# Markov Chain and Adaboost Image Saliency Detection Algorithm Based on Conditional Random Field

Biao Lu Information Engineering Department, Suzhou University, Suzhou 234000, Anhui, China

Nannan Liang\* Information Engineering Department, Suzhou University, Suzhou 234000, Anhui, China \*Corresponding author

Chengfang Tan Information Engineering Department, Suzhou University, Suzhou 234000, Anhui, China

Zhenggao Pan Information Engineering Department, Suzhou University, Suzhou 234000, Anhui, China

Received: February 28, 2021. Revised: July 15, 2021. Accepted: July 27, 2021. Published: July 30, 2021.

Abstract—The traditional salient object detection algorithms are used to apply the underlying features and prior knowledge of the images. Based on conditional random field Markov chain and Adaboost image saliency detection technology, a saliency detection method is proposed to effectively reduce the error caused by the target approaching the edge, which mainly includes the use of absorption Markov chain model to generate the initial saliency map. In this model, the transition probability of each node is defined by the difference of color and texture between each super pixel, and the absorption time of the transition node is calculated as the significant value of each super pixel. A strong classifier optimization model based on Adaboost iterative algorithm is designed. The initial saliency map is processed by the classifier to obtain an optimized saliency map, which highlights the global contrast. In order to extract the saliency region of the final saliency map, a method using conditional random field is designed to segment and extract the saliency region. The results show that the saliency area detected by this method is prominent, the overall contour is clear and has high resolution. At the same time, this method has better performance in accuracy recall curve and histogram.

*Keywords*—Saliency Detection; Absorbing Markov Chain; Adaboost; CRF.

# I. INTRODUCTION

With the progress of the times and the rapid development of high technology, the channels and forms of human access to information have become complex and diverse. In the information age, the way to obtain information is not limited to traditional media such as text, voice, image, video, etc. augmented reality, virtual reality and other technologies have emerged. In the face of a lot of miscellaneous information, it has become the most practical problem to obtain the interested information more effectively in the shortest time. With the development of multimedia information technology and the maturity of mobile network architecture, the image has become the main carrier of information, and the effective extraction of key information in the image has become the focus of research in the field of computer vision.

The reason why human eyes can find the foreground target immediately in the complex and chaotic background is precisely by virtue of the powerful analysis and processing ability of their own visual system. In the face of complex scenes, the human visual system can quickly capture the conspicuous target, while "blind" not interested in information. This process seems natural and easy, but in fact, it involves many complex processes.

Visual attention[1] mechanism gives people the ability to selectively obtain information in the field of vision, so as to quickly screen out important targets. The method of its function is to extract all the scenes seen by both eyes first, and then select the region of interest according to the subjective experience of the observer or the difference between the object and the surrounding environment. Then, through the rotation of the eyeball, this region is placed in the central fovea of the retina of the eye, so that the vision and brain can pay attention to this region, so as to analyze it Li. Through this mechanism, human brain can filter the non important areas in visual field subjectively, and focus on the areas of interest at the same time.

Inspired by the mechanism of human visual attention, scholars who are committed to image processing hope to introduce it into the field of computer vision, design a series of algorithms to make the computer achieve the effect of simulating human eyes, screen out the attractive targets in the image, and ignore the complex and changeable background. This process is defined as significance detection, and the feature that can get human attention is the so-called significance. The significant region is the significant region.

Saliency detection allows the computer to analyze and identify important visual information from the complex image scene, so as to reduce the difficulty of the computer understanding the current image, and provide the required samples for the following various image processing processes. Because the saliency detection technology can help the computer identify the target area of the scene, it can save the waste of time and resources brought by manual calibration to a certain extent. At the same time, more importantly, this technology can be widely used in various image processing technologies, such as image segmentation[2], adaptive target recognition[3], image compression [4] and image retrieval[5].

Image segmentation: The so-called image segmentation refers to dividing the image into several non over-lapping regions according to the features of gray, color, texture and shape, and making these features present similarity in the same region, but obvious difference between different regions. At present, the mainstream segmentation ideas include: threshold based segmentation method, edge based segmentation method, region based segmentation method, graph theory based segmentation method, energy functional based segmentation method and so on.

Target recognition: Simulate the visual attention mechanism of human eyes, and detect the foreground area in the image. Thus, the irrelevant information can be filtered out quickly, and the main information in the scene, that is, the significant target, can be screened out. At present, machine learning methods are commonly used in the field of target recognition.

Adaptive image compression: compressed image can improve the efficiency of image transmission to a certain extent. Through saliency detection, we can choose to compress the unimportant redundant information, and compress the interested information a little or choose not to compress. After such processing, not only the image transmission rate is improved, but also the important information can be highlighted accurately.

Image retrieval: image retrieval is a hot research topic in computer vision. The main idea is to use the computer to simulate the process of human eyes perceiving things, extract the key content features from the whole or part of the image, use the algorithm of artificial intelligence and pattern recognition, transform the features into mathematical description, and compare with the images in the image database, and return the images similar to the query image in turn. In addition, saliency detection is often used in image scaling, medical image denoising and other fields. Image saliency detection has a great range of applications and practical research value, so it is very necessary to study this direction. The existing image saliency detection algorithms can be divided into two categories: top-down method and bottom-up method. The top-down method is usually task driven and has some prior knowledge. The bottom-up method is generally based on low-level visual features (color, texture, spatial distance, etc.) and data-driven. Next, several typical algorithms of the two methods are briefly analyzed.

# A. Itti model

Famous scholars Itti and Koch proposed a model w to simulate the visual characteristics of organisms, because it does not include complex high-level cortical signal processing[6].Therefore, Itti model was the most influential biomimetic significance detection model at that time because of its simple method and fast calculation speed. The algorithm includes feature extraction and saliency map generation.

