Journal of Computers Vol. 34 No. 4, August 2023, pp. 69-82
doi: 10.53106/199115992023083404006

Moving Target Detection Algorithm for Dynamic Image Sequences on The Basis of Artificial Neural Network

Jia-Min Zhang[*], Yan-Xia Chen

School of Mathematics and Information Science, Zhangjiakou University, Hebei 075000, China
zhangjiamin@zjku.edu.cn, chenyanxia@zjku.edu.cn

Received 21 July 2022; Revised 29 December 2022; Accepted 10 February 2023

Abstract. Due to the large changes in dynamic image sequence frames and the complex detection scene, it is difficult to accurately detect moving objects. Therefore, the study proposes a moving target detection algorithm based on artificial neural network. First, the algorithm performs standardized grayscale processing and gamma correction processing on the dynamic image to eliminate the noise interference of the dynamic image. After that, the model calculates the gradient of the dynamic image in order to complete the feature extraction of the dynamic image. Then, according to the result of hog feature extraction, the study adopts the inter-frame calculation method to update the background of the dynamic image. Finally, the principle and structure of the neural network are analyzed experimentally, and a channel attention mechanism is introduced to train dynamic image sequences to obtain MTD results. Experimental results show that the proposed algorithm achieves higher accuracy in MTD than conventional detection algorithms. The calculation efficiency of the algorithm in this paper has significant advantages, and the average detection time is 3.69515ms, which can meet the real-time requirements of MTD.

Keywords: ANN, dynamic image sequence, moving target detection, gamma correction

1 Introduction

With the development of the times and the advancement of science and technology, people are exposed to more and more information in their daily lives, and visual information has attracted more attention. Images and videos are the most intuitive way for humans to obtain information and one of the most important sources of information [1-3]. The human brain analyzes and understands the image information perceived by the human eye, thereby filtering out invalid information and making corresponding judgments on useful information. The target detection task is realized by computer imitating the brain [4]. However, many factors such as background motion, illumination changes, and shadows in complex scenes limit the performance of traditional moving object detection (MTD) algorithms. Many scientists have proposed improvements to the MTD algorithm. Although they have achieved certain results, there are still many limitations. For example, although most research methods can detect moving objects, they are limited to static images, and it is difficult to detect moving objects in dynamic images. Therefore, current research in this field focuses on improving the robustness and detection performance of MTD algorithms in complex scenes [5]. In recent years, Artificial Neural Networks (ANNs) have been successfully applied to many vision tasks. Scholars from all walks of life have proposed improvements or optimizations on the basis of classic ANN to be more suitable for the research field, and its superior performance has promoted the development of various visual fields. Therefore, this research attempts to combine ANN and MTD, and proposes an MTD algorithm based on ANN for dynamic image sequences, aiming at improving the accuracy of moving target detection in complex environments.

2 Related Work

In recent years, the research on artificial neural networks has been quite hot, and many scholars have contributed to this. Asteris and other researchers used the ANN model to predict the compressive strength of concrete in existing structures. The overall model mainly predicts the compressive strength of concrete by means of

[*] Corresponding Author

ultrasonic pulse velocity and Schmidt rebound hammer. The results of ANN test are in good agreement with the actual situation, thus verifying the strong practicability of the ANN model [6]. Scholars such as Rostami S used the ANN model to predict the water nanofluid of multi-walled carbon nanotubes. Experiments predict the behavior of nanofluids by using the method of fitting curves and third-order surface equations. The experiment finally compared the ANN results and the fitting results, and found that the ANN has strong predictive ability, better correlation and smaller prediction error. This verifies the better practical performance of the ANN model [7]. Houssein et al. discussed three deep learning algorithms, Bayesian regularization, Marquardt method, and scaled conjugate gradient, and used them to train nonlinear autoregressive artificial neural networks. After this, the ensemble model is used in the forecasting of the Egyptian Stock Exchange Index. The final experimental results show that the proposed forecasting model can produce better forecasting accuracy than other models in both short-term and long-term forecasting [8]. From the perspective of predicting ground vibration, researchers such as H. Nguyen proposed a hybrid artificial intelligence model based on the combination of hierarchical K-means clustering algorithm and ANN, namely the HKM-ANN model. The model first uses the HKM algorithm to cluster the obtained data, and then uses the ANN algorithm to predict the ground vibration. Experiments compared the performance of the HKM-ANN model with other similar models. The results showed that the HKM-ANN model had the most superior performance in estimating the ground vibration caused by blasting operations [9]. Researchers Stamenković et al. constructed a model for predicting nitrate concentrations in rivers based on the ANN algorithm. The overall model consists of a multi-layer perceptron and a standard three-layer network. The data of ten water monitoring stations are selected in the experiment for the training and testing of the model. The final experimental results showed that the ANN model showed better performance in predicting the nitrate concentration in the river [10].

