Journal of Computers Vol. 32 No. 3, 2021, pp. 135-149
doi:10.3966/199115992021063203010

UAV Aerial Image Target Detection Based on Visual Attention Combined with Dual-weight Adaboost

Pan-Cheng Lu, Yong Ding[*], Chang-Jian Wang

School of Automation, Nanjing University of Aeronautics and Astronautics, No. 29, Jiangjun Avenue,
Jiangning District, Nanjing, 210016, China
{lupancheng, dingyong, cjwang}@nuaa.edu.cn

Received 30 September 2020; Revised 17 December 2020; Accepted 21 January 2021

Abstract. For the problem that UAV (Unmanned Aerial Vehicle) aerial image targets are small and easily interfered by background and illumination changes. This paper proposes a UAV aerial image target detection algorithm combining visual attention with dual-weight adaboost (Dw-adaboost). First of all, a visual attention combined with Dw-adaboost image saliency detection framework is proposed, we selected the brightness, direction, regional contrast, and spatial location features as the main feature channels to generate the saliency map; Secondly, a Dw-adaboost classification algorithm is proposed to determine the optimal weight of the main feature channel; Finally, we use high-efficiency sub-window search on the saliency map to achieve target detection in aerial images. Experiments show that the method in this paper improves the problem that UAV aerial image targets are small and easily interfered by background and illumination changes. It can achieve more accurate detection of UAV aerial image targets in complex scenes.

Keywords: visual attention, saliency map, adaboost, UAV aerial image, target detection

1 Introduction

With the rapid development of UAV technology, UAVs have increasingly rich functions and replaced manned aircraft in many fields. In practical applications, UAV aerial image target detection often has to face more complex scenes, including background dynamic interference, illumination changes, occlusion, and scale changes. Existing target detection algorithms have low detection accuracy and insufficient adaptability in the above complex scenes, which brings great difficulties and challenges to the actual application process of UAVs. UAV target detection algorithms include template matching, key point detection, target segmentation, sliding window, and component integration method [1]. Among them, the way of combining saliency map detection and the sliding window has excellent detection results, it is the current research hotspot of UAV target detection [2].

In the target detection algorithm based on the combination of saliency map detection and sliding window, the quality of target detection greatly depends on the detection result of the saliency map. Saliency map detection methods usually include two types: one is a bottom-up method based on the underlying features [3-6], and the other is a top-down method based on high-level prior knowledge. The low-level features refer to the color and texture of the image, and the high-level prior knowledge refers to the background experience or semantic information. The bottom-up method is mainly based on the underlying data information and does not spontaneously provide visual saliency information, and the detected saliency map is relatively complete. The top-down method faces a variety of salient targets, due to the limitation of prior knowledge, its generalization ability is limited.

[*] Corresponding Author

In recent years, bottom-up saliency research has attracted the attention of domestic and foreign scholars. ITTI et al. [7] proposed a saliency detection algorithm based on visual attention. Ma et al. [8] proposed an image saliency analysis based on the contrast-based fuzzy growth method. Harel et al. [9] proposed a graph-based saliency detection algorithm. The algorithm used the features of the ITTI algorithm in the feature extraction process, and it introduced the Markov chain to calculate the saliency map. Hou et al. [10] used spectral margin to detect significance. The focus of these algorithms is on the salient points in the image, so the salient objects in the saliency map are blurred.

Later, many scholars paid attention to salient objects or regions in the image. Achanta et al. [11] used the difference of Gaussian method to extract the full-resolution saliency map. This method is simple and efficient, but it is not conducive to saliency object detection in complex scenes. Stas et al. [12] proposed a context-aware saliency detection algorithm. The edge features of the detected saliency objects are more obvious and the consistency is not high. On the basis of the context-aware saliency detection algorithm, paper [13] proposed a random visual saliency detection algorithm, and the algorithm has a faster calculation speed. Guo et al. [14] proposed a salient area detection algorithm based on color contrast. Existing relevant target detection algorithms have insufficient performance in target detection accuracy, tracking accuracy and adaptability in complex scenarios, which brings huge difficulties and challenges to the practical application of UAVs.

Therefore, the current image saliency area detection research has achieved many new results, but it is still worthy of further study, especially in the complex scenes of UAV aerial image target detection. Paper [15] proposes a joint saliency detection method to detect common saliency objects in the image set. Nonclercq et al. [16] used the Bayesian framework to extract salient areas of the image. This algorithm improves the detection effect, but the detection effect is not good when the salient points are not around the salient objects. Cheng et al. [17] proposed a method to detect the saliency of images, and the method combined with regional color distance metric and spatial distance metric, but the detection effect on images with complex background needs to be improved.

Different from above relevant works, this paper proposes a UAV aerial image target detection algorithm combining visual attention with Dw-adaboost, the algorithm in this paper can realize UAV aerial target detection in complex scenarios. For the actual needs of UAV target detection, the underlying feature information is extracted through the visual attention, and the Dw-adaboost determines the weight of each feature channel, then we use efficient sub-window search on the saliency map to achieve detection result.

For the problem that UAV aerial image targets are small and easily interfered by background and illumination changes, the main contribution of this paper are summarized as follows.

(1) We propose an image saliency detection algorithm based on visual attention combined with Dw-adaboost, this algorithm makes the underlying information of the image better expressed.

(2) A Dw-adaboost classification algorithm is proposed. This algorithm improves the learning ability of the adaboost algorithm for multiple clusters distribution data set.

(3) This paper algorithm improves the problem that UAV aerial image targets are small and easily interfered by background and illumination changes. It can achieve more accurate detection of UAV aerial image targets in complex scenes.

The remainder of this paper is structured as follows: In Section 2, we introduce the saliency map detection algorithm of visual attention combined with Dw-adaboost in this paper. In Section 3, we use high-efficiency sub-window search on the saliency map to achieve target detection in UAV aerial images. Section 4 is the algorithm flow chart of this paper. In Section 5, we provide the experimental setting, results and analyses. Finally, we conclude the paper and talk about the future work.

2 Visual Attention Combined with Dw-adaboost Image Saliency Detection Algorithm

In this section, we first introduce visual attention combined with Dw-adaboost image saliency detection framework. Then based on visual attention, the feature extraction methods of brightness, direction, regional contrast, and spatial location features are introduced respectively. Finally, we introduce Dw-adaboost and gave the algorithm steps of it.

2.1 Visual Attention Combined with Dw-adaboost Image Saliency Detection Framework

Feature extraction is the basis for analyzing saliency maps. Three typical characteristic channels of brightness, color and direction are used in the ITTI visual attention model. In actual UAV aerial image target detection, the target in the UAV scene is small, and the target is easily affected by background factors and illumination changes. Visual attention combined with Dw-adaboost image saliency detection framework is shown in Fig. 1.