Feature extraction: firstly, the input image is represented as a 9-layer Gaussian pyramid. The first layer is the input image, and then the input image is filtered by a 5 \* 5 Gaussian filter in the remaining eight layers, and then sampled, with the size of 1/2 to 1/256 of the input image respectively. Then, RGB, direction, brightness and other features are extracted from the nine layer pyramid to construct the pyramid with various features. In order to simulate the characteristics and structure of biological vision, Itti algorithm makes multi-scale subtraction of various features in the pyramid. The difference results in the comparison of local directional features between the center (scale C) and the periphery (scale s).

Saliency map generation: In order to minimize the error caused by feature amplitude difference and filter the noise to highlight the salient area, each feature map is normalized and then convoluted. Then, the initial feature is superimposed with the convolution result of the previous step to make the same feature compete in space in the way of side suppression. After several convolution processes, only a few of the most significant pixels are evenly distributed on the feature map. In the case of few significant points, multiple feature maps are superimposed to highlight the salient points of each feature. Then normalize the feature maps corresponding to the direction, color, brightness and other features, and sum them point by point (sampling to the fourth scale). In this way, saliency maps of various features can be obtained. Finally, saliency map s corresponding to the input image can be obtained by synthesizing the saliency maps corresponding to each feature.

In order to speed up the calculation, this model has many steps, and some parts of the model are not completely similar to the biological model. In addition, the result of operation depends heavily on the selection of parameters. In order to find the salient regions of different images, it is necessary to set different parameters. At the same time, due to the continuous use of neighborhood interpolation, the resolution of saliency map is reduced.

# B. Fourier spectrum residual model

Hou and Zhang proposed a saliency analysis algorithm based on spatial frequency domain analysis[7](SR algorithm). They proposed a simple calculation model of image visual saliency. This model and the model proposed by Itti are two different models. Itti model mainly focuses on the prominent part of the whole image for image visual saliency, and extracts saliency through the fusion of various features The focus of Hou's model is not on the prominent part of a picture, but on the background. Observe whether the background of most pictures meets any change in a certain space. Finally, remove the background, and naturally only the prominent part of the picture is left.

In this method, the input image is divided into two parts, one is prior information, the other is novel information. In this method, the input image is analyzed by logarithmic spectrum, the prior information in logarithmic frequency is removed, and the spectrum information of the target is obtained. Finally, the significance mapping is obtained by using inverse Fourier transform, that is, the target region corresponding to its spatial domain is obtained. The advantage of SR algorithm lies in its fast operation speed, and the saliency map obtained in the experiment is consistent with human eye perception.

#### C. Frequency tuning salient region detection model

Achanta and Hemami [8] proposed five indicators for significance detection. We should emphasize the most significant object. Highlight the whole significant are. Establish a clear boundary of significant object. We should discard the high-frequency information generated by texture and noise, because these information have no actual contribution to the significance detection, but will disturb the significance detection. Output full resolution saliency map effectively. The theory also plays a certain role in the development of significance detection in the future.

According to these five indexes, the basic idea proposed by the author is to use multiple band-pass filters to filter the continuous frequency band from low frequency to high frequency of the image, combine all the outputs as the final saliency map, and strive to produce full resolution saliency map. The specific method is as follows: for an image, Gaussian filtering is used to get the filtered image, and the lab value  $I_{wbc}(x, y)$  of its pixel points is taken, and then the average color of each pixel point of the input image is calculated in the lab space  $I_{\mu}$ ; the saliency value of each point in the image is defined by the following formula.

$$S(x, y) = \parallel I_{\mu} - I_{wbc}(x, y) \parallel$$

#### D. Significance detection model of global contrast

This model is an image saliency detection method based on histogram contrast proposed by Cheng[9]. In essence, this method is similar to LC model[10], but it takes color information into consideration instead of using only the gray level in the image as LC does. Based on RGB color space, the input image can have at most 256 \* 256 \* 256 colors, so it is not wise to use histogram to process it directly.

But this is only an extreme case. When dealing with images, we will not encounter so much color information. Therefore, this paper quantizes the three color channels in RGB from 256 to 12, that is, the processed input image has only 12\*12\*12 colors at most, so as to reduce the number of colors and facilitate subsequent processing. At this time, the use of histogram technology is not so complicated. Considering that excessive quantization amplitude may cause certain errors, this paper uses a smoothing process to optimize it. Finally, different from LC, it is processed in lab color space, but there are some differences between lab and RGB. The author chooses RGB space to complete color quantization. In order to add spatial relationship, the input image is divided into several regions by using graph based image segmentation method, and then the color histogram is established for each region. For each region in the image, its saliency is calculated by measuring its color contrast with other regions of the image.

#### E. Context aware saliency detection model

Gofermanwl [11] proposed that local information and global information should be considered at the same time. For example, attention should be focused on some specific areas rather than scattered, regions with different colors and patterns should have high significance values, uniform or fuzzy regions should have low significance values, and so on. Under the guidance of this criterion, the author designs the specific algorithm as follows: using super-pixel segmentation to process the image, comparing the ith super pixel PI with all other super pixels PJ in the lab color space. If the difference between PI and other super pixels is too large, it means that it is more significant than other super pixel regions.  $d_{color(Pi, Pj)}$  is the Euclidean distance between two super pixels.  $d_{\text{position}}(P_i, P_j)$  is the space distance between two super pixels. At the same time, considering the influence of color and distance, the difference between two super pixels is defined as:

$$\Box d(P_i, P_j) = \frac{d_{\text{color}}(P_i, P_j)}{1 + c^* d_{\text{position}}(P_i, P_j)}$$

Where C is the control constant, take 3. It can be seen from the formula that the larger the color distance between two super pixels and the smaller the position distance, the greater the difference between them. If for any super pixel PJ, the difference between PI and it is very large, then Pi is considered to be significant.

