selection of relevant pixels during label allocation. In this algorithm,

I_i represents the i-th pixel point, and the label classification criteria for I_i are denoted in equation (1).

$$L_i = \arg\min_{L_j} D\left(I_i, S_{L_j}\right), I_j \in A_i \tag{1}$$

In equation (1), L_i means the label of I_i. A_i denotes a set, which contains the four neighboring pixels of I_i; j stand for the sequence number, and S_{L_j} represents the seed point of I_j. D indicates the distance. For I_j and S_{L_j}, the expression for calculating the distance between the two is expressed in equation (2).

$$D\left(I_i, S_{L_j}\right) = \sqrt{d_c^2 + \left(\frac{d_s + m}{N_s}\right)^2} \tag{2}$$

In equation (2), it sets the weight of the spatial distance term to m, sets the clustering center to K. N_s refers to the edge length of the initial super-pixel block. It sets the color distance in Lab space to d_s, and sets the spatial distance to d_c. According to equation (1), designing a joint update strategy is beneficial for more precise execution of label allocation and update steps. After I_i is processed, its label changes to L_j, and the current seed point is updated through equation (3).

$$S_{L_i} = \frac{S_{L_i} * \left|\varsigma_{L_i}\right| - I_i}{\left|\varsigma_{L_i}\right| - 1} \tag{3}$$

In equation (3), ς_{L_i} stands fro a super-pixel block, and the amount of pixels in ς_{L_i} is labeled as $\left|\varsigma_{L_i}\right|$. S_{L_j} is updated, and the relevant calculation expression involved are shown in equation (4).

$$S_{L_j} = \frac{S^*_{L_j} \left|\varsigma_{L_j}\right| - I_i}{\left|\varsigma_{L_j}\right| - 1} \tag{4}$$

By updating equations (3) and (4), the arithmetic operations of these two equations are relatively simple and can be effectively executed. Color features are extracted and spatial color contrast is calculated in super-pixel blocks. Firstly, the input image is analyzed using the FLICS algorithm to obtain a uniform and compact super-pixel region r_k. By calculating the color distance between two super-pixel regions r_k and r_i, the significant value of r_k can be obtained. The method for calculating the color distance r_k between r_k and r_i is indicated in equation (5).

$$D_r\left(r_k, r_i\right) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} f\left(c_1, i\right) f\left(c_2, i\right) D\left(c_1, i, c_2, j\right) \tag{5}$$

In equation (5), n_k means the amount of colors in r_k, $k \in [1, 2]$; In r_k, the frequency of the ith color $c_{k,i}$ appearing is $f\left(c_k, i\right)$. In the process of calculating the significance value of r_k, the position information of the super pixel region is introduced to enhance the effective impact of the spatial information of the region on the significance detection results, to enhance the influence of the regions closer to the region and weaken the influence of the regions farther away. It sets the strength of spatial weights to σ_s^2, and the larger the value, the smaller the impact of spatial weights. The value of σ_s^2 in the study is 0.4. Using the control term $w_s\left(r_k\right)$ can control the attractiveness of non image boundary regions to human vision, and the calculation method is shown in equation (6).

$$w_s\left(r_k\right) = \exp\left(-9d_k^2\right) \tag{6}$$

In equation (6), in the normalized coordinate system, the average Euclidean distance between pixels in super-pixel region r_k and the center of the image is d_k. The saliency map of image color features is obtained by calculating the color distance and spatial distance weighting of the super pixel region. When extracting texture features from images, Gabor filters are used [17], [18], [19]. Firstly, a Gabor filter is designed that includes 24 Gabor filters that can be tuned to different frequencies and directions. In the input image, it designs these frequencies and directions as different, orthogonal subsets of localization frequency and direction information. According to the direction, the Gabor filter is divided into four groups, with directions of 0°, 45°, 90°, 135°, respectively. In each saliency, the filter wavelength starts from $4/sqrt\,(2)$, and the latter filter wavelength is twice the wavelength of the previous filter until the last wavelength is $2^5 * 4/sqrt\,(2)$. In a smaller input image, its maximum wavelength is the diagonal length of the image. The amplitude response of each Gabor filter is extracted. The outputting of this filter is the response of the corresponding frequency information in the image, and the output of this filter is selected as the feature for classification. Among the selected classification features, Gaussian smoothing is required to add relevant spatial information to the feature set, reshape the feature set, and make it suitable for principal component analysis and Kmeans function. It standardizes feature information to convert it into the form of common variance and mean. A Gaussian low-pass filter is utilized to solve the problem of local changes in the amplitude response of Gabor filters. Mapping of position information is added in XY space and Gabor feature sets are established for classification. The data are reshaped to make it suitable for the K-means algorithm. The K-means is repeated five times to prevent the occurrence of local minima. After following these steps, the texture feature map of the original image can be gained.

By using the FLICS algorithm to segment the texture feature map of the original image, uniform and compact super-pixel blocks are obtained. The ratio of the length between the image region and boundary to the square root of the area of the region is defined as connectivity, and its geometric explanation is shown in Figure 3.

In Figure 3, the yellow region is represented as a prominent region, which corresponds to a smaller connectivity of 0.28. The purple and blue areas are set as non-prominent areas, which are more related to the image boundary and have a higher connectivity. The boundary connectivity of texture feature maps is calculated. Firstly, in the Lab color space, p and q represent two adjacent super-pixel blocks, and the average color distance $d_{geo}\left(p, q\right)$ between these two

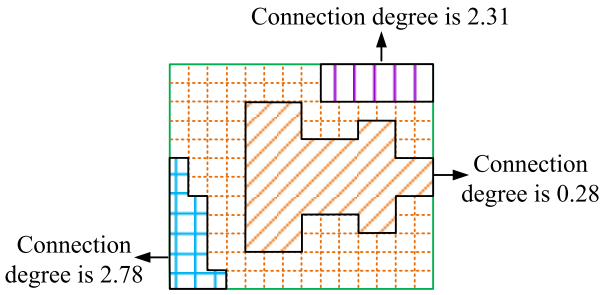

FIGURE 3. Geometric interpretation of boundary connectivity in image regions.

super-pixel blocks is calculated using equation (7).

$$d_{geo}(p,q) = \min_{p_1=p,p_2,\cdots,p_n=q} \sum_{i=1}^{n-1} d_{app}(p_i, p_{i+1}) \tag{7}$$

In equation (7), n means the amount of super-pixel blocks. It sets the contribution of super-pixel p_i to p to $S'(p, p_i)$, with a value range of 0 to 1. When both p_i and p are located in the smooth region of the texture, $d_{ist}(p, q)$ value is 0, $S'(p, p_i) = 1$, to ensure that the area of p_i is fully added to p. When p_i and p are located in different regions of the image, there is at least one larger boundary distance d_{app} in the shortest path boundary distance between p_i and p, where $d_{app} \gg 3\sigma_{clr}$ and $S'(p, p_i)$ are approximately 0. σ_{clr} represents a parameter. After experiments, the effect of $\sigma_{clr} \in [5, 15]$ is good, and for this, it sets $\sigma_{clr} = 12$. The boundary length of p is set to $Len_{bnd}(p)$, and for p, its new boundary connectivity is indicated in equation (8).

$$BndCon(p) = \frac{Len_{bnd}(p)}{\sqrt{Area(p)}} \tag{8}$$

In equation (8), Bnd refers to the set of all image boundaries; $BndCon(p)$ denotes the boundary connectivity of p; The generalized area of p is marked as $Area(p)$. When calculating equation (8), the Johnson algorithm is used to solve the shortest path between all super-pixel blocks [20]. Based on the above boundary connectivity, it adds background weighted contrast. Through this contrast, effective background information can be provided for significance estimation. The weighted contrast of the background is represented as $Ctr(p)$, and the background probability ω_i^{bg} is mapped from the connected values of the image boundary of p. The larger the boundary connectivity value, the closer the ω_i^{bg} value is to 1. The relevant mathematical expression of $Ctr(p)$ is shown in equation (9).