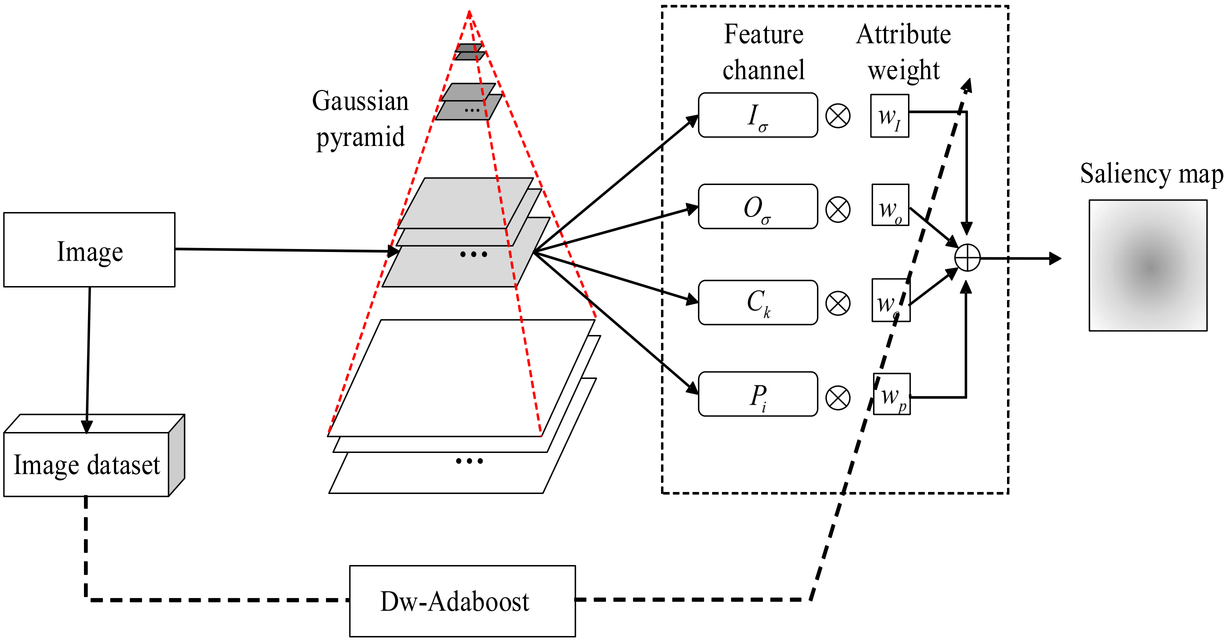


Fig. 1. Visual attention combined with Dw-adaboost image saliency detection framework

As shown in Fig. 1, the entire frame is divided into two main parts.

In the first part, the input image passes through the Gaussian pyramid. According to the needs of UAV target detection, this paper selects brightness (I_σ), direction (O_σ), regional contrast (C_k) and spatial location features $(P)_i$ as the main feature channels to generate saliency map.

In the second part, in order to improve the accuracy of target detection in the same scene, the input image is composed of a data set. Dw-adaboost uses this data set for training, determines the weight of each feature channel and finds its optimal value, thereby achieving image saliency detection.

2.2 Feature Extraction Based on Visual Attention

In UAV aerial images, the size of the target is quite different. When the vision system analyzes unknown complex scenes, it is often unable to obtain scale information of the target. In order to describe the target at multiple scales, it is necessary to sample the image using a Gaussian pyramid [18].

The image is first smoothed by a low-pass filter, and then different scales filters are combined to form a Gaussian pyramid structure. The entire pyramid is divided into nine layers. In x and y directions, the zoom factor is the 1/2 of the previous layer, and we use Gaussian blur down-sampling to decrease the resolution, and the resolution of the smallest layer is 1/256 of the original image. In the Gaussian pyramid, the tier σ of the Gaussian pyramid image $G_k(x,y)$ can be expressed as:

$$G_k(x,y)=\sum_{m=-2}^{2}\sum_{n=-2}^{2} w(m,n)\,G_{k-1}(2x+m,2y=n), \tag{1}$$

where $w(m,n)$ is the low-pass window function of 5×5, which can be expressed as $w(m,n) = h(m)\cdot h(n)$, and h is the Gaussian density distribution function, and it satisfies normalization and symmetry, so $w(m,n)$ can be expressed as:

$$w(m,n)=\begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix}. \tag{2}$$

After constructing the Gaussian pyramid, different scales of feature maps are obtained, and it requires a center-edge operation to calculate the saliency map. The center edge operation is as follows:

(1) First of all, we transform the small scale into the large scale through interpolation.

(2) Secondly, we perform point-to-point subtraction on feature map [19]. Assuming that the tier c of the Gaussian pyramid represents the central layer, the value can be $\{2,3,4\}$. Tier $c+s$ represents the peripheral layer of the pyramid structure, and the value s can be $\{3,4\}$. The following describes the brightness, direction, regional contrast, and spatial location feature.

Brightness Feature. Brightness is the perceptual property of the visual system's radiation or luminescence of visible objects. Without brightness, other underlying features such as shape and color cannot be expressed. Brightness is an important underlying feature that affects the extraction of the image region of interest. The brightness layer I_σ can be obtained by performing multi-layer Gaussian filtering on the brightness feature, it can be expressed as:

$$I_\sigma = \frac{r+g+b}{3} * G_\sigma(x,y), \tag{3}$$

where r, g, b is the red, green, and blue color channel feature, and they are extracted from image.

Direction Feature. When constructing direction features, because the Gabor function has frequency localization and directivity, a Gabor filter with better direction selectivity is selected to extract direction features. The tier σ Gabor function of the Gaussian pyramid can be expressed as:

$$G(x_0,y_0,\theta) = e^{-\frac{x_0}{2\sigma_x^{\,2}}-\frac{y_0}{2\sigma_y^{\,2}}}(\cos(2\pi f x_0) + j\sin(2\pi f y_0)), \tag{4}$$

where $x_0 = x\cos\theta + y\sin\theta,\ \ y_0 = y\cos\theta - x\sin\theta,\ \ \theta_k = \frac{\pi}{n}(k-1), k=1,2,...,n.$ θ is the main frequency direction of the filter, f represents the center frequency of the filter, and σ_x, σ_y represents the Gaussian variance in the horizontal and vertical directions. When $n=4$, the output of the Gabor function in the four directions of $0^\circ, 45^\circ, 90^\circ, 135^\circ$ as the direction feature, the direction feature $O_\sigma(\theta)$ can be expressed as:

$$O_k(\theta) = I_k * G(\theta), \theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}. \tag{5}$$

Regional Contrast Feature. For an image, people will pay more attention to the area where the contrast is very different from the surrounding objects [20]. This paper considers the influence of color distance in detecting the saliency area, the image area contrast feature function C_k can be expressed as:

$$C_k = \sum_{\substack{m=1 \\ m\neq k}}^{n} D_r(r_k, r_m), \tag{6}$$

where r_k, r_m is the k-th and m-th area in the image, and C_k is the contrast feature function of the k-th area r_k. $D_r(r_k, r_m)$ is the contrast between r_k and r_m, which can be expressed as:

$$D_r(r_k, r_m) = \sum_{i=1}^{n_k} \sum_{j=1}^{n_m} \frac{p_k(i) \cdot p_m(j) \cdot \| r_k(i) - r_m(j) \|^2}{1 + d_r(r_k, r_m)}, \tag{7}$$

where $d_r(r_k, r_m)$ is the distance between r_k and r_m. $r_k(i)$ and $r_m(j)$ are the color vectors of the regions r_k and r_m in the RGB color space, and n_k, n_m are the total number of colors contained in r_k and r_m, n is the total number of image areas, $p_k(i)$ is the probability of $r_k(i)$ appearing in the image area r_k, and $p_m(i)$ is the probability of $r_m(i)$ appearing in the image area r_m.