In order to improve the computational efficiency, we only need to calculate the first k (k = 65) blocks most similar to PI. Finally, the formula of significance value is obtained.  $\Box$ 

$$S_{i}^{r} = 1 - \exp\left\{-\frac{1}{k}\sum_{k=1}^{k}d(p_{i}^{r}, q_{k}^{r})\right\}$$

In order to ensure the comprehensiveness and robustness of saliency detection. The input image in multiple scales according to the previous formula to calculate its significance value, as long as the calculation of the significance value of multiple scales, and then take the average of the results. The formula is as follows.

$$\bar{S}_i = \frac{1}{M} \sum_{r \in R} S_i^r$$

Finally, context correction is added: the saliency threshold is set to extract the most noticeable local region from the saliency map, and the saliency value of the pixel outside the region of interest is weighted by the Euclidean distance between the nearest saliency pixel to get a new saliency value. Thus, the saliency value near the salient target is improved and the saliency value of the background area is reduced. The calculation formula is as follows:

$$S_i = \overline{S}_i (1 - d_{foci(i)})$$

Where  $d_{foci(i)}$  is the Euclidean distance between pixel I and the nearest significant pixel.

Compared with the bottom-up model, the traditional top-down model mainly applies the influence of advanced factors such as task and prior knowledge. Zhang proposed a term down saliency detection model[12]. Firstly, Bayesian framework was used to extract statistical information from a series of natural images, and difference of Gaussian (dog) and independent component analysis (ICA) were used respectively Two image saliency algorithms based on eye movement are proposed. As a top-down model, Jiang presented another more intuitive algorithm in his article, considering the image saliency from the perspective of regression.

The specific algorithm is as follows: the first step is to get positive and negative samples from many artificially marked images; the second step is to select a total of three kinds of information such as regional contrast as features to describe the samples; the last step is to use positive and negative samples to train the model, iterate the training process, and get the final model after reaching the standard, which is used to detect the significance of the input image. In the mainstream image database, this method has achieved good accuracy and recall, but because the first step of the method needs to get the sample set from many manually marked images for the next training, if there is a big difference between the selected image and the input image in the final classification process, it will cause some errors. This is also a common problem for this kind of method.

In a word, the bottom-up saliency detection method can get more attention and research because of its fast speed and using visual features. At the same time, as the basis of many research directions in computer vision, image saliency detection plays an important role in the development of image compression, image segmentation, target recognition and other related fields.

# II. DESIGN OF MARKOV CHAIN AND ADABOOST IMAGE SALIENCY DETECTION BASED ON CONDITIONAL RANDOM FIELD

The traditional salient object detection algorithm is used to apply the underlying image features and some prior knowledge, or use a large number of manually labeled sample images to train the machine learning model, The training of machine learning model also costs a lot of time and human resources. In order to correctly highlight the saliency target close to the image edge and reduce the time consumption, this paper designs Markov chain[13, 14] and Adaboost image saliency detection technology based on conditional random field. The comparison of saliency detection methods is shown in Fig.1.



Fig.1. Comparison of significance detection methods

In Fig.1, the order from left to right is: original image, saliency map generated by traditional method and saliency map generated by this method. It can be seen that the traditional method mistakenly gives higher saliency value to the background pixel in the center of the image according to prior knowledge, while the foreground target is not highlighted correctly, In order to highlight the salient objects close to the edge of the image correctly, a Markov chain and Adaboost image saliency detection method based on conditional random field is proposed. The overall framework design of the algorithm is shown in Fig.2.



CRF Salient region extraction

Fig.2. Framework of Markov chain and Adaboost image saliency detection based on CRF

The initial saliency detection model based on Markov chain can detect the saliency objects close to the edge more accurately, which is 1-4 steps

(1) The original image is segmented by SLIC;

(2) The edge super pixels are the virtual absorption nodes, and the other super pixels are the transfer nodes;

(3) The absorption time of each transfer node is calculated and assigned to the corresponding super-pixel as the saliency value to get the saliency map.

(4) The initial saliency map is obtained after background filtering and smoothing.

The optimization model based on Adaboost can optimize the initial saliency map of Markov chain to enhance the contrast between foreground and background in saliency map, namely 5-9 steps. The steps are as follows:

(5) SVM with multiple single kernel and single feature is selected as the weak classifier;

(6) The training samples are extracted from the initial saliency map by setting high and low thresholds;

(7) Based on Adaboost iterative algorithm, training samples are used for training, and a strong classifier optimization model is obtained by integrating multiple weak classifiers;

(8) The initial saliency map is optimized by constructing a strong classifier, and the optimized saliency map with better contrast is obtained;

(9) By fusing the initial saliency map and the optimized saliency map, the final saliency map is obtained.

The initial saliency map of Markov chain is good at capturing local details, and the optimized saliency map of Adaboost can highlight global information. Combining the two, the final saliency map is obtained. This processing makes the overall algorithm maintain good universality and robustness for all kinds of images.

Machine learning algorithm is introduced into the field of image processing, and a salient region extraction method based on conditional random field (CRF) is proposed. The CRF model can accurately extract the salient region and ensure the consistency of the inner region of the contour.

### A. Initial saliency detection model based on Markov chain

For an input image, SLIC method is used to segment the super-pixel. The image after super-pixel segmentation is regarded as an Absorbing Markov chain model. Each super-pixel in the graph is regarded as a node in the graph model, and the line between adjacent nodes is regarded as a boundary. Next, copy the edge super-pixel and set it as the virtual edge absorption node. The saliency value of the super pixel is defined as the expected time that the corresponding node starts from itself and travels on all the transfer nodes on the path before reaching all the absorption node. If the edge node is directly set as the absorption node without copying, its own significance value cannot be calculated. Copy the edge super pixel as the virtual absorption node, as shown in Fig.3.