$$\omega Ctr(p) = \sum_{i=1}^{n} d_{app}(p, p_i)\omega_{spa}(p, p_i)\omega_i^{bg} \tag{9}$$

In equation (9), ω, ω_{spa}, σ_{spa} refer to parameters; $\omega_{spa}(p, p_i) = \exp\left(-d_{spa}^2(p, p_i) \Big/ 2\sigma_{spa}^2\right)$, $d_{spa}(p, p_i)$ mean the distance between p and the center of p_i, and the value of σ_{spa} is taken as 0.25. On this basis, the saliency map of the image texture feature channel can be got. In the saliency map with color features, its contrast map is higher, and the saliency map with texture features can accurately locate the prominent target position. For this, these two types of graphs are merged, and the relevant calculation method is shown in equation (10).

$$S^* = \alpha S_1 + (1 - \alpha) S_2 \tag{10}$$

In equation (10), the final integrated saliency map is marked as S^*; S_1 and S_2 stand for the texture and color feature saliency map. In fact, S^*, α mean parameters, $\alpha \in [0, 1]$, through the contrast between the linear weighted color feature channel and the texture feature channel. On this basis, the obtained fusion saliency map is optimized. First, it sets the objective function J, and uses J to get the optimal saliency map. See equation (11) for the relevant mathematical expression.

$$J = \sum_{i=1}^{n} \omega_i^{bg} s_i^2 + \sum_{i=1}^{n} \omega_i^{fg} (s_i - 1)^2 + \sum_{i=1} \omega_{i,j} \left(s_i - s_j\right)^2 \tag{11}$$

In equation (11), j represents the j super-pixel. s_i means the significant value of the ith super-pixel, and the probability that the p_i in the target area belongs to the target is expressed as ω_i^{fg}. The larger the ω_i^{fg}, the closer the s_i is to 1. In the background area, ω_i^{bg} is larger and s_i is closer to 0. $\sum_{i=1} \omega_{i,j} \left(s_i - s_j\right)^2$ means a smoothing term, which ensures the continuity of significance values. $\omega_{i,j}$ expresses the weight value, which is used to eliminate small noise in the background and foreground through σ_{clr}, with a value of 10.

B. APPLICATION OF FLICSP ALGORITHM IN ISD

Considering that there are relatively few methods for combining bottom-up and top-down models, from the perspective of combining the two models, the saliency map obtained by the FLICS algorithm is used as a prior information map, and deep convolutional networks are used to extract deep uncertain features of the image. Adopting a guided learning approach to further obtain more accurate saliency maps. Given the high universality of the Visual Geometry Group (VGG) 16 model in image feature extraction, it is applied in research [21], [22], [23]. Based on the actual situation of the study, a correlation saliency detection model is constructed with the VGG16 model, as denoted in Figure 4.

In Figure 4, the original image and its true value image are utilized as inputs, and the input for network testing is the original image. The first three pooling layers are retained, and after each module's pooling processing, the output feature map size gradually decreases. Through deconvolution operation, the data of each convolution module is sampled to ensure that the output feature model feature map and input image have the same size. The feature map is transformed into a feature vector through two convolutional layers containing 512 neurons. It sets this network as the feature mapping function ϕ with parameter θ. In image X_m, the pixel is set

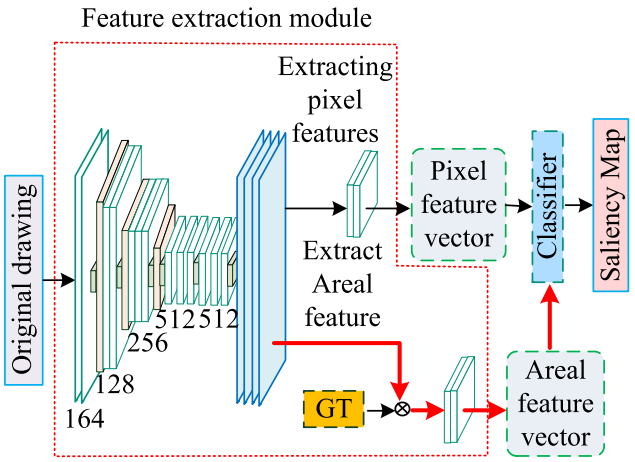

FIGURE 4. Correlation significance detection model.

to x_{mn} and the feature vector to $\phi_{mn} = \phi(x_{mn}, \theta)$. The true value image area is set to C_{mk}. By using this network to extract features, the subsequent convolutional layer settings are the same as the pixels, and the corresponding feature vectors are obtained. The mathematical expression is shown in equation (12).

$$\lambda_{mk} = \varphi(C_{mk}; \eta) \tag{12}$$

In equation (12), η means the parameter of the network, $k \in [1, 2]$. The feature vector of C_{mk} is represented as λ_{mk}, and φ expresses the feature mapping function. In X_m, its foreground area is set as C_{m1} and its background area as C_{m2}, $X_m = C_{m1} \cup C_{m2}$. After these feature extraction steps, the x_{mn} and C_{mk} features are mapped to the feature metric space of the same dimension respectively. Due to $x_{mn} \in C_{m1}$ or $x_{mn} \in C_{m2}$, the calculated Euclidean distance between ϕ_{mn} and λ_{mk} can be used as the membership degree of the pixel belonging to the region, and its mathematical expression is shown in equation (13).

$$p(C_{mk}|x_{mn}) = \frac{\exp\{-d(\phi_{mn}, \lambda_{mk})\}}{\sum_{j'} \exp\{-d(\phi_{mn}, \lambda_{mj})\}} \tag{13}$$

In equation (13), $p(C_{mk}|x_{mn})$ means the membership degree of x_{mn} belonging to C_{mk}, and $k \in [1, 2]$ denotes the corresponding area as background or foreground, j' represents the total number of image regions. When training the network, it is necessary to continuously optimize the objective function Γ through gradient descent algorithm to achieve or approach the optimal $p(C_{mk}|x_{mn})$. The relevant mathematical expressions involved are indicated in equation (14).

$$\Gamma = \sum_{m,n} t_{mn} \log p(C_{m1}|x_{mn}) + (1 - t_{mn}) p(C_{m2}|x_{mn}) \tag{14}$$

In equation (14), the saliency label of x_{mn} is marked as t_{mn}, and in $x_{mn} \in C_{m1}$ and $t_{mn} = 1$, x_{mn} belongs to the saliency region. In $x_{mn} \in C_{m2}$ and $t_{mn} = 0$, x_{mn} belongs to the image background. In the model constructed in Figure 4, its core is the feature extraction module. It contains two main

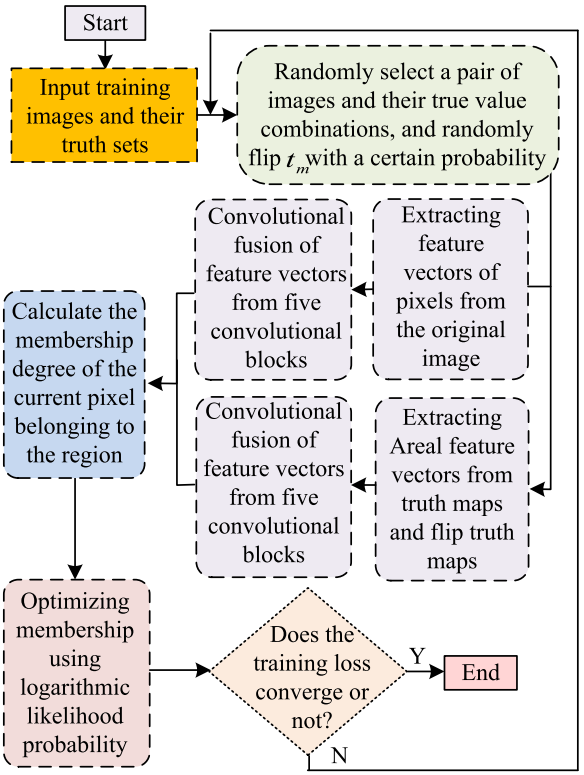

FIGURE 5. Training process.