Spatial Location Feature. In visual attention, humans often turn the angle of view to focus the object. Similarly, in the process of image acquisition, people usually focus on the object that attracts attention [21]. This paper introduces the location information of the image area into the saliency detection process. Specifically, the spatial location feature function is constructed using the distance relationship between the region and the image center, so that the area near the center of the image has high saliency, and the region farther from the center has less saliency. The Spatial location feature p_i of the normalized image area is defined as follows:

$$p_i = 1 - \sqrt{\left(\frac{x_i}{h} - \frac{1}{2}\right)^2 + \left(\frac{y_i}{w} - \frac{1}{2}\right)^2}, \tag{8}$$

where p_i is the location characteristic function of the i-th area r_i, $p_i \in [0, 1]$. (x_i, y_i) is the centroid coordinate of the i-th area r_i, and h, w is the height and width of the entire image.

2.3 Dw-adaboost Classification Algorithm

Based on extracting each feature of the image, to make the UAV aerial image target detection task perform well in different scenarios, we uses the adaboost algorithm to train the image data set to obtain the optimal weight [22]. The adaboost algorithm is the most representative and popular boosting algorithm in current ensemble learning. Once its multiple weak classifiers are trained, the weights of the weak classifiers are fixed. Due to the preference of classifiers, although some weak classifiers have good classification ability on the entire training set, they may lack the learning of a certain local area, resulting in low classification ability for this area, especially for some showing multiple clusters distribution data set [23-25].

To the above problems, we proposes a Dw-adaboost classification algorithm. Based on a weak classifier with fixed weights, it adds dynamics that can measure the adaptability of the weak classifier to the sample. Dw-adaboost is shown in Fig. 2. The left of Fig. 2 is the image data set, we use D_{w1} to train the weak classifier h_1, then update the sample weight distribution to obtain D_{w2}, and then use D_{w2} to train the classifier h_2. By iterating T times in this way, T weak classifiers with fixed weight α are obtained and β is the dynamic weight of the weak classifier, and β is the classification accuracy of multiple training samples around the sample to be tested. There are many ways to combine weights, many experiments have found that different combinations of weights have little effect on the classification results. Therefore, this paper uses the simplest multiplicative normalization method, and the voting weight of each weak classifier is $\alpha \times \beta$.

The Dw-adaboost classification algorithm training process in this paper retains all the weak classifiers trained. In the classifier combination stage, we add dynamic weights that can measure the adaptability of the sample and each weak classifier, the overall weight of the weak classifier is obtained by multiplying the static weight and the dynamic weight of the sample. We use the overall weight to determine the category of the sample. The Dw-adaboost algorithm process is described in Algorithm 1:

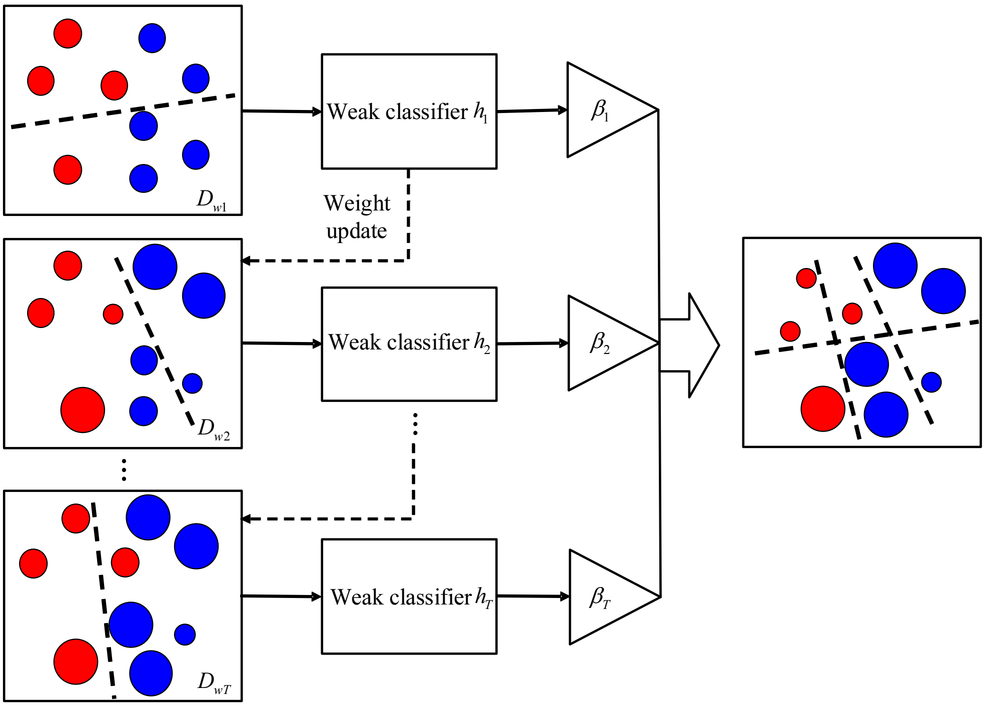


Fig. 2. Schematic diagram of the Dw-adaboost classification algorithm

Algorithm 1. The Dw-adaboost classification algorithm process

Input: Set D=$\{(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)\}$, among them, $x_i \in \chi \subseteq \mathrm{R}^n, y_i \in Y = \{-1, +1\}$.
Output: Classifier collection $f(x)$.