Fig.3. Copy edge super pixel as virtual absorption node

In the established graph, each transfer node is only connected with its adjacent nodes, and there is no association between the absorption nodes. The feature difference between two adjacent nodes determines the boundary weight between them. Specifically, the expression of boundary weight  $W_{ij}$  between adjacent node i and node j is as follows:

$$w_{ij} = e^{-\frac{\sum d_k(c_i, c_j)}{\sigma^2}}$$

Where  $\sigma$  is the constant controlling the weight. It represents the difference between the feature mean value c of node i and the feature mean value cj of node j in the k-th feature space. For the difference between the n-dimensional feature vectors cin and cjn, we use Euclidean distance to measure it. The specific expression is shown in:

$$d_k(c_i, c_j) = \sqrt{\sum_{n=1}^{N} (c_{in}, c_{jn})^2}$$

It is not difficult to see from the expression that the boundary weight reflects the similarity between the nodes to a certain extent. The smaller the difference between the two characteristics, the greater the weight between the nodes, which means the closer the connection. On the contrary, the larger the feature difference between the super pixels, the smaller the weight, indicating that the connection between the corresponding nodes is not strong.

The boundary weights of all nodes are calculated by the above formula, and then the corresponding matrix of the established graph is rearranged, so that t transition nodes are in the front and r absorption nodes are in the back. The elements of the rearranged incidence matrix M represent the relevance between the corresponding nodes. The details are as follows:

$$M_{ij} = W_{ij} \quad j \in N(i), 1 \le i \le t$$
$$M_{ij} = 1 \quad i = j \text{ or } \quad M_{ij} = 0 \quad other$$

Where N(i) represents the set of other nodes adjacent to node i. It can be seen that transfer node i only has weight with adjacent nodes, and does not consider the relationship with other nodes. At the same time, the sum of the weights connected to each node is calculated, that is, the sum of each row of matrix M is calculated, and then the incidence matrix is normalized as the transition probability in the absorption Markov chain. The sum of boundary weights of each node is represented by diagonal matrix D.

$$D = diag(\sum_{j} m_{ij})$$

The transfer matrix t of the corresponding graph is obtained by multiplying the diagonal matrix and the incidence matrix. It can be seen that its elements are the normalized boundary weights.

$$T = D^{-1} * M$$

T is also a sparse matrix. The matrix Q is extracted by formula, and then the basic matrix n is obtained by formula  $N = (I - Q)^{-1}$ . The elements in matrix N are called the time consumed by the random walkers from the transition node i to the state j. Therefore, each row of the basic matrix represents the expected time for the random walker to start from the transition node i, go through all the transition states, and finally reach the absorption state. Then, the absorption time of each transfer node can be calculated

$$y = N * e$$

Where e is the unit column vector of the T dimension.

Finally, the absorption time y in the above formula is normalized to obtain the saliency value of the corresponding super pixel of each node.  $\Box$ 

$$S(i) = \gamma(i), i = 1, 2, 3, \dots, t$$

Where I is the transition node in the graph and Y is the absorption time after normalization.

Starting from a transition node, the random walker walks along the boundary between adjacent nodes and obeys the transition probability. After swimming for a period of time, he will be absorbed by any node in the virtual boundary absorption node and stay. Of course, the probability and time that he is absorbed to each virtual absorption node are not the same, and the absorbed time is the expected time of all transfer nodes on the path. Taking the absorbed time as the saliency measure can suppress the background area and highlight the whole target. The process of saliency detection algorithm based on Markov chain is as follows:

Saliency detection algorithm based on Markov chain
Input: image to be detected
Output: initial saliency image
step1: The input image is segmented by super pixels, and the feature vector of
each super pixel is calculated
step 2: The edge super pixels are the virtual absorption nodes, and the other
super pixels are the transfer nodes
step 3: Calculation of incidence matrix M and transfer matrix T
step 4: Extract matrix Q from transfer matrix T and calculate basic matrix
step 5: The absorption time of each transfer node is calculated and assigned to
the corresponding super-pixel as the saliency value to obtain the initial
saliency graph S
step 6:Establish a histogram for S and calculate Y. if Y < 2, smooth S and then
output

step 7. Otherwise, S is filtered and smoothed before output

Because the proposed saliency detection method based on Markov chain uses super-pixel features to generate saliency map, it tends to detect image details and capture local structure information more easily. However, the ability to grasp the overall picture is slightly insufficient, and the image effect is not good for the uneven color distribution of the foreground target. Based on this model, an optimization model based on Adaboost iterative algorithm is proposed. The model focuses on the global information and can generate the optimized saliency map. The contrast of background filtering effect is shown in Fig.4.



Fig.4. Comparison of background filtering effect

# iterative algorithm

As a simple and practical classifier, support vector machine (SVM) is often used in various image processing problems, and has achieved good results. However, in the process of applying SVM classifier, one of the most noteworthy problems is that it is difficult to select a consistent and applicable kernel function for different samples to be classified, and the selection of kernel function becomes more difficult for complex and changeable images encountered in image processing.

The ensemble learning method is used to combine multiple classifiers into strong classifiers, so as to obtain stronger generalization performance than a single classifier. The properties of weak classifiers are various, and the accuracy of classification is different for the same problem. Moreover, there is a contradiction between the accuracy and diversity of individual classifiers.

The saliency graph optimization method based on Adaboost uses the idea of Adaboost to get a strong classifier through iterative training. The support vector machines with multiple single kernel and single feature are not simply superimposed, but are assigned different weights. In this way, a strong ensemble is constructed by multiple classifiers with general generalization performance, which makes the combination generate a strong classifier with better classification effect.

At the same time, traditional methods first select a large number of images from the database, and then extract features and positive and negative samples for training. This requires manual calibration of positive and negative samples, consumes a lot of human resources, and requires a lot of calculation. Moreover, due to the differences between training samples and test samples, the traditional classification methods can not always get the expected results for images from different databases or different categories of images.