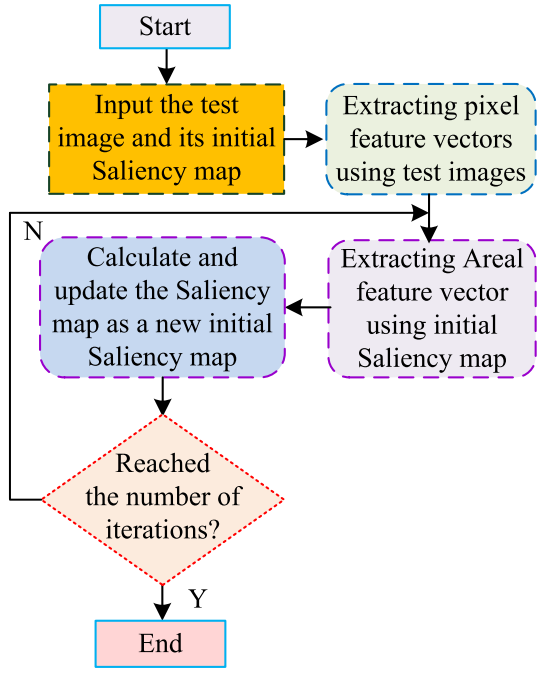

FIGURE 6. Network testing process.

operations, namely 1 * 1 convolution and deconvolution. In the convolutional kernel of 1*1, due to its size being only 1*1, the influence of neighboring pixels is not considered when performing related operations. By adjusting the amount of feature maps output by the convolutional verification

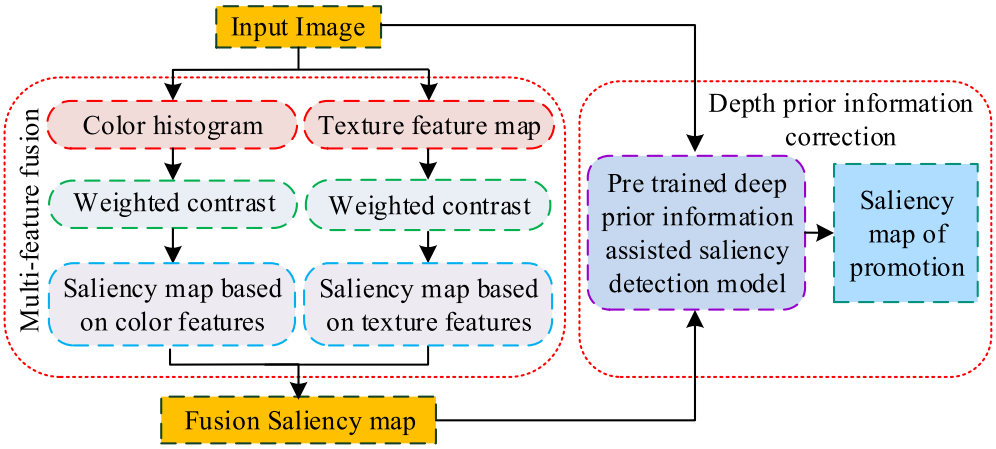

FIGURE 7. FLICSP algorithm process.

TABLE 1. Information on relevant algorithms.

Project	MR algorithm	RC algorithm	SF algorithm	FLICS algorithm
Processing unit	Super-pixel	Super-pixel	Super-pixel	Super-pixel
Contrast ratio	Overall situation	Local and Global	Local and Global	Local and Global
Color space	Lab	Lab	Lab	Lab
Resolving power	1	1	1	1
Spatial relationship	Have	Have	Have	Have
Project	DSR algorithm	GS algorithm	HC algorithm	FLICSP algorithm
Processing unit	Super-pixel	Super-pixel	Whole image	Super-pixel
Contrast ratio	Local	Local and Global	Local	Local and Global
Color space	Lab	Lab	Lab	Lab
Resolving power	1	1	1	1
Spatial relationship	Have	Have	Not have	Have

module, the corresponding pixel points on different feature channels are non-linearly combined. Through this operation, the nonlinear expression ability of the network is enhanced, achieving the goal of dimensionality reduction while reducing parameters. In deconvolution, it is the back-propagation of the convolution. It is worth noting that after the deconvolution operation, the input image that is exactly the same as the convolution operation cannot be obtained. It can only increase the size of the output feature map. In the process of deconvolution implementation, the input image is first filled, and the convolution kernel is utilized for convolution to obtain an output twice the size of the original image through cropping operations. It selects PyTorch and trains and tests the model using a server with 256GB of memory. In network training, the parameters for loading five convolutional blocks are specified in PyTorch, the newly added convolutional layers are randomly initialized, and feature vectors are fused. The original and truth images of the DUT dataset are separated into training and validation sets, and trained 32 epochs. The training is expressed in Figure 5.

In Figure 5, the training image and its truth set are used as input for training. A pair of images and their truth combinations are randomly selected, and t_m is randomly flipped with a certain probability to extract the feature vectors of pixels and regions. Convolution fusion is performed on the feature vectors of five convolutional blocks, and the membership degree of the current pixel belonging to the region is calculated. The membership degree is optimized through logarithmic likelihood probability to determine whether the training loss converges. When it converges, the training ends, and vice versa, a pair of images and their true value combinations are selected again. In actual training, the initial learning rate is set to 0.001, the random turnover probability to 0.05, and the batch-size to 3. When conducting network testing, the testing process is shown in Figure 6.

In Figure 6, the original image and the initial saliency map are taken as the input of the network, and pixel feature vectors are extracted through the test image. The initial saliency map is selected to extract the areal feature vector. For the given initial saliency map, the new one is constructed, and iterative correction is made through equation (15). The related calculation method is shown in equation (15).

$$V_m^{(t+1)} = \frac{t}{t+1} V_m^{(t)} + \frac{1}{t+1} U_m^{(t)} \tag{15}$$

In equation (15), the iteration times is set to t; The input initial saliency map is set to $V_m^{(t)}$; Saliency map is built according to the initial saliency map, and it expressed as $U_m^{(t)}$; $V_m^{(t+1)}$ denotes the newly constructed saliency map. When the iteration times is reached, it needs to stop the operation, or re-extract the areal feature vector. Due to the unsatisfactory testing results, a hybrid saliency detection