1. Initialize sample weights : $D_1 = (w_{11}, ..., w_{1i}, ..., w_{1N}),\ w_{1i} = \frac{1}{N}, i = 1, 2, ..., N$.
2. Do the following operations on $m = 1 \sim T$:
   (a) Use the training set of weighted redistribution D_m to learn the weak classifier $h_m(x): \chi \to \{-1, +1\}$.
   (b) Calculate the classification error of $h_m(x)$ on the training set:

$$e_m = P(h_m(x_i) \neq y_i) = \sum_{i=1}^{N} w_{mi} I(h_m(x_i) \neq y_i).$$

   (c) Calculated fixed weight h_m :

$$\alpha_m = \frac{1}{2} \ln \frac{1 - e_m}{e_m}.$$

   (d) Update the training set weight distribution:

$$w_{m+1,i} = \begin{cases} \frac{w_{mi}}{Z_m} e^{-\alpha m}, h_m(x_i) = y \\ \frac{w_{mi}}{Z_m} e^{\alpha m}, h_m(x_i) \neq y \end{cases},$$

   where Z_m is the norm factor, $Z_m = \sum_{i=1}^{N} w_{mi} \exp(-\alpha_m y_i h_m(x_i))$.
3. Train set to get T weak classifiers $H = \{h_1, h_2, ..., h_T\}$.
4. Find multiple samples from the training set that are similar to the sample to form a temporary test set D.

5. For $i = 1, 2, ..., T$, use the weak classifier h_m to classify and test D respectively, and calculate the dynamic weight:
$$\beta = \frac{N_k}{Num},$$
where N_k is the number of correctly classified samples in k neighboring points, and Num is the number of neighboring samples.
6. Combine fixed weights and dynamic weights to obtain a linear combination classifier for the sample:
$$f(x) = \sum_{m=1}^{T} \alpha_m h_m(x) \beta_m .$$

3 High-Efficiency Sub-Window Search

After generating a saliency map based on visual attention combined with Dw-adaboost. The efficient sub-window search with the maximization of significant area density is used to achieve target detection [26].

The density of the salient target area is greater than the density of the background area, the objective function can be designed to search for the area with the largest salient density. The search window W^* is designed as:

$$W^* = \arg\max_{W \subseteq I} f(W), \tag{9}$$

where W is the search window and $f(W)$ is a target search function.

It is a waste of time to traverse and search W^* from the image, and an efficient sub-window method can be used. Let $W = \{W_1, W_2, ..., W_i\}$ denote the region set, where $W_i \subseteq I$. Suppose that there are two regions $W_{\min}$ and $W_{\max}$ in W, so that $W_{\min} \subseteq W_i \subseteq W_{\max}$. f is the last estimate of the optimal solution, and it can be expressed as:

$$f(W) = \frac{\sum_{(x,y)\in W_{\max}} S(x,y)}{\sum_{(x,y)\in I} S(x,y)} + \frac{\sum_{(x,y)\in W_{\max}} S(x,y)}{C_0 + W_{\min}}, \tag{10}$$

where C_0 is a positive constant, it is used to balance the influence of area $W_{\min}$, $S(x, y)$ is the image saliency map, (x, y) is the pixel in the saliency image, the first term is to let the detection area contain more salient pixels, and the second term is to ensure that the detection area has a high salient density. Therefore, maximizing $f(W)$ balances the relationship between target size and target inclusion significance.

4 Algorithm Flow Chart

The algorithm in this paper is shown in Fig. 3. The whole algorithm is divided into the training phase of Dw-adaboost and the test phase of the input image.

The training phase is the process of Dw-adaboost using image data sets for training and learning. First, we initialize the sample weights, and get the fixed weights of the weak classifier after training on the training set; Then, we update the weights of the training set to obtain the dynamic weights; Finally, we get the feature channel linear classifier $f(x)$.

The test phase is the process of object detection on the input image. Firstly, a Gaussian pyramid is constructed to extract the underlying features, and the channel feature values are normalized sampling; Secondly, the feature channel feature values are converted into column vectors, and the linear combination classifier $f(x)$ in the training phase is used to determine the feature channel weights; Finally, we use efficient sub-window search method to accurately locate the target position, and realize the target detection of UAV aerial image.