Moving target detection algorithm has attracted the attention of scientists in various fields, so its research has been continuously improved in recent years. Researcher Elhoseny introduces a new multi-target detection and tracking technique, which is constructed based on optimal Kalman filtering. The overall model mainly uses the region growing model to convert video clips into morphological operations according to the number of frames to distinguish objects, after which Kalman filtering is used to track moving objects in multiple video frames. Experimental results show that the model has achieved a tracking accuracy of 86.78% in target tracking and detection, and has good practicability [11]. Sun and other scholars conducted research on object detection of UAV captured traffic information. They found that the traffic images collected by drones have the characteristics of small size and high density, which makes it difficult for most existing algorithms to extract feature information from them. Therefore, the researchers built a small target detection algorithm based on YOLOv3d. The algorithm mainly predicts the position through the shallower feature map. Simulation results show that the proposed model has unsatisfactory detection accuracy for small targets [12]. Jha et al. introduced the target detection algorithm into the real-time monitoring system in the low-end edge computing environment. An N-YOLO detection model was constructed in the experiment. The model combines the detection results of the segmented subimages with the inference results at different times through the correlation tracking algorithm, so as to reduce the calculation amount of target detection and tracking. The simulation experiment results verified the feasibility and practicability of the model [13]. Researchers such as Dai et al. have found that a large amount of redundant data in the Internet of Vehicles will cause delays in data transmission and reduce the accuracy of object detection. Aiming at this defect, they propose an improved Haar-based feature classification algorithm, which can significantly reduce the delay. The final experimental results show that the proposed model can filter out 40% of similar frames, which is about 84 times faster, but the object detection accuracy still needs to be improved [14]. Scholars such as Ammar proposed a moving object detection method combining DeepSphere’s unsupervised anomaly discovery framework and generating adversarial networks. The method considers the morphological operation to achieve better segmentation and extraction of related objects. After this, the segmented objects are classified and recognized using the generative model. The experimental results of this model on the dataset show that it has broad application prospects in segmenting and classifying moving objects in video sequences [15].

To sum up, the ANN model has a wide range of applications in target detection, and its superior performance is favored by many researchers. At the same time, discussions on target detection algorithms have also been quite heated. Although the current research can improve the accuracy of target detection to a certain extent, there are still many limitations in general. Therefore, this study innovatively proposes an ANN-based MTD algorithm for dynamic image sequences, aiming to improve the accuracy of moving target detection in complex environments.

3 MTD in Dynamic Image Sequences

3.1 Dynamic Image HOG Feature Extraction

The feature extraction idea of HOG is to describe the characteristics of the target object in the image through the density distribution in the gradient direction. The HOG method first divides the image into the smallest units, namely cells. These minimal cells are then combined and their histograms of oriented gradients are counted. Finally, the HOG feature vector is obtained by concatenating all histograms of gradient orientations [16]. The statistical edge and gradient information of HOG can help the model extract local features very well. Compared with other feature extraction methods, the operation of HOG is carried out on the local grid of the image. The specific operation steps are shown in Fig. 1: first calculate the gradient direction and gradient size of all pixels in the image, and then divide the image into uniform cells (cells). Voting statistics are performed on the gradient direction in each cell, and each cell obtains a feature vector after voting statistics on nine gradient directions. Use a block containing several cells to delimit the image, and connect the 9-dimensional cell feature vectors in all blocks as the HOG feature vector of the image.

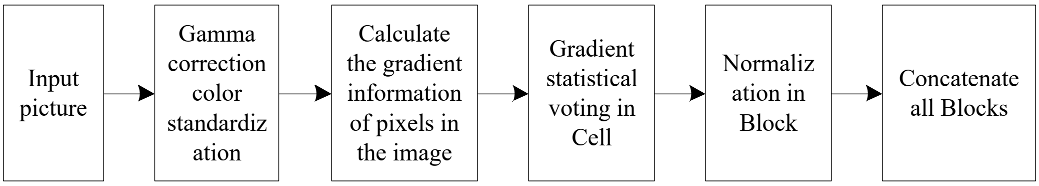

Input picture
Gamma correction color standardiz ation
Calculate the gradient information of pixels in the image
Gradient statistical voting in Cell
Normaliz ation in Block
Concatenate all Blocks

Fig. 1. HOG feature extraction steps

There may be significant differences in the intensity of light in different locations of a dynamic image. Gamma rectification of the image and color space normalization are required for the experiment to remove the negative impact brought on by this disparity. Gamma correction can suppress the interference caused by noise and reduce the impact of local shadows in dynamic images [10]. To calculate quickly and save memory, color images are usually normalized to grayscale images before Gamma processing. The calculation formula of Gamma correction is:

$$I(x,y)=I(x,y)^{gamma}\,. \tag{1}$$

The gradient size of the dynamic image is:

$$G_x(x,y)=H(x+1,y)-H(x-1,y)\,. \tag{2}$$

$$G_y(x,y)=H(x,y+1)-H(x,y-1)\,. \tag{3}$$

$$G(x,y)=\sqrt{G_x(x,y)^2+G_y(x,y)^2}\,. \tag{4}$$

The gradient direction of the dynamic image is:

$$\alpha(x,y)=\arctan\frac{G_y(x,y)}{G_x(x,y)}\,. \tag{5}$$

The gradient of each pixel is determined by its gray value, as can be shown from formula (4). The gradient size is the square of the difference between the gray value of the left and right pixels and the square of the difference between the gray value of the upper and lower pixels, and the sum of the two is then squared [11]. Using the above formula, these components are brought into it, and finally the gradient information of each pixel can be obtained. For each pixel on the image boundary, since there is no pixel outside the boundary for calculation, the value of the pixel on the boundary of the pixel can usually be equal to itself.

An image can consist of several small regions, each called a cell. In each cell, the gradient values of all its pixels are projected in different directions. The angle range of the projection direction may be split into 9 bins for best performance. Each cell contains nine gradient-weighted projections, that is, each cell can obtain a 9-dimensional feature vector. The cell size affects the detection steps after feature extraction, and the feature detection results extracted under different cell sizes are also different.