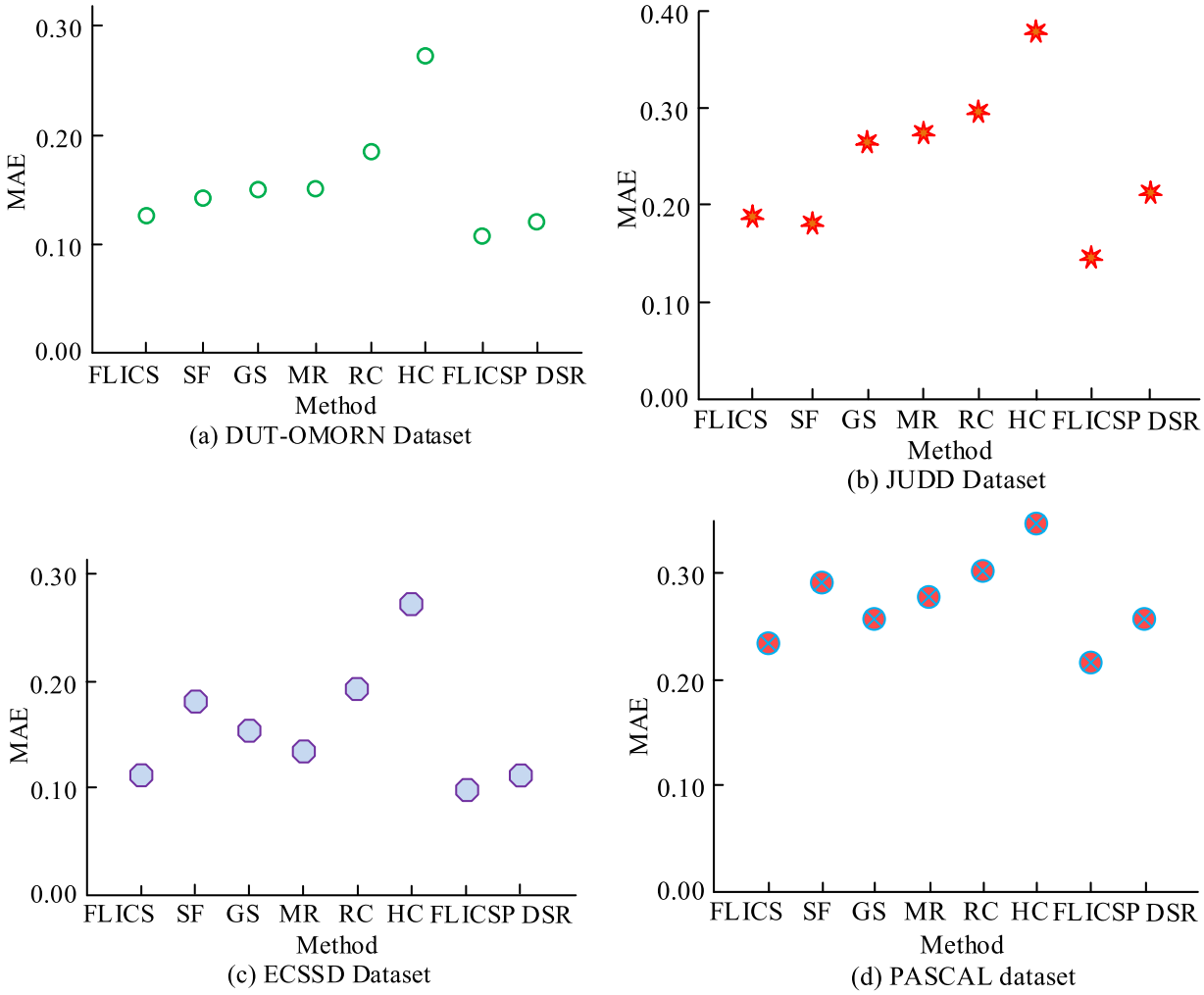

FIGURE 8. MAE values of related algorithms.

algorithm is obtained by combining the FLICS algorithm with the algorithm in Figure 4. For the convenience of writing, the hybrid algorithm is set as the FLICSP algorithm, and its related process is denoted in Figure 7.

In Figure 7, the FLICSP algorithm mainly includes two parts: multi feature fusion and deep prior information correction, which are executed separately. Among them, the FLICS algorithm is used for multi feature fusion, which combines the color feature saliency map and texture feature saliency map of the image. This part is completely bottom-up. The depth prior information correction section is done through pre-training the model, which is top-down. In the pre-training model, the model input is the original image and the fused saliency map. After iteration, the input of saliency map will get better results.

IV. APPLICATION ANALYSIS OF ISD BASED ON FLICSP ALGORITHM

In ISD, to analyze the detection effect of the FLICSP algorithm constructed in the study, a comparative algorithm was selected. In addition to the FLICS algorithm, six algorithms such as the significance detection algorithm based on histogram contrast (HC) were also selected for comparison. Precision Recall (PR) curve and F-value were selected as evaluation indicators, among which, the F value means the weighted average of precision and recall, and this indicator can be applied to evaluate the comprehensive quality of significance testing results. By analyzing the ISD of the algorithm using these indicators, the selected datasets were the DUT-OMRON, JUDD, ECSSD and PASCAL. The DUT-OMRON dataset contained the most natural images, with 5166, followed by the ECSSD dataset, all of which contained natural images and their truth images.

A. COMPARISON OF MAE AND PR CURVES AND F-VALUES OF DIFFERENT SALIENCY DETECTION ALGORITHMS

The mean absolute error (MAE) and PR curves of different significance detection algorithms were analyzed. Among them, the relevant information of the algorithm is shown in Table 1.

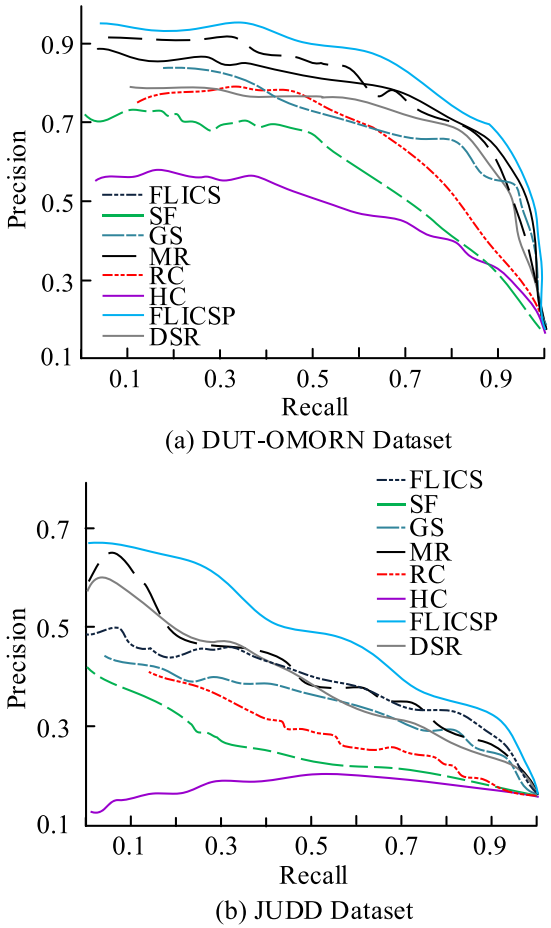

FIGURE 9. PR curve of algorithm.