In order to improve the efficiency and quality of classification, training and testing are limited to one picture, and effective training samples are extracted from the saliency map obtained by Markov chain model, which reduces the time complexity. At the same time, because of the previous background filtering and image smoothing, the saliency map has high accuracy, and the samples provided are reliable and robust, which effectively improves the classification accuracy. The saliency graph optimization based on Adaboost is shown in Fig.5.



Fig.5. saliency graph optimization method

For any input image, using Markov chain model, we select two appropriate thresholds, and mark the significant value lower than the low threshold as negative samples, and the significant value higher than the high threshold as positive samples. Super pixels between high and low thresholds are not selected as training samples. Then the sample set is extracted from it, where Si represents the ith sample and N represents the total number of training samples. The linear sum of multiple SVM kernels can be expressed as (1).

$$\Box \Box k(s, s_i) = \sum_{n=1}^{M} \beta_n k_n(s, s_i), \sum_{n=1}^{M} \beta_n = 1 \quad (1)$$

Where  $\beta_n$  is the weight of the corresponding kernel function, M is the number of weak classifiers, M = the number of features \* the number of kernel functions. For different feature sets, the objective function can be written as follows:

$$Y(s) = \sum_{i=1}^{M} \beta_{n} \sum_{i=1}^{N} \alpha_{i} l_{i} k_{n}(s, s_{i}) + \overline{b}$$
(2)

Where  $\alpha_i$  is the Lagrange multiplier and *b* is the bias constant of SVM. Adaboost enhancement algorithm is used to replace the single combination of traditional SVM. Formula (3) can be written as follows:

$$Y(s) = \sum_{n=1}^{M} \beta_n(\alpha^T k_n(s) + \overline{b_n})$$
(3)

In order to calculate the coefficient  $\beta_n$ , Adaboost method

is used, and H is the number of iterations in the training process. Each single kernel SVM is regarded as a weak classifier, and the final strong classifier is the weighted combination of all single kernel SVM. The specific process is as follows:

$$\varepsilon_m = \frac{\sum_{i=1}^{N} D(i) |Z_n(S_i)| \operatorname{sgn}(-l_i z_n(S_i)) + 1) / 2}{\sum_{i=1}^{N} D(i) |Z_n(S_i)|}$$
(4)

When the classification is correct, the expression value in the symbolic function in equation (4) is negative, then the value of the whole symbolic function is -1, and the corresponding error rate is reduced. When the corresponding classifier is wrong, the situation is opposite, which makes the error rate increase. Find the minimum error rate and the corresponding SVM objective function, and then calculate the weight of the *j*-th kernel function according to the following formula.

$$\beta_{j} = \frac{1}{2} \log \frac{1 - \varepsilon_{j}}{\varepsilon_{j}} * \frac{1}{2} (\operatorname{sgn}(\log \frac{1 - \varepsilon_{j}}{\varepsilon_{j}}) + 1 \qquad (5)$$

It can be seen from the expression that only when the error rate is less than 0.5 can the combination coefficient  $\beta > 0$  be guaranteed, and then the sample weight is updated according to the error rate.

$$D_{j+1}(i) = \frac{D_j(i)e^{-\beta_j l_j z_j(s_j)}}{2\sqrt{\varepsilon_j(\varepsilon_j - 1)}}$$
(6)

When the j-th iteration, if the sample is correctly classified and the product is positive, the (6) molecule will decrease, that is, the weight of the sample will decrease in the j + 1 iteration. If it is wrongly classified, the corresponding weight is increased, which makes itself more concerned by the classifier in the next iteration. After J training iterations, J coefficients and objective functions can be obtained. According to equation , a strong classifier based on Adaboost principle can be constructed. Next, we can use the segmented super pixels in the current input image as the input test samples to get an optimized saliency map. The following figure summarizes the specific algorithm flow of Adaboost optimization model construction algorithm.

Optimization model construction algorithm flow
based on Adaboost
Input :
Training sample set as $\{S_i, I_i\}_{i=1}^N$ ; Training times is J
Output : $Y(s) = \sum_{h=1}^{H} \beta_h z_h(s)$
step1: D(i)=1/N,i=1,2,3N
step2: for j=1,2,3J do
step 3: $\varepsilon_m = \frac{\sum_{i=1}^N D(i)  Z_n(S_i)  \operatorname{sgn}(-l_i z_n(S_i)) + 1) / 2}{\sum_{i=1}^N D(i)  Z_n(S_i) }$
step 4: if $\mathcal{E}_m > 0.5$ then break
step 5: $\beta_j = \frac{1}{2} \log \frac{1 - \varepsilon_j}{\varepsilon_j} * \frac{1}{2} (\operatorname{sgn}(\log \frac{1 - \varepsilon_j}{\varepsilon_j}) + 1)$
step 6: $D_{j+1}(i) = \frac{D_j(i)e^{-\beta_j l_j z_j(s_j)}}{2\sqrt{\varepsilon_j(\varepsilon_j - 1)}}$
step 7: End For

The core idea of Adaboost algorithm is to increase the weight of samples with wrong classification and reduce the weight of samples with correct classification. In this way, the samples with wrong classification by weak classifier are more obvious. In the next training process, the classifier will pay attention to the wrong samples, so as to constantly update the new sample distribution. The next time the classifier classifies the wrong points, it will improve the overall error rate, resulting in the weight of the weak classifier that determines the error smaller, and increase the weight of the good classifier. Finally, a reasonable weight distribution of the weak classifier is obtained through the iterative process, and a strong classifier is generated, whose saliency map can highlight the global information of the input image. The algorithm rendering is shown in Fig.6.