Interference caused by factors such as lighting and shadows will affect the gradient histogram calculated by each cell, resulting in large differences between adjacent cells. It is particularly important to further standardize batteries. In addition, normalization can further reduce the influence of confounding factors. Before normalization, a block needs to be defined, which contains several consecutive cells on the image. The normalization process operates as follows: By summing the nine eigenvalues of the blocks contained in the block, the nine total eigenvalues of the block can be obtained. For each cell, divide the 9 eigenvalues in each cell by the above value to achieve normalization. In addition, since the usual detection window is sliding, there will be many overlapping blocks. That is, the same cell feature will appear multiple times in the final feature vector summary, and the value of each occurrence is different. The block block obtained here is the HOG feature descriptor. All the obtained HOG feature descriptors are connected to form a one-dimensional feature vector, which is the final HOG feature vector.

3.2 Background Update

The method of inter frame calculation is adopted to calculate the “sum” of two images far apart to obtain the background image [17]. This method can raise the quality and clarity of the background model. The specific steps are as follows:

1. Select 10 consecutive images from the first frame as the sample set $M(x)$ initialized by the background modeling, and each image is $p_i (i = 1,...,10)$;

2. Select the image with a large distance between two frames, and calculate the “sum”, $N_j(x) = \left(p_j \cap p_{11-j}\right)(j = 1,2,3,4)$. Removing the adjacent 5th and 6th frames of images can ensure the high quality of background modeling initialization, and at the same time reduce the complexity of the algorithm, and then get the background of removing the moving target;

3. Calculate the “sum” of the two obtained images to the result $N_j(x)(j = 1, 2, 3, 4)$ of “sum”, $M(x) = N_1(x) \cup N_2(x) \cup N_3(x) \cup N_4(x)$. Since the two images are separated by a large distance, there will be no overlap where there is no background. This can be used to prevent incomplete background models caused by objects that have been stationary for a long time becoming moving objects. Through the above method, a complete and accurate background model can be obtained.

In the process of foreground detection, judging whether the pixel band is a moving area pixel is the key, and whether the moving object extraction is completed is also a key evaluation index of foreground detection. The moving target is extracted twice during the background modeling initialization procedure, and then an accurate and comprehensive background model is built. The steps of foreground detection are as follows:

1. Canny operator is adopted for detecting the current frame edge to simplify the target area and reduce the computational complexity of moving target extraction;

2. Due to the high accuracy and integrity of the background model, the background difference method is adopted for extracting the moving region edge;

3. Use morphological expansion to connect the broken edges of the moving region; After corrosion, the moving area is filtered out to eliminate the thin objects around the edge and make the boundary smooth;

4. Finally, fill the target movement area. First, search for each line of the extracted moving edge, find the first and last edge pixel of each line, and set all the middle pixels to 1. Then, find the top and bottom edge pixels of each column, and set the middle pixels to all 1s. Finally, the two motion areas are “summed” to complete the filling of the motion area.

The high-frequency interference in the backdrop becomes a significant element impacting MTD since the background often varies with time, light, and camera shaking. As time goes by, it is first necessary to determine whether to update the current background model. The strategy for detecting whether the background model needs to be updated is:

1. Calculate the difference between the current frame's background image of the moving target and the background model. If it is greater than a certain threshold $R_{\min}$, the background model changes, and the background model needs to be re established by using the continuous images from the current frame according to the background modeling initialization method; If it is less than a certain threshold $R_{\min}$, it indicates that the background model has not been changed due to the influence of environmental factors, as shown below:

$$C_i(x) = \left\| V_i(x) - M_i(x) \right\| . \tag{6}$$

$$\begin{cases} 1, if \quad C_i(x) \geq R_{\min} \\ 0, if \quad C_i(x) < R_{\min} \end{cases} . \tag{7}$$

In the formula (6), $V_i(x)$ represents the background image excluding the current frame of the moving target, $M_i(x)$ represents the background model, $C_i(x)$ represents the difference between the two background images, and $R_{\min}$ is the minimum threshold.

2. The binary image is obtained by removing the difference between the current frame's background image of the moving target and the background model. The number of pixels in the binary image is counted as 1. When the percentage of the number of pixels with value of 1 in the number of all pixels in the whole image is larger than the set threshold, it indicates that the background has changed on a large scale. If the percentage in adjacent images is still large, the background model will be reestablished from the current frame to the first frame according to the method of background modeling initialization in this paper.

3.3 MTD Algorithm for Dynamic Image Sequences on the Basis of ANN

The human brain from the perspectives of bionics, neurology, information and communication, automatic control principles, and organizational collaboration, and have constructed bionic neurons. In the process of exploring the human brain and studying artificial neurons, a new interdisciplinary technical field, namely "neural network", was gradually established. There are many fields involved in neural networks, and they are interconnected, interpenetrated, and mutually promoted. Scientists in various fields study neural networks from different perspectives according to the interests and characteristics of their respective disciplines [18-19].

ANN is a new force rising in the field of artificial intelligence. Its research can be traced back to the 1980s. The artificial neural network is based on the biological neural network, imitating the thinking principle of animals to process information. As an operational model, a neural network consists of multiple interconnected neurons. The network output varies depending on how the network is connected and the activation function. According to the complexity of the system, the artificial neural network completes information processing by adjusting the interconnection between multiple nodes in different network layers.