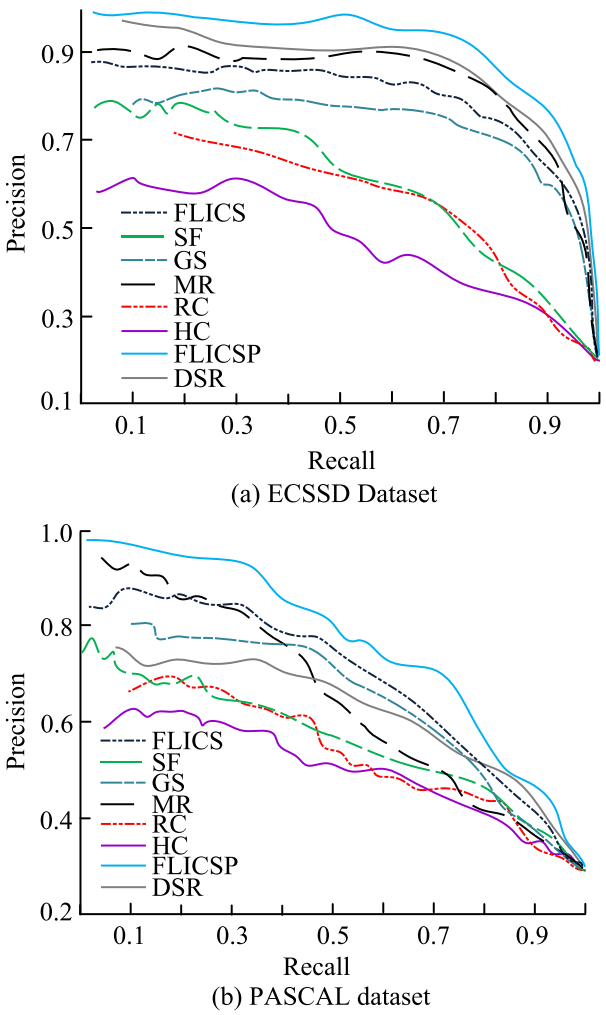

FIGURE 10. PR curve of algorithm.

In Table 1, the RC algorithm represents a detection algorithm based on regional contrast [24]. The SF algorithm represents a saliency detection algorithm based on contrast filtering [25]. The MR algorithm denotes a saliency detection algorithm based on graph theory manifold sorting [26]. The GS algorithm means the geodesic saliency detection algorithm [27]. The DSR algorithm expresses a saliency detection algorithm based on dense and sparse reconstruction [28]. There were some differences in the relevant information of the six comparison algorithms. By selecting four datasets and analyzing the MAE values of different saliency detection algorithms, the relevant results are shown in Figure 8.

In Figure 8, there were differences in MAE values for different algorithms under different datasets. In Figure 8 (a), overall, the FLICSP algorithm had the smallest MAE value, while the HC algorithm had the largest MAE value. Among them, the MAE value of the FLICS algorithm was 0.122, which was 0.019 less than the SF algorithm, while the MAE value of the latter was 0.141. The MAE value of MR was 0.152, which was 0.121 less than that of HC algorithm, while the MAE value of the latter was 0.273. The MAE value of the FLICSP algorithm was 0.115, which was 0.007 less than the FLICS algorithm. In Figure 8 (b), the MAE value of the HC algorithm was 0.370, and the MAE value of the FLICSP algorithm was 0.143. The former was 0.227 larger than the latter. The MAE value of the RC algorithm was smaller than that of the HC algorithm, with the former being 0.288. The MAE value of the FLICS algorithm was 0.177. In Figure 8 (c), the minimum MAE value of the FLICSP algorithm was 0.094, followed by the DSR algorithm with a MAE value of 0.115. In Figure 8 (d), the MAE values of the DSR algorithm and the FLICS algorithm were 0.257 and 0.238, respectively, while the MAE values of the FLICSP algorithm were 0.213. Compared to other algorithms, the FLICSP algorithm had better detection performance. The PR curves of different algorithms were analyzed, and the findings are indicated in Figure 9.

In Figure 9, compared to the JUDD dataset, the PR curve of the algorithm in the DUT-OMRON dataset was closer to the upper right corner of the graph. Unlike other algorithms, overall, the FLICSP algorithm achieved first place on two datasets. In Figure 9 (a), the PR curve of the FLICSP algorithm was located above the PR curve of other algorithms, followed by the MR algorithm. The area of the PR

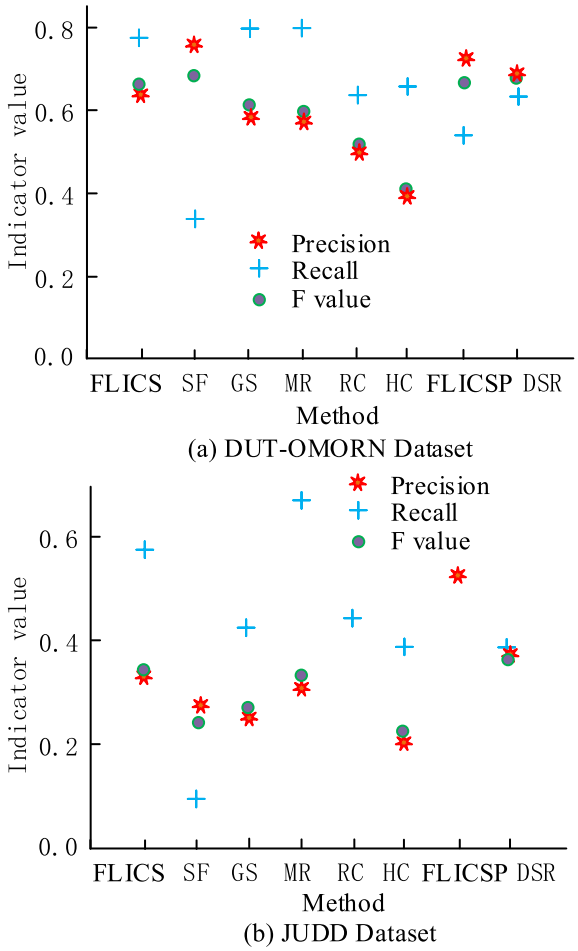

FIGURE 11. The F-values and other situations of different algorithms.

curve was only second to that of the FLICSP algorithm, while the PR curve of HC was located at the bottom, with the smallest area. Compared to Figure 9 (a), there was a certain variation in the size relationship between the PR curve areas of different algorithms in Figure 9 (b). In Figure 9 (b), the PR curve area of the GS algorithm was larger than that of the RC algorithm, while the PR curve of the FLICSP algorithm remained above other algorithms, and ranked fourth. The FLICSP algorithm had the best performance. It selected the ECSSD dataset and PASCAL dataset to analyze the PR curve of the algorithm, as shown in Figure 10.

In Figure 10 (a), different algorithms resulted in differences in the corresponding PR curves. The PR curve of the HC algorithm was at the bottom, followed by the RC algorithm, while the PR curve of the FLICSP algorithm was at the top. Compared to the MR algorithm, the PR curve of the DSR algorithm ranked higher, second only to the FLICSP algorithm. The FLICS algorithm's PR curve ranked fourth. In Figure 10 (b), the position of the algorithm's PR curve changed, with the FLICS algorithm's PR curve ranking second and the FLICSP algorithm's PR curve still at the top. The PR curve of the DSR algorithm ranked fifth. It analyzed the F-values of different algorithms, and the results are shown in Figure 11.