Fig.6. Effect of Adaboost algorithm

The saliency map obtained by image saliency detection is a gray image which indicates the saliency of all pixels in the input image. However, the optimized saliency map can only see the general distribution of each saliency region in the input image, and can not accurately segment the saliency target from the image, which is not conducive to the subsequent development of other image processing work. Aiming at the problem of target extraction in saliency map, a saliency region extraction method based on conditional random field algorithm is proposed.

#### C. Segmentation of significant regions based on CRF

Conditional random fields (CRF) is a machine learning algorithm developed rapidly in recent years. As a method based on statistical analysis, it is widely used in semantic segmentation, data mining, image processing and other computer fields. Conditional random field shows excellent performance and application potential for all kinds of sample classification.

In order to reflect the differences between pixels, it is necessary to select suitable features and design the expressions of local potential function and interactive potential function to reflect the relationship between individual pixel and marker value. The parameters are trained according to the features, and the model is established based on the maximized conditional probability distribution. In this way, the label values of each pixel can be obtained by inputting the observation sequence of the sample, and then the saliency target can be correctly segmented from the saliency map.



Similar to the way that Markov random field defines joint probability, CRF uses potential function and clique in graph structure to define conditional probability P(Y|X). On the

premise of obtaining the observation sequence, CRF model mainly involves two kinds of marker variables, namely single marker variable and adjacent marker variable. The expression of conditional a posteriori probability distribution is as follows:

$$P(Y \mid X) = \frac{1}{Z} \exp \left\{ \sum_{i \in S} A(y_i, X) + \sum_{i \in S} \sum_{j \in N_j} I(y_i, y_j, X) \right\}$$
(7)

 $A(y_i, x)$  represents the local potential function, which is used to represent the relationship between the i-th pixel feature and the tag value, and I(Y, X) represents the interactive potential function, which is used to represent the feature difference between the pixel I and its adjacent pixel J, as well as the relationship between the tag value. Z is the normalized constant to make the conditional a posteriori result between 0 and 1.

According to the principle and mechanism of CRF, the interaction function is designed mainly around two aspects. Firstly, the interaction function should be able to show the interaction between two adjacent pixels. Secondly, it should reflect the relationship between the extracted features and the sequence of pixel markers. Here, the logistic function is used to construct the interaction function, and the expression is as follows:

$$I_{ij}(y_i, y_j, x_i, x_j) = y_i y_j \upsilon^T \mu_{ij}(X) = \log \frac{1}{e^{-(y_i y_j \upsilon^T \mu_{ij}(X))}}$$
(8)

In the above formula,  $\mathcal{U} = (\mathcal{U}_0, \mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_n)$  is the balance parameter, and  $\mu_{ij}(X)$  is the feature difference between pixel i and j, the expression is.

$$\mu_{ij}(X) = \Box f_i(x) - f_j(x) \Box$$
(9)

After establishing the model, we need to use samples for training. Here, the CRF model is trained by using several manually calibrated images in the image database as samples. A common method to estimate the class conditional probability is to assume that it has a certain form of probability distribution, and then estimate the parameters of the probability distribution based on the samples. In the training process, the maximum likelihood estimation is used to obtain the optimal estimation of each parameter. The expression of likelihood function is as follows.

$$L(\theta) = \prod_{m=1}^{M} \prod_{j \in S} P(y_i^m / x_i^m, \theta)$$
(10)

Where  $\theta = \{\omega, \upsilon\}$  and m are the number of samples in the training process. The maximum likelihood function  $L(\theta)$ can estimate the optimal solution of parameter set  $\theta$ . The expression is as follows.

$$\stackrel{\scriptscriptstyle \Delta}{\theta} = \arg\max\prod_{m=1}^{M}\prod_{j\in S}P(y_i^m x_i^m \,/\, \theta) \quad (11)$$

By solving the partial derivatives of the parameters, the optimal solutions of the model parameters  $\omega$  and  $\upsilon$  are obtained. At this time, the region extraction model based on conditional random field is completed. After the CRF model is constructed, the test samples can be used for reasoning operation. The model outputs the labels of all samples to

generate a binary image, and then extracts the pixels covered by the white area in the binary image corresponding to the input image, which is the foreground target.

This paper briefly reviews the salient region extraction method based on CRF: the salient value of each pixel in the saliency map is taken as the observation sequence, and the marker sequence value is +1 or -1, which respectively represents that the pixel is in the salient region or the background region, The model parameters are obtained by maximum likelihood estimation. After the model is constructed, the input image is processed by our proposed significance detection method based on Markov chain. The saliency map is taken as the input, and the saliency value on the saliency map is taken as the observation sequence. Through the CRF model, the labeled sequence is obtained. The positive pixels in the labeled sequence are white, and vice versa, A marked binary image can be obtained. If the white area in the binary image corresponds to each pixel in the original image, the target area can be extracted. The CRF method is compared with the traditional Otsu method and adaptive threshold segmentation method. At the same time, the truth value image is added to facilitate the comparison and observation. The salient area extraction method based on CRF is shown in Fig.8(a.b.c.d).





Fig.8. Salient region extraction method based on CRF

From left to right, the segmentation results are as follows: original image, Otsu, adaptive threshold segmentation, CRF method, and manual labeling image. It is not difficult to find that compared with the other two classical methods, the CRF method proposed in this paper can extract salient objects more completely. From the segmentation results of Otsu model, it can be seen that in some cases, this method can only obtain the profile of salient objects, In the experiment, because the gray distribution of each image is not the same, some background noise will also lead to the low accuracy of the threshold, Adaptive threshold segmentation can extract the outline of the salient region, but referring to the artificially marked segmentation image, it sometimes generates a little more background information in the image. On the contrary, our method can not only correctly extract the foreground region, but also better suppress the background noise around the target.

# III. EXPERIMENT RESULTS AND ANALYSIS

This chapter will verify the proposed saliency detection algorithm from qualitative and quantitative aspects, evaluate the saliency algorithm on four public image databases ASD,

MSRA, SED and SOD, and compare with several existing mainstream algorithms. We use PR curve and AUC, F-measure histogram to compare the algorithms quantitatively.