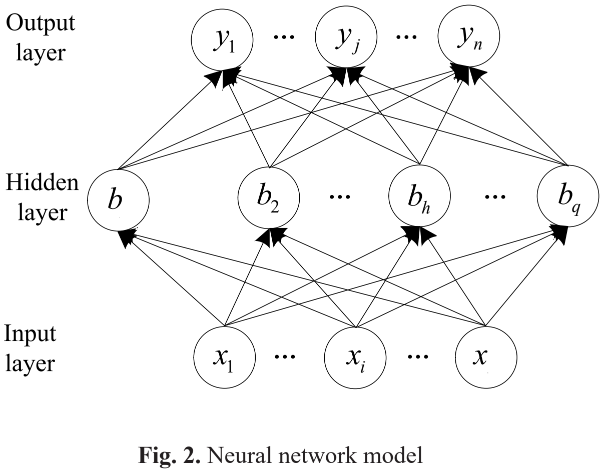


Fig. 2. Neural network model

A particular form of neural network known as an ANN learns and self-trains the network parameters to achieve the best desirable output outcomes given the input circumstances. The basic BP algorithm includes two processes: forward signal transmission and reverse error transmission [20]. In the process of backpropagating the error layer by layer, each layer shares the error and adjusts the weight so that the error decreases along the gradient direction. Until the error is minimized, the model stops learning and determines the network parameters. Fig. 2 shows the basic structure of the BP neural network model.

If the original dynamic image is X, H_i is adopted to indicate the characteristic graph of layer i of the neural network ($H_0 = X$). Then:

$$H_i = f(H_{i\ 1} \otimes W_i + b_i)\ . \tag{8}$$

In the formula (8), W_i indicates the convolution and weight vector of layer i . The operation symbol $\otimes$ indicates the convolution operation. b_i indicates the offset of layer i .

Through repeated convolution and alternating transmission of sampling layers, the ANN depends on the last full connection layer to reduce the dimension and match the categories. Finally, it classifies according to the extracted features so as to get the probability distribution Y based on input (where y_i represents the i -th label category):

$$Y(i) = P(L = y_i \,|H_0 ; (W, b))\ . \tag{9}$$

The ANN's training objective is to reduce the loss function $L(W, b)$ to the minimum. It is defined to measure the difference between the output result and the expected result through the network structure. Common loss functions is composed of mean square error (MSE) and negative log likelihood (NIL):

$$MSE(W,b) = \frac{1}{|Y|}\sum_{i=1}^{|Y|}(Y(i) - \hat{Y}(i))^2\ . \tag{10}$$

$$NIL(W,b) = \sum_{i=1}^{|Y|}\log Y(i)\ . \tag{11}$$

In the process of model training, i.e. parameter training, gradient descent algorithm is mainly adopted. Back propagation is carried out through gradient descent. Each trainable parameter of the convolution layer is updated level by level.

$$W_i = W_i - \eta\frac{\partial E(W,b)}{\partial W_i}\ . \tag{12}$$

$$b_i = b_i - \eta\frac{\partial E(W,b)}{\partial b_i}\ . \tag{13}$$

In the formula (12) and (13), η indicates the learning rate, that is, the gradient descent step, and is adopted to represent the strength of back propagation.

There will be a significant amount of duplicate features produced throughout the ANN training process. The model includes an attention mechanism to concentrate on significant characteristics. CBAM can improve the quality of features by assigning different weights without increasing computational cost. CBAM consists of channel attention mechanism and spatial attention mechanism. Fig. 3 shows its structure.

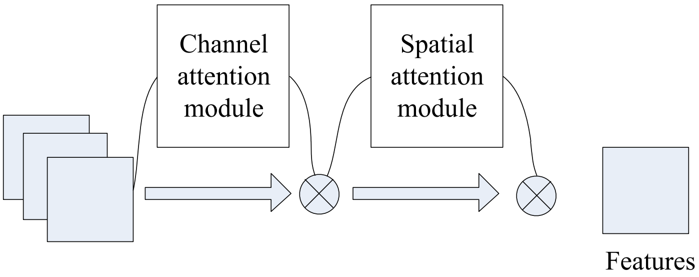


Fig. 3. CBAM structure

Fig. 4 shows the operation flow of channel attention mechanism.

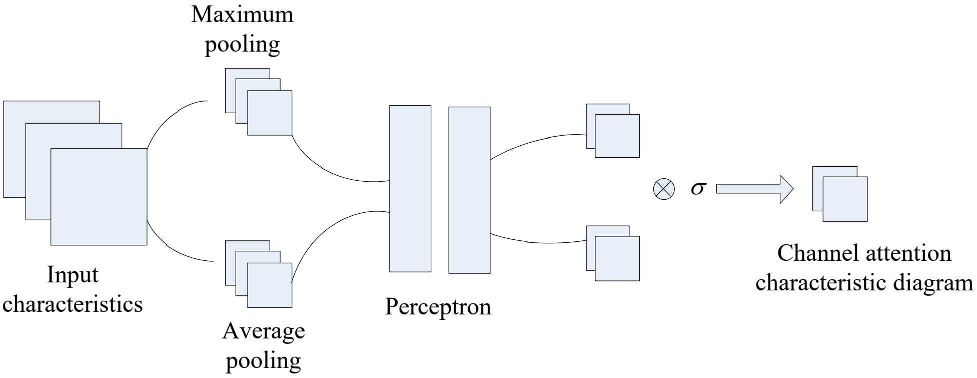


Fig. 4. Channel attention module

The feature map is first maximum and average pooled in Fig. 4, and two vectors of the same dimension are then produced. These two vectors are respectively input into the same multi-layer perceptron for learning. The output results are added one by one and then sent to the sigmoid function to activate the attention vector.