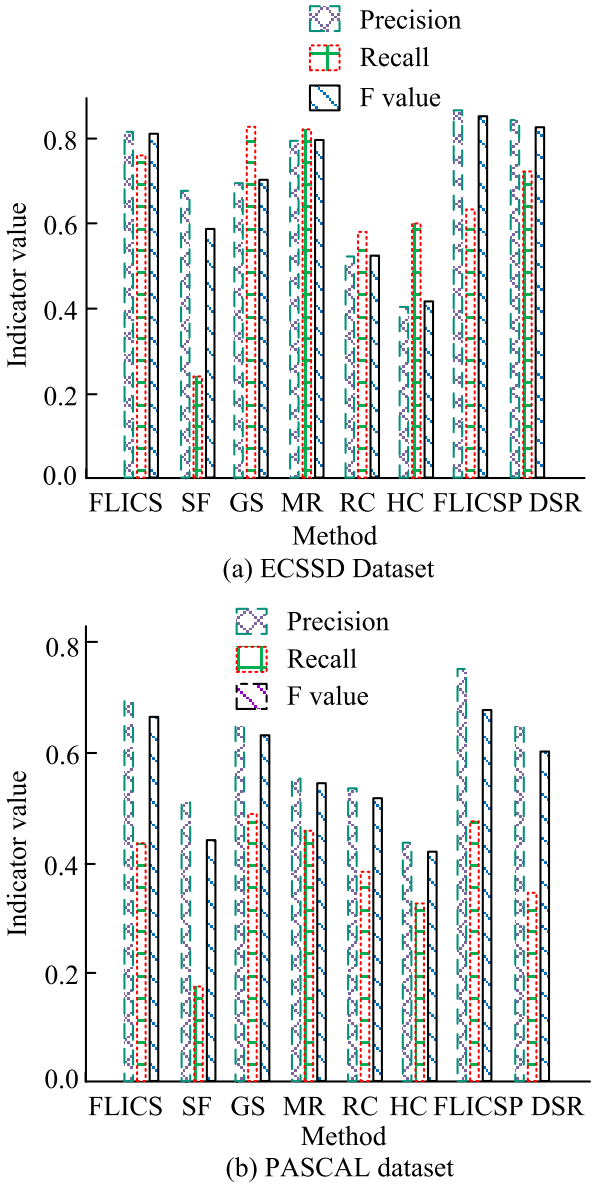

FIGURE 12. F-values and other situations of different algorithms under different datasets.

In Figure 11, there were differences in the precision, recall, and F-value of different algorithms, and the precision to some extent determined the size of the F-value. In Figure 11 (a), the HC algorithm had the lowest precision and F-value, with 0.391 and 0.403 respectively, and its recall rate was 0.657. The FLICSP algorithm ranked second with a precision of 0.736, 0.049 higher than the DSR algorithm, while the latter had a precision of 0.687. The maximum precision of the SF algorithm was 0.768, with an F-value of 0.686. In Figure 11 (b), the FLICSP algorithm had the highest precision and F-value, with 0.534 and 0.506, respectively. Its precision was 0.205 higher than the FLICS algorithm, which was 0.329. The F-value of the DSR algorithm ranked second, at 0.367. According to the results in Figure 11, the FLICSP algorithm performed well. It selected the ECSSD dataset and

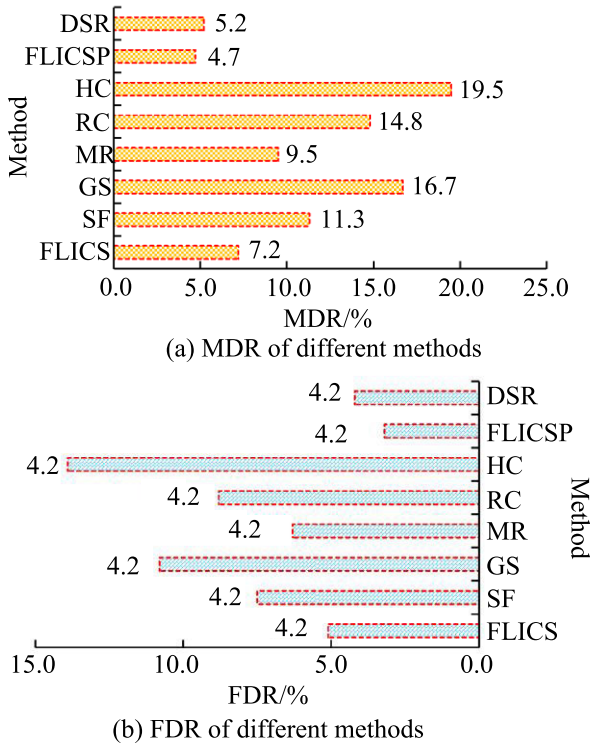

FIGURE 13. The FDR and MDR of the algorithm.

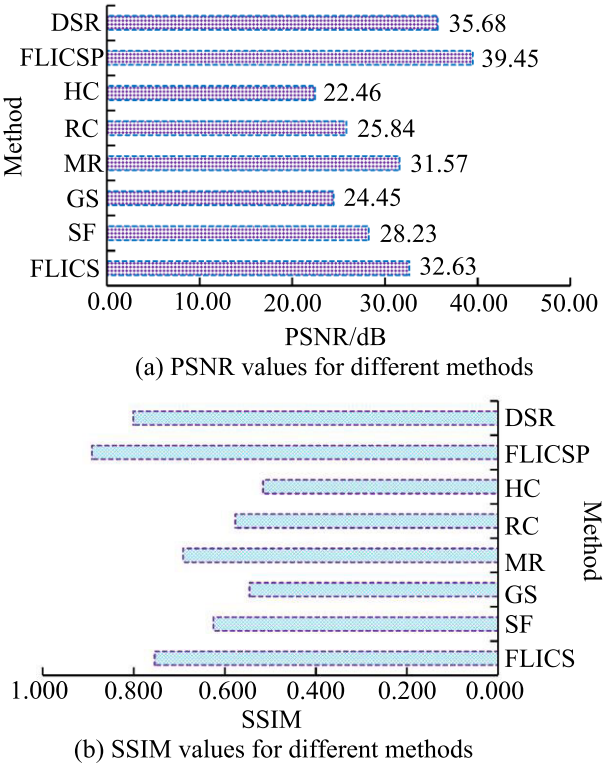

FIGURE 14. Image quality evaluation.

PASCAL dataset to analyze the F-value of the algorithm, as shown in Figure 12.

In Figure 12, the FLICSP algorithm had the highest precision and F-value, indicating good detection performance. In Figure 12 (a), the precision and recall of the FLICS algorithm were 0.823 and 0.776, respectively, which were 0.304 higher than the precision of the HC algorithm, which was 0.519. The FLICSP algorithm had the highest precision at 0.874, with an F-value of 0.851, while the DSR algorithm ranked second with a precision of 0.842. The precision of the MR algorithm ranked fourth with a value of 0.797, an F-value of 0.800, and a recall rate of 0.828. In Figure 12 (b), the FLICSP algorithm had the highest precision and F-value, followed by the FLICS algorithm, the DSR algorithm ranked third, and the HC algorithm ranked last. In algorithm precision, the FLICSP algorithm had a precision of 0.786, which was 0.086 higher than the FLICS algorithm, while the latter had a precision of 0.700. The precision of the DSR algorithm was 0.665, 0.07 higher than the GS algorithm, which was 0.0658. In F-values, the F-values of FLICSP and FLICS algorithms were 0.754 and 0.678, respectively, with the former being 0.076 larger than the latter. The HC algorithm had the smallest F-value of 0.425, which was 0.094 smaller than the RC algorithm, while the latter had an F-value of 0.519. The F-value of the GS algorithm was 0.634. In the recall rate, the minimum recall rate of the SF algorithm was 0.182. Through the analysis of a large number of experimental results, it was found that only extracting image features that have been discovered and can be modeled cannot obtain more advanced image features. This indicated that classic bottom-up detection algorithms have reached a bottleneck period, and the rise of deep learning has brought possibilities for the representation and extraction of advanced image features.

B. DETECTION EFFECTIVENESS AND IMAGE QUALITY EVALUATION OF DIFFERENT SALIENCY DETECTION ALGORITHMS

It analyzed the detection effect of different saliency detection algorithms and selected the JUDD dataset for testing. The image content in this dataset was relatively complex, and its color tone was relatively single. It evaluated the detection performance of the algorithm through false detection rate (FDR) and missed detection rate (MDR). The specific results are denoted in Figure 13.