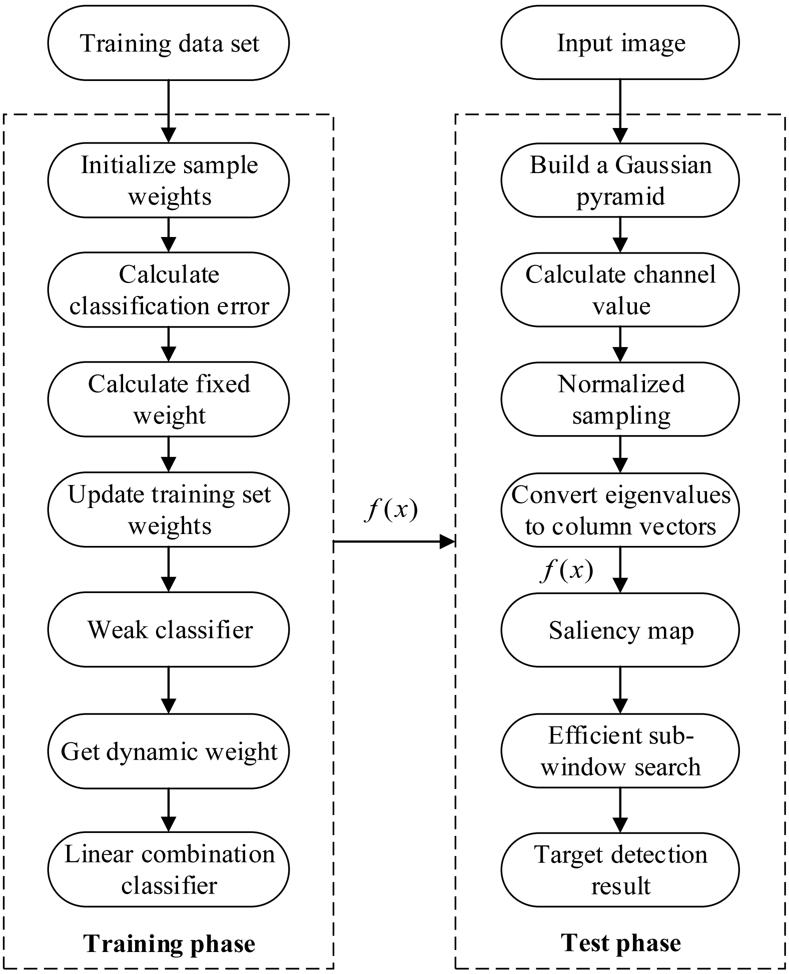


Fig. 3. Algorithm flow chart of this paper

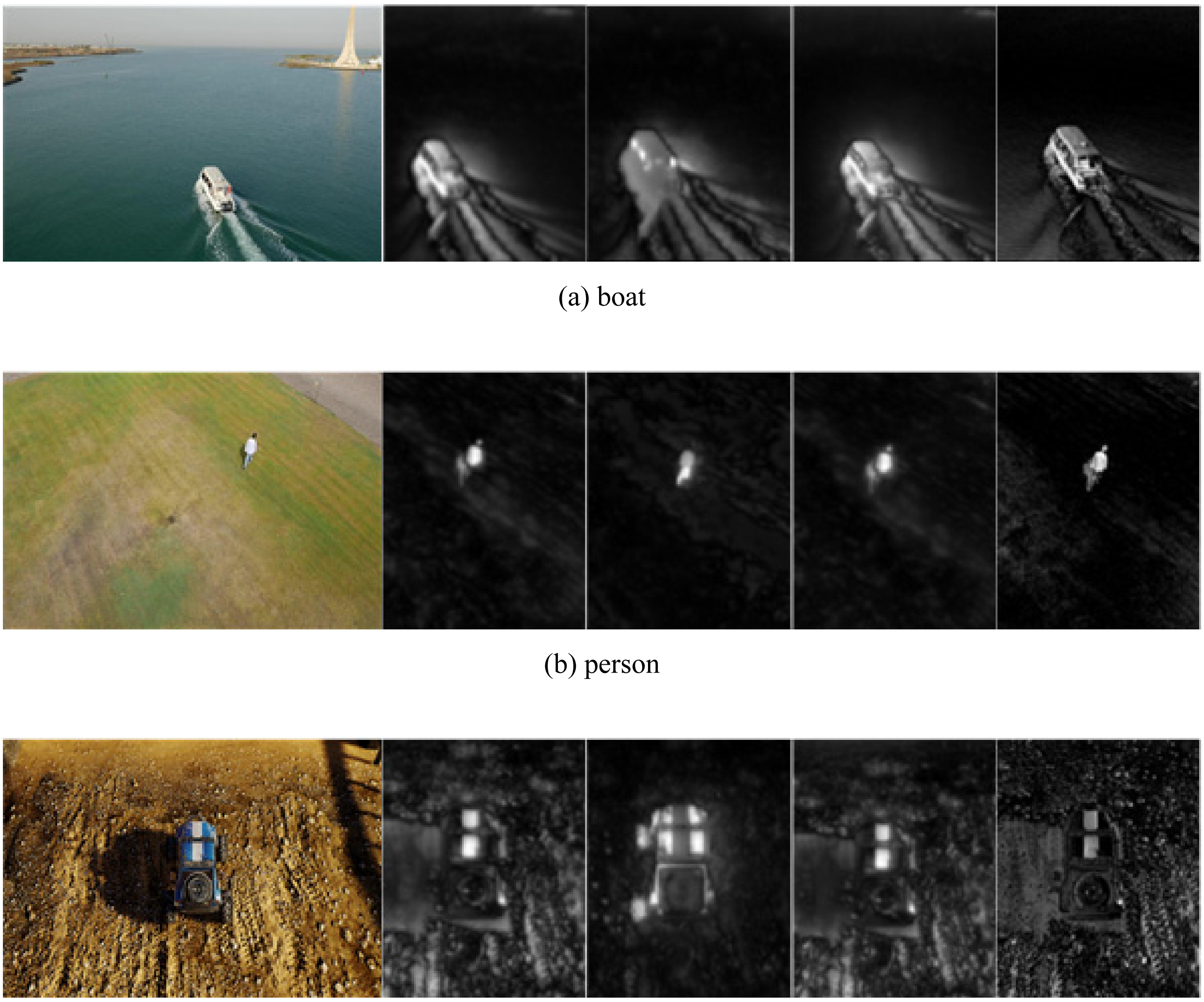
5 Analysis of Experimental Results

To effectively evaluate the performance of the algorithm in this paper, the comparison algorithms in the experiment all come from the public code of the corresponding author. All experiments are performed on Inter(R) Core(TM) i5-2450M CPU@ 2.50GHZ, 4GB memory, NVIDIA GeForce GTX 1660 Ti computer. The experiment verifies from three aspects: feature extraction comparison, saliency map algorithm performance comparison, and target detection results.

5.1 Feature Extraction Comparison

Aiming at the complex scenes of UAV aerial images, we selected the UAV123 dataset and the video image sequence collected by the UAV as the test dataset [27]. The UAV123 data set contains twelve attribute changes involving small targets, background dynamic interference, in-plane rotation, and occlusion et al. These aerial images are taken by UAV in a low-altitude environment, and they can meet the needs of the algorithm simulation.

Fig. 4 shows the saliency map of the UAV123 dataset. The first column is the original aerial image, and from left to right are the brightness feature, direction feature, regional contrast feature, and spatial location feature. Fig. 4(a) is an aerial image of a ship target in the background of the island shore, there is clutter interference from the wave. In the process of extracting the bottom layer feature of the target, it is easy to cause inconsistent highlighting between the target and the background; Fig. 4(b) is the case where the target is small in the aviation background, and the target occupies a small proportion in the entire detection screen; Fig. 4(c) is the case where target is affected by light in a complex scene, the target and the background have a high regional contrast.



(a) boat

(b) person

(c) car

Fig. 4. Saliency map of the aerial image feature extraction under UAV123 dataset

From the results of the bottom-level feature extraction, the brightness feature, direction feature, regional contrast feature, and spatial location feature of the algorithm in this paper all have good and significant effects on aerial images, as follows:

(1) Brightness Feature. As a necessary feature for bottom-level feature extraction, for the case where the clutter interference of the ship background of the boat sequence in Fig. 4(a), and the target of the person sequence in Fig. 4(b) is small, the brightness feature can suppress the clutter interference of the background and highlight smaller targets.

(2) Direction Feature. As shown in the car sequence in Fig. 4(c), the direction feature highlights the blue part of the car. This feature can replace the color feature. At the same time, this feature performs better when the target is affected by the light. To a certain extent, the problem that the target is affected by light is solved.