The weight vector calculation formula is:

$$\begin{aligned} M_c F &= \sigma MLPAvgPoolF + MLPMaxPoolF \\ &= \sigma W_1 W_0 F^c_{avg} + W_1 W_0 F^c_{\max} \end{aligned} . \quad \textbf{(14)}$$

Multiply the weight vector shown in formula (14) by the channel of the feature map to obtain the feature map to enhance attention, as shown in formula (15):

$$F^{'} = M_c F \otimes F . \quad \textbf{(15)}$$

In the formula (15), F represents the input characteristic graph, σ represents the activation function sigmoid, F^c_{avg} and $F^c_{\max}$ represent the characteristic graph after the average global pooling and maximum global pooling respectively, W_1 and W_0 represent the weight of the full connection layer, *MLP* represents the perceptron, $F^{'}$ represents the characteristics of the channel attention, and the obtained M_cF is the channel weight vector.

The precise working of the spatial attention module is shown in Fig. 5.

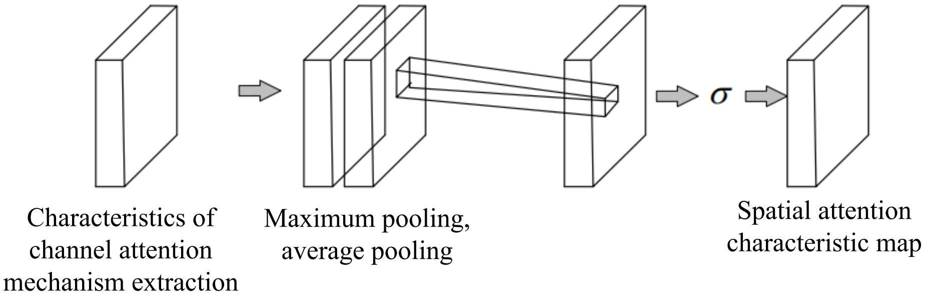

Characteristics of channel attention mechanism extraction
Maximum pooling, average pooling
σ
Spatial attention characteristic map

Fig. 5. Spatial attention module

The outcome of the channel attention mechanism's max-pooling and average-pooling operations are shown in Fig. 5. To produce the weight matrix of spatial attention, the outcomes of the two pooling procedures are stitched together, and the stitched feature maps are convolved. After propagating along the channel direction, the feature map is multiplied element by element to obtain a feature map with enhanced spatial attention, as shown in formula (16):

$$F^{''} = M_s F^{'} \otimes F^{'} . \quad \textbf{(16)}$$

In the formula (16), $F^{''}$ represents the feature map refined by spatial attention, and $M_sF^{'}$ represents the spatial weight vector.

ELM has been widely used in image classification. In the training process, few parameters need to be set, high efficiency, and good generalization performance. The output result is:

$$f(x) = h(x)H^T(1/C + HH^T)^{-1}L . \quad \textbf{(17)}$$

In the formula (17), $f(x)$ indicates the objective function, C indicates the regularization coefficient, x indicates the input vector, $h(x)$ indicates the output of the hidden layer, and L indicates the expected output.

After introducing the radial basis kernel function into the limit learning machine, formula (17) can be transformed into:

$$f(x) = [K(x,x_1);....;K(x,x_n)]/C + \Omega_{ELM}{}^{-1}L . \quad \textbf{(18)}$$

$$\Omega_{ELM} = HH^T = h(x_i)h(x_j) = K(x_i,x_j) . \quad \textbf{(19)}$$

$K(x_i, x_j)$ is the radial basis kernel function, which is expressed as:

$$K(x_i,x_j) = \exp\left(-\frac{\left\|x_i - x_j\right\|}{\gamma^2}\right). \quad \textbf{(20)}$$

In the formula (20), the kernel parameters γ reflect the distribution characteristics of the data samples.

Sparrow search algorithm is an optimization algorithm that transforms the foraging and anti-predation behavior rules of sparrows in nature. The sparrow group has two roles : discoverer and participant. Finders are adaptable and search a wide range, leading sparrow populations in search of food. Participants follow the finder for food. At the same time, in order to increase their own predation rate, participants will monitor the competition for food around the finder. The function of the early warning mechanism in the sparrow search algorithm is that when the population is threatened by predators, it will immediately conduct anti-predation behavior. The finder's update formula is:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t} \cdot \exp\left(-\dfrac{i}{a \cdot T}\right), R_2 < S \\ X_{i,j}^{t} + Q \cdot L, R_2 \geq S \end{cases} . \tag{21}$$

In the formula (21), $X_{i,j}^{t}$ indicates the information of the i -th sparrow of generation t in the j dimension, T indicates the maximum number of iterations, R_2 indicates the warning line, S indicates the safety value. In case of warning, the sparrow approaches the safety direction, a is the random value within the range of [0, 1], Q is the random value, which satisfies the normal distribution, and L is a matrix of all 1. The update formula of the participant is:

$$X_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\dfrac{X_{worst} - X_{i,j}^{t}}{t^2}\right), i > n/2 \\ X_P^{t+1} + \left|X_{i,j}^{t} - X_P^{t+1}\right| \cdot A^{+} \cdot L, i \leq n/2 \end{cases} . \tag{22}$$

In the formula (22), X_P indicates the optimal position, A indicates the matrix of all -1 or -1, $A^{+} = A^{T} \cdot AA^{T-1}$, and X_{worst} indicates the worst position.

In a sparrow population, 20% of the individuals can sound the alarm. When facing danger, its position update formula is as follows:

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^{t} + \beta \left|X_{i,j}^{t} - X_{best}^{t}\right|, f_i > f_g \\ X_P^{t+1} + K \left[\dfrac{\left|X_{i,j}^{t} - X_{worst}^{t}\right|}{f_i - f_w + \varepsilon}\right], f_i = f_g \end{cases} . \tag{23}$$

In the formula (23), β represents a random number, which is adopted to control the step length, and K is -1 or 1, which is adopted to control the direction. f_i represents fitness, f_g and f_w are the best fitness and the worst fitness, and ε is a constant.