In Figure 13, different algorithms had different FDR and MDR. In Figure 13 (a), the MDR of the FLICSP algorithm was 4.7%, which was 2.5% lower than the FLICS algorithm, which was 7.2%. The MDRs of DSR algorithm and SF algorithm were 5.2% and 11.3%, respectively, with the former being 6.1% lower than the latter. The algorithm with the highest MDR was the HC algorithm, which had a MDR of 19.5%, 14.8% higher than the FLICSP algorithm, and the latter had the lowest MDR. In Figure 13 (b), the FLICSP algorithm had the lowest FDR, followed by the DSR algorithm, and the HC algorithm had the highest FDR. The FDR of the DSR algorithm was 4.2%, while the FDR of

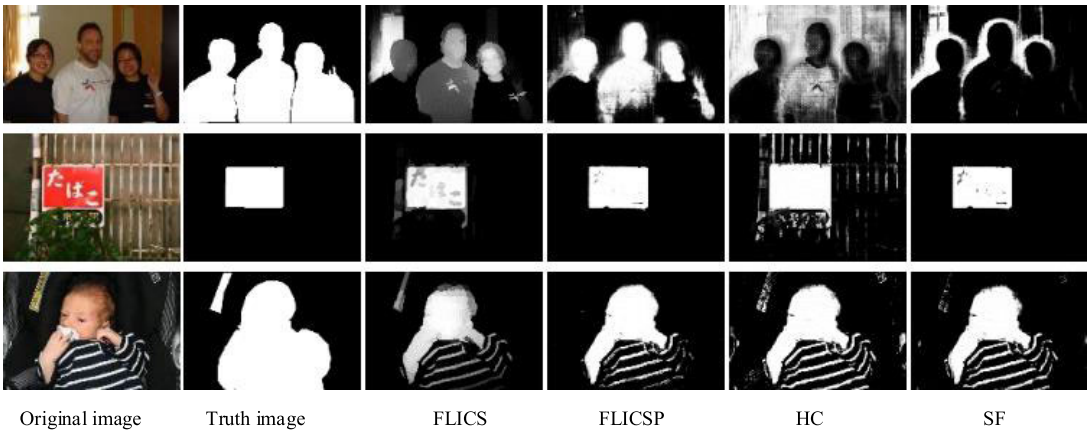

FIGURE 15. Related result graph.

the FLICSP algorithm was 3.2%, which was 1.0% lower than the former. The FDR of GS algorithm was 10.8%, which was 3.1% lower than HC algorithm. Overall, the FLICSP algorithm performed well in detection. It selected peak signal-to-noise-ratio (PSNR) and structural similarity index measure (SSIM) to evaluate the significance image quality under different algorithms, as shown in Figure 14.

In Figure 14, the PSNR and SSIM index values of saliency images varied under different algorithms. For Figure 14 (a), in the PSNR indicator, the HC algorithm had a minimum image PSNR of 22.46dB, followed by the GS algorithm, which had an image PSNR of 24.45 dB; The image PSNR of FLICSP and DSR algorithms was 39.45 dB and 35.68 dB respectively, with FLICSP algorithm having the highest PSNR. In Figure 14 (b), the maximum SSIM value of the FLICSP algorithm was 0.892, which was 0.138 higher than the FLICS algorithm. The SSIM value of the SF and HC algorithms was 0.625 and 0.516, respectively. The saliency image quality was the best under the FLICS algorithm. Select two datasets, ECSSD and PASCAL, and select different algorithms to process the original images in the dataset. The results of the original images, truth images, and algorithms are compared, as shown in Figure 15.

From Figure 15, it can be seen that each result graph of the FLICSP algorithm is significantly closer to the true value than the FLICS, HC, and SF algorithms, which to some extent indicates that the FLICSP algorithm can more fully utilize image features. And from the extraction of significant regions in the figure, it can be seen that the FLICSP algorithm performs better, that is, its performance is better. The research method has solved the problem of insufficient utilization of image features in previous significance detection, and improved the quality of significance detection.

V. CONCLUSION

To improve the effect of ISD, FLICS algorithm was used to segment the image, extract color and texture features, and obtain the corresponding saliency map. Contrast calculation was carried out, and linear fusion and saliency optimization were performed on the obtained saliency map. Then the fused feature saliency map was obtained. On this basis, the algorithm was combined with a saliency detection model based on deep prior information assistance to obtain the FLICSP algorithm, and experimental analysis was conducted. The experimental results showed that compared with other algorithms, the FLICSP algorithm had higher detection accuracy, better detection effect, and better image quality. The FLICSP algorithm had the smallest MAE value in the DUT-OMRON dataset, which was 0.115 and 0.007 less than the FLICS algorithm. Compared with other algorithms, FLICS algorithm had higher accuracy and F-value. In the JUDD dataset, the FLICSP algorithm had the highest accuracy and F-value, with values of 0.534 and 0.506, respectively, which was 0.205 higher than the FLICS algorithm. Unlike other algorithms, FLICS algorithm had a lower MDR of 4.7%, which was 2.5% lower than FLICS algorithm. The SSIM value of the FLICSP algorithm was the highest. The maximum SSIM value of FLICSP algorithm was 0.892, which was 0.138 higher than FLICS algorithm, while the SSIM value of HC algorithm was 0.516. From this, the research method had good application effects. There are certain shortcomings in the research, and only the saliency detection methods have been explored. In the future, the research methods will be applied to tasks such as image compression.

REFERENCES

[1] X. Li, W. Zhang, X. Sun, and X. Gao, "Semantic-meshed and content-guided transformer for image captioning," *IET Comput. Vis.*, vol. 16, no. 5, pp. 431–444, Aug. 2022, doi: 10.1049/cvi2.12099.

[2] T. Cao, W. Wang, S. Tighe, and S. Wang, "Crack image detection based on fractional differential and fractal dimension," *IET Comput. Vis.*, vol. 13, no. 1, pp. 79–85, Jan. 2019, doi: 10.1049/iet-cvi.2018.5337.

[3] S. Oslund, C. Washington, A. So, T. Chen, and H. Ji, "Multiview robust adversarial stickers for arbitrary objects in the physical world," *J. Comput. Cogn. Eng.*, vol. 1, no. 4, pp. 152–158, Sep. 2022, doi: 10.47852/bonviewJCCE2202322.

[4] S. S. Patterson, J. A. Kuchenbecker, J. R. Anderson, M. Neitz, and J. Neitz, "A color vision circuit for non-image-forming vision in the primate retina," *Current Biol.*, vol. 30, no. 7, pp. 1269–1274, Apr. 2020, doi: 10.1016/j.cub.2020.01.040.