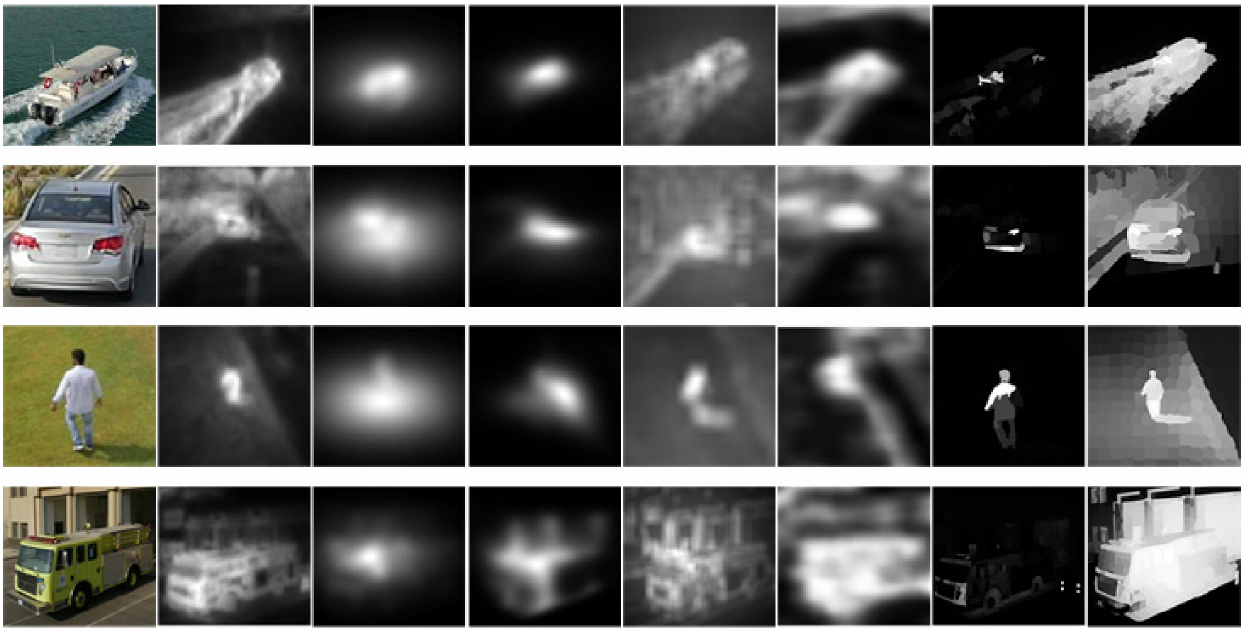
(3) Regional Contrast Feature. Different from the brightness and direction features, the regional contrast feature has a partial improvement for background clutter interference, small targets and targets affected by illumination.

(4) Spatial Location Feature. As can be seen from Fig. 4, the spatial location feature can well suppress the clutter interference of the background, make the target clear, and effectively distinguish the location information of the target and the background to a certain extent.

5.2 Performance Comparison of Saliency Map Algorithms

Brightness, direction, regional contrast and spatial location feature can highlight the image target, but this does not guarantee the effect of the salient map. In order to ensure the final quality of the saliency map. This paper uses Dw-Adaboost for data set learning to determine the weight of each feature.

In order to fully illustrate the superiority of the image saliency detection algorithm of visual attention combined with Dw-adaboost in this chapter, we select six typical saliency map algorithms (CA, COV, HFT, ITTI, MC, SF) [28-29] and four data set sequences (boat, car, person, truck) to verify the algorithm. Fig. 5 shows the saliency map detection results of a total of seven methods including the algorithms in this paper.



(a) Origin (b) CA (c) COV (d) HFT (e) ITTI (f) MC (g) SF (h) Ours

Fig. 5. Comparison of the algorithm in this paper with six typical saliency map detection methods

The experimental results of the four sequences as follows:

(1) boat. It can be seen from the results that the saliency map for boat. The image results of CA, COV, HFT, ITTI, and MC are all too blurred. At the same time, the CA, ITTI, and MC algorithms cannot effectively suppress background interference. The image results of SF are not blurred, but the image cannot be uniformly highlighted. The algorithm in this paper can effectively suppress the background interference on both sides of the boat, and it has good results.

(2) car. For the saliency map of car, the image results of CA, ITTI, and MC are too blurry, COV and HFT cannot accurately determine the target position, and SF cannot achieve consistent image highlighting. The algorithm in this paper can effectively distinguish between target and background.

(3) person. For the saliency map of person, the image results of CA, COV, HFT, ITTI, and MC are all too blurry, and the image results of SF cannot fully display the target. The algorithm in this paper highlights the goal.

(4) truck. For the saliency map of truck, the image results of CA, COV, HFT, ITTI, MC are also too blurry, and the SF target is too dark. The algorithm in this paper can realize the discrimination of the target.

In summary. Compared with the other six methods, the algorithm in this paper is clearer than the other six methods. The target and background are not blurred, and it has target background clutter suppression. The target saliency map can achieve uniform highlighting.

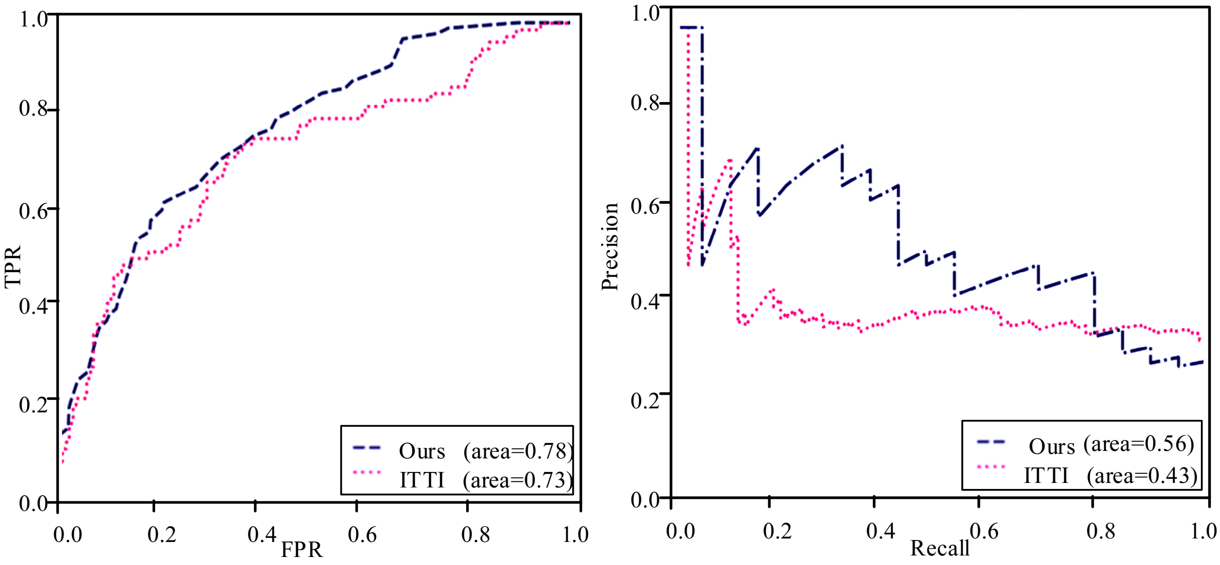
5.3 Analysis of Target Detection Results

Based on the good detection effect of a single sequence target, in order to measure the performance of the target detection algorithm on the UAV123 data set, the target to be detected is defined as a positive sample and the background is defined as a negative sample, and the target detection algorithm is tested. Fig. 6 shows the comparison of ROC and PRC curves between the algorithm in this chapter and the target detection algorithm of the ITTI model.

The ROC curve uses FPR as the abscissa and TPR as the ordinate, as shown in Fig. 6(a) is the ROC curve. The ROC curve must pass through points $(0,1)$ and $(1,0)$. The more convex, and the curve is toward the upper left corner, the better the classification effect of the target detector.

The PRC curve is shown in Fig. 6(b), in the target detection, the change of the recognition threshold will cause the precision and recall values to change. Take recall as the abscissa and precision as the ordinate to get the PRC curve.