In the previous paper, we have briefly introduced and analyzed the early classical algorithms and the better methods in recent years. Now we compare the proposed method with several representative methods mentioned above for quantitative and qualitative evaluation. Compared with other methods in the experimental group, our proposed method comprehensively considers the color and texture features, and highlights the local contrast and global information, Multi scale processing is used to reduce the error of significance detection caused by the uneven size of significant area. Figure 9 shows the comparison of several significance detection methods, and the order from left to right is: original image, text model, DSR model, GC model, GMR model, HS model, RC model, SF model, In this paper, the proposed method, DSR, GMR, Hsu, RC, SF. The performance of the algorithm in this paper is comparable to that of the excellent algorithm in recent years, and some images even have better effect. This fully shows that the idea of using Adaboost to optimize the saliency map of image is correct and reasonable under the premise of Absorbing Markov chain model.



Fig.9. Comparison of results of several significance detection methods

From the saliency map, compared with other methods, this algorithm is more accurate for foreground target extraction and ideal for background noise suppression. This is because we use graph cut and background filtering method to make saliency map more accurate and uniform to highlight foreground targets. At the same time, it is also more effective for targets with a variety of textures and colors, Other algorithms also have good results, but there are still some limitations. For example, RC model, this method uses super-pixel segmentation to input image, then accelerates histogram, takes color histogram and spatial position difference as super-pixel saliency value, and then obtains saliency map

In the third image, RC model mistakenly gives a high saliency value to the zebra crossing on the road. GMR also uses the super-pixel technology, Then, the feature difference between the super pixel and other super pixels in its neighborhood is taken as the measurement basis of its significant value. The disadvantage of this idea is that if the super pixel segmentation distributes several pixels with little color difference in the same foreground object in different super pixels in the field, the feature difference between this super pixel and other super pixels will be smaller, However, the super-pixel contained in the foreground object should have a high saliency value

Secondly, if the color difference between the salient region and the background is not obvious, it will also affect the saliency image generated by the algorithm, In the saliency map corresponding to SF model, it can be seen that this method can outline the outline of the foreground target, but the interior of the target area is not uniform. In the map, the saliency values of the pixels contained in the petals and the lower body of pedestrians are wrongly suppressed, which is easy to cause incomplete foreground target in the subsequent threshold segmentation Due to the use of clustering algorithm in the model, the effect of saliency image generated for a certain type of image is not ideal, and there is a certain degree of error.

In the DSR model, the background dictionary is composed of four sets of image blocks in the boundary region of the image. Although the boundary super-pixel is used as the virtual absorption node in this method, based on the nature of Markov chain, it can still better highlight the salient objects close to the edge. However, when the DSR method is used to process the

penultimate image, the red markers are not highlighted correctly, We will compare the accuracy recall curve, F value and AUC value of each algorithm quantitatively. The comparison of the results of several saliency detection methods on the public database is shown in Fig.10(a.b.c.d).









(d)

Fig.10. Comparison of several saliency detection methods on public databases□

Table 3.1 and table 3.2 list the specific evaluation data of AUC value and F-measure of each algorithm in four image databases. The last behavior is the comparison of each method on sed database after using this model optimization

Table 5.1 Comparison of AUC evaluation value	Table 3.1	Comparison	of AUC	evaluation	values
--	-----------	------------	--------	------------	--------

	ours	DSR	GC	GMR	HS	RC	SF
MSRA	0.8532	0.8196	0.7632	0.8452	0.8157	0.8287	0.6735
ASD	0.9717	0.9612	0.9552	0.9702	0.9711	0.9013	0.9128
SED	0.9385	0.9215	0.9113	0.9013	0.9146	0.9213	0.9136
SOD	0.8215	0.8132	0.8203	0.7918	0.8014	0.8167	0.8189
RE	0.9412	0.9213	0.8921	0.7981	0.8412	0.8835	0.9124

Table 3.2 Comparison of F-measure evaluation values

	ours	DSR	GC	GMR	HS	RC	SF
MSRA	0.691	0.672	0.631	0.655	0.641	0.631	0.668
ASD	0.913	0.873	0.867	0.875	0.893	0.885	0.837
SED	0.694	0.667	0.597	0.672	0.653	0.664	0.668
SOD	0.663	0.652	0.642	0.661	0.562	0.598	0.468
RE	0.694	0.623	0.628	0.685	0.653	0.673	0.684

It can be seen from the comparison results that the performance of this algorithm is even better than several excellent saliency detection models published in recent years. Among them, the method proposed in this paper has the best effect on MsrA database. Although RC algorithm has a high recall rate, it considers the global information, Color features and spatial distance are used to define the pixel saliency value. However, the saliency image is not ideal when the foreground background color is similar or the target is close to the image edge.

The method in this paper uses the principle of Absorbing Markov chain, which can suppress the background noise better and has stronger robustness. The algorithm DSR and the algorithm in this paper produce very similar histogram results, and the two are on the same level. But from the comparison results of accuracy recall curve, it can be seen that the curve effect of the algorithm in this paper is better than that of DSR in most cases, In ASD, the algorithm performs well in both AUC and F-measure, and the AUC reaches 0.9717.