[5] M. Zhang, J. Chen, P. Li, M. Jiang, and Z. Zhou, "Topic scene graphs for image captioning," *IET Comput. Vis.*, vol. 16, no. 4, pp. 364–375, Feb. 2022, doi: 10.1049/cvi2.12093.
[6] Y. Zhang, Q. Guo, Y. Zhang, and C. Zhang, "Fast and robust superpixel generation method," *IET Image Process.*, vol. 14, no. 17, pp. 4543–4553, Dec. 2020, doi: 10.1049/iet-ipr.2020.1179.
[7] Z. Lin, Z. Yingjie, D. Bochao, C. Bo, and L. Yangfan, "Welding defect detection based on local image enhancement," *IET Image Process.*, vol. 13, no. 13, pp. 2647–2658, Nov. 2019, doi: 10.1049/iet-ipr.2018.5840.
[8] K. Xu, H. Gong, and F. Liu, "Vehicle detection based on improved multitask cascaded convolutional neural network and mixed image enhancement," *IET Image Process.*, vol. 14, no. 17, pp. 4621–4632, Dec. 2020, doi: 10.1049/iet-ipr.2020.1005.
[9] Y. Zhou, X. Liu, Y. Zhang, H. Yin, and Y. Lu, "Salient object detection via reliability-based depth compactness and depth contrast," *IET Image Process.*, vol. 14, no. 14, pp. 3623–3631, Dec. 2020, doi: 10.1049/iet-ipr.2019.1495.
[10] G. Wang, Z. Wang, K. Jiang, B. Huang, Z. He, and R. Hu, "Silicone mask face anti-spoofing detection based on visual saliency and facial motion," *Neurocomputing*, vol. 458, pp. 416–427, Oct. 2021, doi: 10.1016/j.neucom.2021.06.033.
[11] M. Singh, M. C. Govil, E. S. Pilli, and S. K. Vipparthi, "SOD-CED: Salient object detection for noisy images using convolution encoder–decoder," *IET Comput. Vis.*, vol. 13, no. 6, pp. 578–587, Sep. 2019, doi: 10.1049/iet-cvi.2018.5814.
[12] M. Singh, M. C. Govil, and E. S. Pilli, "CHACT: Convex hull enabled active contour technique for salient object detection," *IEEE Access*, vol. 6, pp. 22441–22451, 2018, doi: 10.1109/ACCESS.2018.2826924.
[13] W. Cui, Q. Zhang, and B. Zuo, "Deep saliency detection via spatial-wise dilated convolutional attention," *Neurocomputing*, vol. 445, pp. 35–49, Jul. 2021, doi: 10.1016/j.neucom.2021.02.061.
[14] S. Serte and A. Serener, "Graph-based saliency and ensembles of convolutional neural networks for glaucoma detection," *IET Image Process.*, vol. 15, no. 3, pp. 797–804, Feb. 2021, doi: 10.1049/ipr2.12063.
[15] X. Lin, Y. Tang, H. Tianfield, F. Qian, and W. Zhong, "A novel approach to reconstruction based saliency detection via convolutional neural network stacked with auto-encoder," *Neurocomputing*, vol. 349, pp. 145–155, Jul. 2019, doi: 10.1016/j.neucom.2019.01.041.
[16] L. Wei, S. Zhao, O. E. F. Bourahla, X. Li, F. Wu, and Y. Zhuang, "Deep group-wise fully convolutional network for co-saliency detection with graph propagation," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 5052–5063, Oct. 2019, doi: 10.1109/TIP.2019.2909649.
[17] A. K. Samantaray and A. D. Rahulkar, "New design of adaptive Gabor wavelet filter bank for medical image retrieval," *IET Image Process.*, vol. 14, no. 4, pp. 679–687, Mar. 2020, doi: 10.1049/iet-ipr.2019.1024.
[18] H. Luqman, E.-S.-M. El-Alfy, and G. M. BinMakhashen, "Joint space representation and recognition of sign language fingerspelling using Gabor filter and convolutional neural network," *Multimedia Tools Appl.*, vol. 80, no. 7, pp. 10213–10234, Mar. 2021, doi: 10.1007/s11042-020-09994-0.
[19] T. Zhang, P. Zhang, W. Zhong, Z. Yang, and F. Yang, "JL-GFDN: A novel Gabor filter-based deep network using joint spectral–spatial local binary pattern for hyperspectral image classification," *Remote Sens.*, vol. 12, no. 12, p. 2016, Jun. 2020, doi: 10.3390/rs12122016.
[20] A. Ali, V. R. Hulipalled, S. S. Patil, and R. A. Kappaparambil, "DPCCG-EJA: Detection of key pathways and cervical cancer related genes using enhanced Johnson's algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 28, no. 1, pp. 124–138, Jan. 2019.
[21] S. Pasban, S. Mohamadzadeh, J. Zeraatkar-Moghaddam, and A. K. Shafiei, "Infant brain segmentation based on a combination of VGG-16 and U-Net deep neural networks," *IET Image Process.*, vol. 14, no. 17, pp. 4756–4765, Dec. 2020, doi: 10.1049/iet-ipr.2020.0469.
[22] D. Lakshmi, K. P. Thanaraj, and M. Arunmozhi, "Convolutional neural network in the detection of lung carcinoma using transfer learning approach," *Int. J. Imag. Syst. Technol.*, vol. 30, no. 2, pp. 445–454, Jun. 2020, doi: 10.1002/ima.22394.
[23] S. Pang, F. Meng, X. Wang, J. Wang, T. Song, X. Wang, and X. Cheng, "VGG16-T: A novel deep convolutional neural network with boosting to identify pathological type of lung cancer in early stage by CT images," *Int. J. Comput. Intell. Syst.*, vol. 13, no. 1, pp. 771–780, Jun. 2020, doi: 10.2991/ijcis.d.200608.001.
[24] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, Mar. 2015, doi: 10.1109/TPAMI.2014.2345401.
[25] R. P. Rangkuti, V. Dewanto, and W. Jatmiko, "Utilizing Google images for training classifiers in CRF-based semantic segmentation," *J. Adv. Comput. Intell. Intell. Informat.*, vol. 20, no. 3, pp. 455–461, May 2016, doi: 10.20965/jaciii.2016.p0455.
[26] L. Wang, B. Jiang, Z. Tu, A. Hussain, and J. Tang, "Robust pixel-wise saliency detection via progressive graph rankings," *Neurocomputing*, vol. 329, pp. 433–446, Feb. 2019, doi: 10.1016/j.neucom.2018.10.061.
[27] X. Wang, H. Ma, and X. Chen, "Geodesic weighted Bayesian model for saliency optimization," *Pattern Recognit. Lett.*, vol. 75, pp. 1–8, May 2016, doi: 10.1016/j.patrec.2016.02.008.
[28] H. Lu, X. Li, L. Zhang, X. Ruan, and M.-H. Yang, "Dense and sparse reconstruction error based saliency descriptor," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1592–1603, Apr. 2016, doi: 10.1109/TIP.2016.2524198.

ZHIXIAN LI was born in Henan, China, in 1972. He received the B.S. degree in computer science and its application from Henan Normal University, Henan, in 1994, and the M.S. degree in computer application technology from the Xi'an University of Architecture and Technology, China, in 2005.

Since 2010, he has been an Associate Professor with the School of Computer and Information Engineering, Luoyang Institute of Science and Technology, Henan. He is the author of one book and more than 16 articles. His research interests include image processing, computer vision, big data analysis, and machine learning.

JIANWEI WU was born in Henan, China, in 1976. He received the B.S. degree in computer science and its application from the Henan University of Science and Technology, China, in 1998, and the M.S. degree in computer application technology from the Wuhan University of Technology, Hubei, China, in 2007.

Since 2011, he has been an Associate Professor with the School of Computer and Information Engineering, Luoyang Institute of Science and Technology, Henan. He is the author of two books and more than ten articles. His research interests include image processing, computer vision, big data analysis, and machine learning.

GUOQIANG WANG received the B.S. degree from the Zhengzhou University of Technology, Zhengzhou, China, in 1999, the M.S. degree from the Henan University of Science and Technology, Luoyang, China, in 2004, and the Ph.D. degree from the Dalian University of Technology, Dalian, China, in 2008. From June 2015 to June 2016, he was a Visiting Scholar in electrical and computer engineering with the University of Miami, Coral Gables, FL, USA. He is currently a Professor with the Luoyang Institute of Science and Technology, China. His main research interests include image processing, pattern recognition, and machine learning.