As shown in Fig. 6, the area under the ROC curve is AUC, and it is a probability value. The larger the AUC value, the more likely the current classification algorithm will rank the positive samples in front of the negative samples, that is, the classification algorithm can better achieve classification. AP is the area under the PRC curve. The better the target detector, the greater the AP value. It can be seen from the curve that the AUC and AP values of the algorithm in this paper are 0.78 and 0.56. Compared with the ITTI method, the AUC and AP values are better than the ITTI method.



(a) ROC

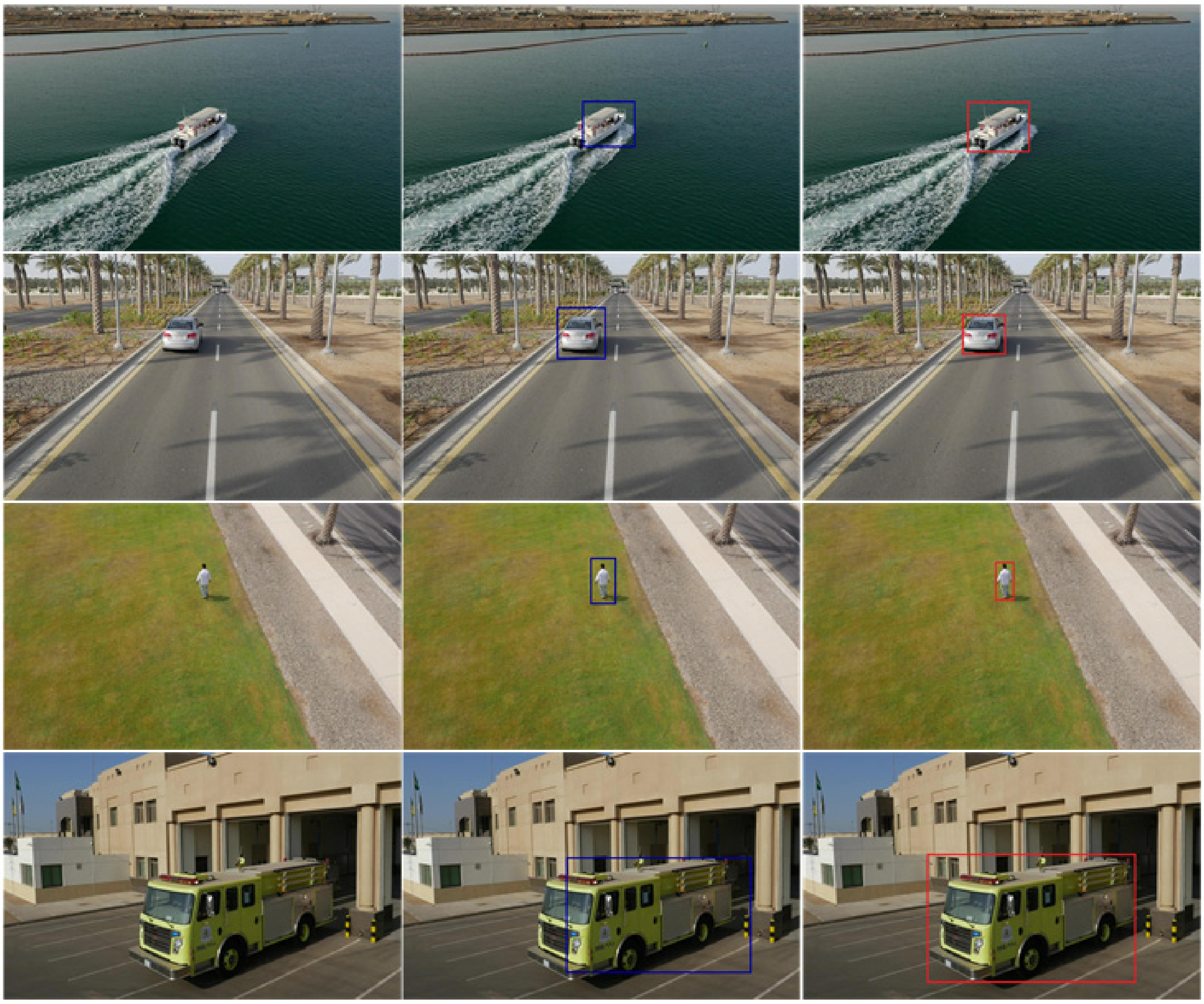
(b) PRC

Fig. 6. Comparison of ROC and PRC curves

Based on the saliency detection results of the method in this paper, an efficient sub-window search method is used for target detection. As shown in Fig. 7, the ITTI method does not perform well in target detection in island shore scenes and architectural scenes. The target parts of ships and trucks are incomplete and lost. The target detection effect of cars and pedestrians has improved, but the bounding box covers the background area.

To more intuitively verify the results of the target detection experiment, we use the center point distance error method to objectively measure the ITTI method and the method in this paper, and draw the data into a histogram for comparison. The following is the center distance error data of a typical image.

It can be seen from Table 1 and Fig. 8 that the center distance error value of the method in this paper is less than that of the ITTI method. For the situation where there is background clutter interference, the target is small and easily affected by light changes. The method in this paper has great detection results. For boat, car, person and truck sequences, compared to ITTI method, the center distance error of this algorithm is reduced by 25.37, 35.55, 35.09 and 8.49.



(a) Original image (b) ITTI algorithm (c) Ours

Fig. 7. Detection results of different salient target detection methods

Table 1. The center distance error between the method in this paper and the ITTI method

Image	Method	GT	PGT	Error
boat	Ours	539 298 722 424	539 296 721 426	3.00
	ITTI		567 300 723 428	28.37
car	Ours	508 177 648 289	511 179 645 298	10.15
	ITTI		499 149 613 286	45.81
person	Ours	618 173 678 286	620 177 672 282	8.49
	ITTI		655 196 679 286	43.58
truck	Ours	402 275 1069 668	406 277 1065 665	6.71
	ITTI		387 274 1068 670	15.20

In summary, the algorithm in this paper can achieve target detection of UAV aerial images in complex scenes. The main advantages are shown in the following aspects:

(1) The brightness, direction, regional contrast and spatial location features selected in this paper can better express the image target information.

(2) The saliency map of the algorithm in this paper is clearer, which can effectively distinguish the target and the background information, and the target is uniformly highlighted in the saliency map.

(3) The algorithm in this paper can achieve accurate target detection in complex scenarios where the target is small and the target is easily interfered by background factors.