Through the above calculations, the detection of moving objects in dynamic images can be completed.

4 Experimental Verification

4.1 Experimental Data

To verify the put forward algorithm of MTD in dynamic image sequences on the basis of ANN, comparative experiments are carried out. All experiments in this paper are performed on 15 complex scene videos in the public datasets I2R and CDnet2014, respectively: AirportHall, Curtain, Bootstrap, Boats, Fountain, Campus, Lobby, ShoppingMall, WaterSurface, Escalator, Fall, Canoe, Fountain01, Founatin02 and Overpass. The I2R dataset provides 20 frames per video as Groundtruth, and the CDnet2014 dataset provides Groundtruth for each frame in a video sequence to evaluate performance. The videos contain a variety of complex scenes, such as various kinds of moving targets (campus, boats, fall), crowded people (airporthall, bootstrap), sudden lighting (lobby), shadows (airporthall, bootstrap, shoppingmall), and dynamic backgrounds (curtain, watersurface, escalator, campus, fountain, boats, fall, canoe, fountain01, fountain02, override).

From the above data sets, dynamic images are selected for MTD. Some experimental sample images are shown in Fig. 6.

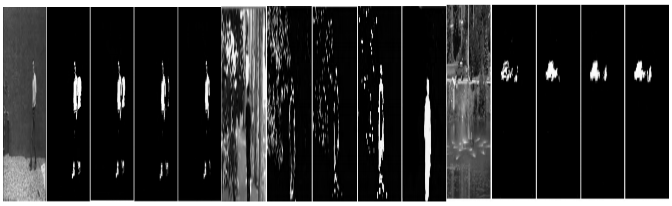


Fig. 6. Image of experimental sample

4.2 ANN Training

The training of the ANN uses 600 groups of data. For different dynamic images, the statistical characteristics of the edge trend of the moving target and its corresponding binary number combination are used as the network input and the network output. The parameters of ANN are shown in Table 1.

Table 1. Parameters of ANN

Network structure and activation function	Parameter
Input layer	16 nodes
Hidden layer	6 nodes
Output layer	2 nodes
Hidden layer activation function	Sigmoid
Output layer activation function	Purelin
Training function	Trainrp

Fig. 7 displays the MSE network alterations that occurred throughout training. Fig. 7 demonstrates that the network is constrained and the MSE curve begins to decrease very rapidly. After a small number of iterations, the MSE reaches the ideal level. After that, the trained model is used to test 542 sample images, including 224 positive samples and 318 negative samples. The model correctly identified 520 samples and misidentified 22 samples. Therefore, the recognition accuracy of the model is 95.94%. According to the above recognition accuracy and mean square error curves, it can be concluded that when the convolution kernel size is 9*9 and the network depth is 7, the recognition efficiency of ANN is ideal, and the accuracy rate on the test set composed of 542 samples is 95.94%.

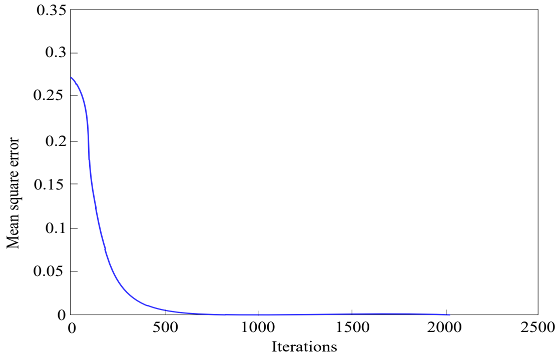
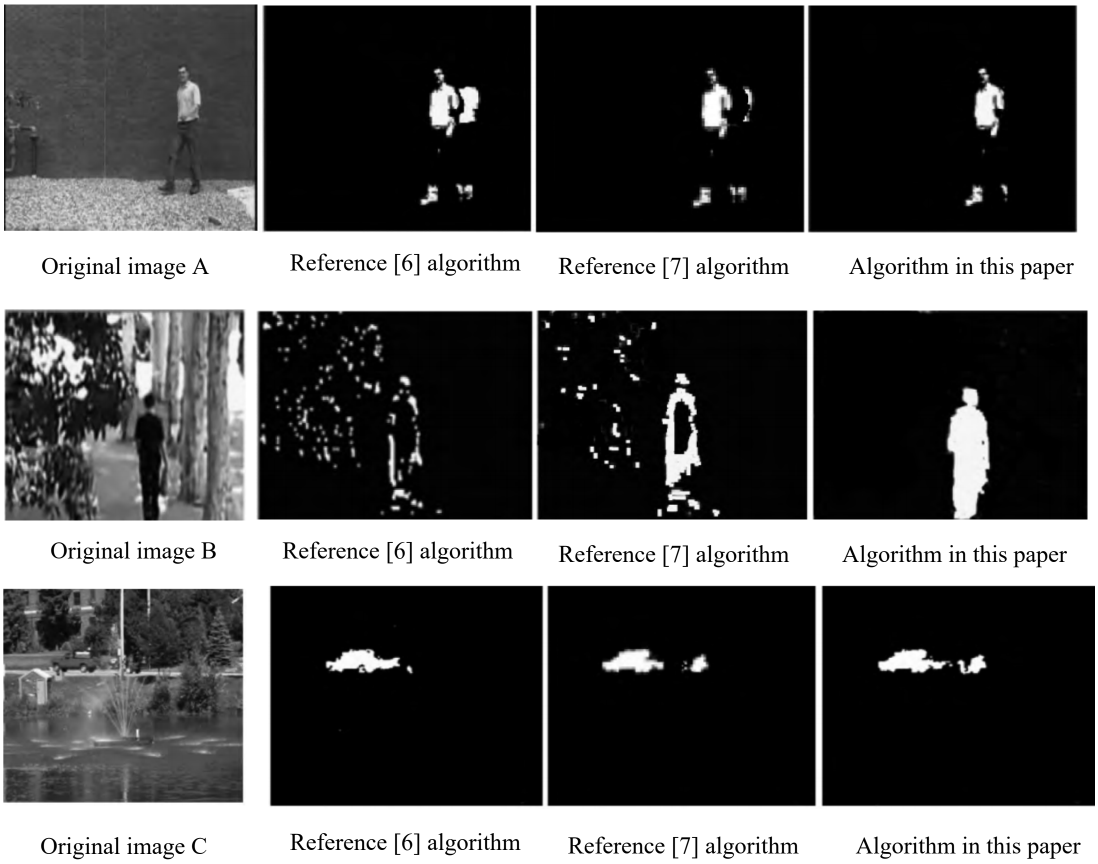