By comparing the detection results, it can be found that the saliency map produced by this algorithm shows strong background suppression ability, unlike other algorithms, such as HS model, when the texture of the target is complex, the effect is general. This is because this paper constructs a strong classification model based on Adaboost, which makes reasonable use of the global information and enhances the contrast between the target and the background, Thus, the saliency map not only suppresses the background, but also highlights the target area

Fig.10 shows the accuracy recall curve and AUC, F-measure histogram of this method and other significance detection methods in SOD database. As mentioned in the previous database introduction section, although SOD only contains 300 pictures, it is difficult compared with other databases. Quantitatively, when comparing the PR curves of each algorithm on the database s0d, the, The performance of the algorithm in this paper is slightly better than that of other models. This shows that the saliency described by the absorption Markov chain and Adaboost is more accurate. Of course, it must be pointed out that from the experimental results of each group, we can see that although there are a variety of images in the dataset, including multiple targets, different sizes of targets, close to the edge and complex texture of salient targets, But the good performance of this algorithm shows the robustness of the algorithm

Through the data shown in Table 3.1 and table 3.2, it is not difficult to find that the algorithm in this paper performs better than other methods. At the same time, the data in the "sed (RE)" row shows that we can use our optimization algorithm to improve other models, that is, in the SED database, the saliency map generated by other methods is smoothed and optimized by the strong classifier proposed in this paper, According to the comparison of AUC evaluation values in table 4.1, the AUC value of the two methods after our optimization is more than 0.9, in which the AUC value of SF after optimization is 6% higher than that of the original, and the highest AUC value of 0.9218 is obtained by DSR, The optimized algorithms are also greatly improved. It can be seen that the optimization effect of the proposed method for saliency map is obvious.

# IV. CONCLUSION

The Markov principle is reviewed, including the absorption time and absorption probability in the absorption Markov chain, and then the algorithm details are described: including the establishment of graph model, saliency detection based on the absorption Markov chain. The virtual boundary node is set as the absorption node, and the absorption time from each node to the absorption node in the image is calculated as the saliency value of this node, The high and low thresholds are determined, and positive and negative samples are selected from the initial saliency map obtained by the above method for the classifier training process, which reduces the time consumption caused by manual calibration and a large number of training samples. However, the number of samples provided by a saliency map is limited, so the range of selected samples is expanded to multiple scale saliency maps for the same input image, It not only increases the number of positive and negative samples, but also makes the method more robust in the case of uneven target size. Based on the principle of Adaboost, each single kernel and single feature SVM is regarded as a weak classifier. After several training, the weight of each weak classifier is obtained, which is combined to form the optimal initial saliency map of a strong classifier, Combined with two saliency maps as the final saliency map, so as to highlight their respective advantages. Finally, the saliency region extraction method based on conditional random field is used to accurately extract the saliency target in the picture.

#### ACKNOWLEDGMENT

This work was supported in part by (Natural science research project of Anhui Colleges (NO.2008085QD192) (Excellent young talents support program in Colleges and Universities (gxyqZD2020048) (Horizontal project of Suzhou University (No.2020xhx087) (Quality engineering project of Suzhou University(No.szxy2020xgjy05 (Excellent academic and technical backbone University (No.2020XJGG01) (Key scientific research project of Suzhou University in 2020 (No. 2020yzd01); The Key Research and Technology Development Projects of Anhui Province (No. 202004a06020045); The Key Natural Science Project of Anhui Provincial Education Department (No. KJ2019A0668); (Key scientific research project of Suzhou University in 2020 (No. 2017yzd19); Leader of computer science and technology in Suzhou University (2019XJZY23); Collaborative education project in 2020 (202002154001).

#### REFERENCES

 Boothe R. "Perception of the Visual Environment". Springer Berlin, vol. 65, no.3, pp. 29-43, 2015.

- [2] Han J,Ngan K N,Li M,et al., "Unsupervised extraction of Visual attention objects color images", *IEEE Transactions on Circuits&Systems for Video Technology*, vol. 16, no.1, pp.131-139, 2016.
- [3] Rutishauser U, Koch C., "Is bottom-up attention useful for Object recognition", Proceedings of the 2016 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.136-142, 2016.
- [4] Christopoulos C,Ebrahimi T., "The JPEG2000 still image coding system:An overview", *IEEE Transactions on Consumer Electronics*, vol. 46, no.4, 103-127, pp. 2008.
- [5] Chen T,Cheng M., "Sketch2 Photo:internet image montage", ACM Transactions on Graphics (TOG), vol.16, no. 5, pp.181-192, 2019.
- [6] Itti L, Koch Cs Niebur E., "A model of saliency based Visual attention for rapid scene analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.20, no.11, pp.18-29, 2018.
- [7] Hou X., "Saliency Detection: A spectral residual approach", IEEE Conference on Computer Vision and Pattern Recognition, pp.121-129, 2017.
- [8] Achanta R, Hemami S, Estrada F, et al., "Frequency-tuned salient Region detection", *IEEE Conference on Computer Vision and Pattern Recognition*, pp.177-184, 2019.
- [9] Cheng M,Mitra N J., "Global contrast based tregion detection", Proceedings of the 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 409-416, 2011.
- [10] Zhai Y, Shah M., "Visual attention detection in video sequences using spatiotemporal cues", ACM Internationa Conference on Multimedia, pp.115-124, 2016.
- [11] Goferman S, Zelnik-Manor L9 Tal A., "Context-Aware Saliency Detection", *IEEE Transactions on Software Engineering*, vol. 34, no. 10, pp.135-146, 2017.
- [12] Zhang L,Tong M H,Marks T K, et al. "SUN: A Bayesian framework for saliency using natural statistics", *Journal of Vision*, vol. 8, no. 7, pp.132-141, 2016.
- [13] Benmachiche A., Makhlouf A., "Optimization of Hidden Markov Model with Gaussian Mixture Densities for Arabic Speech Recognition", WSEAS Transactions on Signal Processing, vol.15, pp. 85-95, 2019.
- [14] Lyandres, V., "On the Correlation Function of an Arbitrary Distributed Continuous Markov Process", WSEAS Transactions on Circuits and Systems, vol. 19, pp. 105-110, 2020.

# Creative Commons Attribution License 4.0 (Attribution 4.0 International , CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0

https://creativecommons.org/licenses/by/4.0/deed.en US