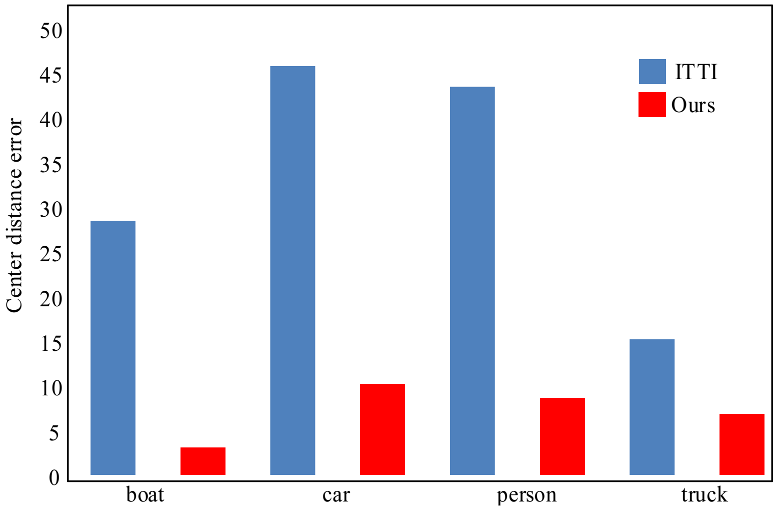


Fig. 8. Comparison of the center distance error

6 Conclusion

In this paper, we use visual attention combined with the Dw-Adaboost for UAV aerial image target detection. According to the actual needs of UAV target detection, firstly, the visual attention mechanism is used to extract the underlying feature information, and then the Dw-Adaboost is proposed to determine the weight of each feature channel, and it improves the effect of feature extraction and fusion. Finally, we use efficient sub-window search on the saliency map to achieve aerial image target detection. The experiments at the end of the article show that this method can achieve accurate detection of UAV aerial image targets in complex scenes, effectively suppress clutter interference, and detect targets more accurately.

In future work, the algorithm in this paper can realize the detection of a single target in a complex scene, but for a single target that needs to be tracked specifically among multiple targets, tracking based on the saliency map method cannot accurately locate the target, so the detection problem for a specific target is still further research is needed.

References

[1] B. Hariharan, P. Arbelaez, R. Girshick, J. Malik, Simultaneous detection and segmentation, in: Proc. 2014 European Conference on Computer Vision, 2014.

[2] K. He, G. Gkioxari, P. Dollar, R. Girshick, Mask RCNN, in: Proc. 2017 IEEE International Conference on Computer Vision, 2017.

[3] C.-L. Guo, Q. Ma, L.-M. Zhang, Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform, in: Proc. 2008 IEEE international Conference on Computer Vision, 2008.

[4] H. Xin, H.-Y. Jing, Q. Han, Errata: Salient region detection combining spatial distribution and global contrast, Optical Engineering 51(4)(2012) 47-57.

[5] C. Zhang, X.-C. Zhang, M. Tan, Infrared image segmentation method based on spatial coherence histogram and maximum entropy, The International Society for Optical Engineering 20(2014) 1-8.

[6] M.-M. Cheng, G.-X. Zhang, N.-J. Mitra, Global contrast based salient region detection, in: Proc. 2011 IEEE international Conference on Computer Vision, 2011.

[7] L. Itti, C. Koch, E. Niebur, A model of saliency based visual attention for rapid scene analysis, Pattern Analysis and Machine Intelligence 20(11)(1998) 1254-1259.

[8] Y.-F. Ma, H.-J. Zhang. Contrast-based image attention analysis by using fuzzy growing, in: Proc. 2003 Eleventh ACM International Conference on Multimedia, 2003.

[9] J. Harel, C. Koch, P. Perona, Graph-Based Visual Saliency, in: Proc. 2007 Neural Information Processing Systems Conference, 2007.

[10] X. Hou, L. Zhang, Saliency Detection: A Spectral Residual Approach, in: Proc. 2007 IEEE International Conference on Computer Vision & Pattern Recognition, 2007.

[11] R. Achanta, S. Hemami, F. Estrada, Frequency-tuned salient region detection, in: Proc. 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[12] S. Goferman, M.L. Zelnik, A. Tal, Context-Aware Saliency Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 34(10)(2012) 1915-1926.

[13] H. Zhiyong, H.E. Fazhi, C. Xiantao, Efficient random saliency map detection, China Information 06(2011) 1207-1217.

[14] G. Yingchun, Y. Haojie, W.U. Peng, Image Saliency Detection Based on Local and Regional Features, Acta Automatica Sinica 39(8)(2013) 1124-1129.

[15] Z. Liu, W. Zou, L. Li, Co-Saliency Detection Based on Hierarchical Segmentation, IEEE Signal Processing Letters 21(1)(2013) 88-92.

[16] A. Nonclercq, M. Foulon, D. Verheulpen, C.D. Cock, M. Buzatu, P. Mathys, P.V. Bogaert, Cluster-based spike detection algorithm adapts to interpatient and intrapatient variation in spike morphology, Journal of Neuroence Methods 210(2)(2012) 259-265.

[17] S.-P. Zhang, H.-X. Yao, X. Sun, X.-S. Lu, Sparse coding based visual tracking: review and experimental comparison, Pattern Recognition 46(7)(2013) 1772-1788.

[18] K. He, X. Zhang, S. Ren, J. Sun, Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 37(9)(2015) 1904-1916.

[19] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, Focal Loss for Dense Object Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 42(2)(2020) 318-327.

[20] X. Li, H. Lu, L. Zhang, X. Ruan, M.H. Yang, Saliency Detection via Dense and Sparse Reconstruction, in: Proc. 2013 IEEE International Conference on Computer Vision, 2013.

[21] T. Judd, K. Ehinger, F. Durand, A. Torralba, Learning to predict where humans look, in: Proc. 2010 IEEE International Conference on Computer Vision, 2010.

[22] M. Woniak, M. Grana, E. Corchado. A survey of multiple classifier systems as hybrid systems, Information Fusion 16(2014) 3-17.

[23] Y. Gao, F. Gao, Edited AdaBoost by weighted KNN, NEUROCOMPUTING 73(16-18)(2010) 3079-3088.

[24] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines, ACM Transactions on Intelligent Systems & Technology 2(2011) 1-27.

[25] L. Didaci, G. Giacinto, F. Roli, G.L. Marcialis, A study on the performances of dynamic classifier selection based on local accuracy estimation, Pattern Recognition 38(11)(2005) 2188-2191.

[26] T.-L. Song, X.-L. Zhen, J. Ning, Target Segmentation of Infrared Image Using Fused Saliency Map and Efficient Subwindow Search, Acta Automatica Sinica 44(12)(2018) 2210-2221.

[27] I. Aicardi, F.C. Nex, M. Gerke, A.M. Lingua, An Image-Based Approach for the Co-Registration of Multi-Temporal UAV Image Datasets, Remote Sensing 8(9)(2016) 779.

[28] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[29] J. Li, M.D. Levine, X. An, X. Xu, H. He, Visual Saliency Based on Scale-Space Analysis in the Frequency Domain, IEEE Transactions on Pattern Analysis and Machine Intelligence 35(4)(2013) 996-1010.