Fig. 7. MSE curve trained when the number of layers is 7 and the convolution kernel size is 9*9

4.3 Experimental Results Analysis

Effectiveness of MTD. The methods in references [6] and [7] were contrasted with the approach suggested in this study using the three photos shown in Fig. 6 as the target images for MTD validity evaluation. Fig. 8 displays the outcomes of the three techniques for dynamic target identification. When there is background interference, there is a certain gap between the detection results of the MTD method proposed in reference [6] and reference [7] and the moving target, while the MTD result of the algorithm in this paper is closer to the moving target. For example, in the original image B, the moving objects detected by the algorithms in reference [6] and reference [7] contain more leaves, which leads to the interference of the detection results of moving human objects, so the detection of human motion contours to become blurred. However, our algorithm can clearly detect moving humans and objects.



Original image A Reference [6] algorithm Reference [7] algorithm Algorithm in this paper

Original image B Reference [6] algorithm Reference [7] algorithm Algorithm in this paper

Original image C Reference [6] algorithm Reference [7] algorithm Algorithm in this paper

Fig. 8. MTD results

Algorithm Performance Evaluation. To more accurately evaluate and comprehensively verify the practicability and reliability of the algorithm in this paper, the experiment uses precision rate, recall rate Recall, comprehensive evaluation index F_measure(F), false positive rate (FPR), false positive rate (FPR)),, False Negative Rate (FNR) is used as an evaluation index of algorithm performance. The calculation formula of each index is:

$$P = \frac{T_P}{(T_P + F_P)} . \tag{24}$$

$$R = \frac{T_P}{(T_P + F_N)} \,. \tag{25}$$

$$F = \frac{2 \times P \times R}{(P + R)} \,. \tag{26}$$

$$FPR = \frac{F_P}{(F_P + T_N)} \,. \tag{27}$$

$$FNR = \frac{F_N}{(T_N + F_P)} \,. \tag{28}$$

In the formula(24)-(28), T_P represents the number of foreground pixels correctly detected, T_N represents the number of background pixels correctly detected, F_N represents the number of background pixels incorrectly detected, and F_P represents the number of foreground pixels incorrectly detected. The algorithm put forward in this paper is compared with the algorithms in references [6] and [7]. Table 2 shows the performance verification results of the three algorithms. Observing the performance test results of the algorithm in Table 2, compared with the two algorithms, the performance evaluation indicators of the algorithm in this paper, such as accuracy rate and recall rate, have significant advantages. Compared with the algorithms proposed in reference [6] and reference [7], the comprehensive index is F higher by 0.0630 and 0.0520, respectively. Therefore, the target detection algorithm proposed in this paper has high practicability and reliability.

Table 2. Algorithm performance test results

Test indicators	Test result		
	Algorithm in this paper	Reference [6] algorithm	Reference [7] algorithm
P	0.8134	0.7368	0.7462
R	0.7145	0.6626	0.6743
F	0.7607	0.6977	0.7084
FPR	0.0062	0.0172	0.0089
FNR	0.0102	0.0175	0.0159

Operational Efficiency of the Algorithm. In order to verify the operating efficiency of the algorithm in this paper and the real-time detection function of moving objects, the experiment combined references [6] and [7] to conduct multiple tests on four groups of dynamic images, and obtained the three target detection algorithms required for processing dynamic images. time. T_1, T_2 respectively T_3 represent the time consumption of MTD in the three scenarios. The running efficiency evaluation results of the three algorithms are shown in Table 3. Compared with the algorithms in reference [6] and reference [7], the average running efficiency of the algorithm in this paper is reduced by 0.28788ms and 0.26945ms respectively. It shows that the algorithm in this paper consumes the least time among the three popular algorithms, and has the highest efficiency in processing dynamic images, which reflects the superior performance of the algorithm in this paper. At the same time, it is also verified that the algorithm can meet the real-time requirements of MTD.

Table 3. Evaluation results of algorithm operation efficiency

Test scenario	Test result/ms		
	Algorithm in this paper	Reference [6] algorithm	Reference [7] algorithm
T_1	2.58435	3.59258	2.53285
T_2	3.35471	3.28567	3.77114
T_3	4.37182	4.29627	4.81524
Mean value	3.43696	3.72484	3.70641

5 Conclusion

The advancement of the times and technology has enabled the development of computer vision image information. In order to improve the accuracy of moving object detection in complex environments, an ANN-based MTD algorithm is proposed in experiments. The algorithm effectively processes the dynamic image and completes the feature extraction, and uses the ANN algorithm to realize the detection of the moving target by the model. Finally, simulation experiments were carried out on 15 data sets, which verified the performance of the algorithm both theoretically and experimentally. The experimental selection is compared with the methods shown in reference [6] and reference [7]. The results show that the detection accuracy of the model in this paper is 95.94% on a test set composed of 542 samples. Compared with the detection algorithm based on improved background subtraction, the MTD results of the method in this paper are basically consistent with the actual moving target, and the error is small; compared with the detection algorithm based on GMM and superpixel Markov randomness, the calculation efficiency of the algorithm in this paper is significantly higher Advantage, the average detection time is 3.69515ms. Therefore, the experimental results fully demonstrate that the constructed ANN-based detection method can better meet the requirements of moving object detection in dynamic image sequences. This experiment has basically achieved the purpose of the experiment, but there are few empirical studies on the detection of moving objects in dynamic images, and future research can start from this.

References

[1] H. Mizushina, I. Kanayama, Y. Masuda, S. Suyama, Importance of visual information at change in motion direction on depth perception from monocular motion parallax, IEEE Transactions on Industry Applications 56(5)(2020) 5637-5644.

[2] M. Krestenitis, N. Passalis, A. Iosifidis, M. Gabbouj, A. Tefas, Recurrent bag-of-features for visual information analysis, Pattern Recognition 106(2020) 107380.

[3] A.-D. Cesarei, S. Cavicchi, G. Cristadoro, M. Lippi, Do humans and deep convolutional neural networks use visual information similarly for the categorization of natural scenes? Cognitive Science 45(6)(2021) e13009.

[4] B.-A. Alpatov, P.-V. Babayan, M.-D. Ershov, Approaches to moving object detection and parameter estimation in a video sequence for the transport analysis system, Computer Optics 44(5)(2020) 746-756.

[5] Z. Hu, Y. Wang, R. Su, X. Bian, H. Wei, G. He, Moving object detection Based on non-convex RPCA with segmentation constraint, IEEE Access 8(2020) 41026-41036.

[6] P.-G. Asteris, V.-G. Mokos, Concrete compressive strength using artificial neural networks, Neural Computing and Applications 32(15)(2020) 11807-11826.

[7] S. Rostami, D. Toghraie, B. Shabani, N. Sina, P. Bainoon, Measurement of the thermal conductivity of MWCNT-CuO/water hybrid nanofluid using artificial neural networks (ANNs), Journal of Thermal Analysis and Calorimetry 143(2)(2021) 1097-1105.

[8] E.-H. Houssein, M. Dirar, K. Hussain, W. Mohamed, Assess deep learning models for Egyptian exchange prediction using nonlinear artificial neural networks, Neural Computing and Applications 33(11)(2021) 5965-5987.

[9] H. Nguyen, C. Drebenstedt, X.-N. Bui, D. Bui, Prediction of blast-induced ground vibration in an open-pit mine by a novel hybrid model based on clustering and artificial neural network, Natural Resources Research 29(2)(2020) 691-709.

[10] L.-J. Stamenković, S.M. Kurilić, V.P. Ulniković, Prediction of nitrate concentration in Danube River water by using artificial neural networks, Water Supply 20(6)(2020) 2119-2132.

[11] M. Elhoseny, Multi-object detection and tracking (MODT) machine learning model for real-time video surveillance systems, Circuits, Systems, and Signal Processing 39(2)(2020) 611-630.

[12] W. Sun, L. Dai, X. Zhang, P. Chang, X. He, RSOD: Real-time small object detection algorithm in UAV-based traffic monitoring, Applied Intelligence 52(8)(2022) 8448-8463.

[13] S. Jha, C. Seo, E. Yang, G. Joshi, Real time object detection and tracking system for video surveillance system, Multimedia Tools and Applications 80(3)(2021) 3981-3996.

[14] C. Dai, X. Liu, W. Chen, C. Lai, A low-latency object detection algorithm for the edge devices of IoV systems, IEEE Transactions on Vehicular Technology 69(10)(2020) 11169-11178.

[15] S. Ammar, T. Bouwmans, N. Zaghden, M. Neji, Deep detector classifier (DeepDC) for moving objects segmentation and classification in video surveillance, IET Image Processing 14(8)(2020) 1490-1501.

[16] O. Kaziha, A. Jarndal, T. Bonny, Genetic algorithm augmented convolutional neural network for image recognition applications, in: Proc. 2020 International Conference on Communications, Computing, Cybersecurity, and Informatics, 2020.

[17] E. Lashgari, J. Ott, A. Connelly, P. Baldi, U. Maoz, An end-to-end CNN with attentional mechanism applied to raw EEG in a BCI classification task, Journal of Neural Engineering 18(4)(2021) 0460e3.

[18] P. Ghadekar, R. Kale, N. Agrawal, A. Pophale, S. Rudrawar, Low light and over exposed image enhancement using weight matrix technique, International Journal of Computer Applications 175(25)(2020) 22-26.

[19] J. Zhang, Q. Wang, Y. Yuan, Metric learning by simultaneously learning linear transformation matrix and weight matrix for person re-identification, IET Computer Vision 13(4)(2019) 428-434.

[20] H.-B. Wang, T. Liu, S.-J. Liu, W.-Y. Wei, X.-R. Liu, P.-B. Liu, Y.-Y. Bai, Y.-A. Chen, Implicit progressive-iterative algorithm of curves and surfaces with compactly supported radial basis functions, Journal of Computer-Aided Design & Computer Graphics 33(11)(2021) 1755-1764.