Target Detection for Motion Images Using the Improved YOLO Algorithm

Tian Zhang, College of Computer Science and Technology, ChangChun Normal University, ChangChun, China*

ABSTRACT

The images of motion states are time-varying, and when actually detecting their internal motion targets, the formed detection frames overlap, resulting in small confidence values for the detection frames and low accuracy of the detection results. To address this problem, the authors propose a target detection for motion image using the improved YOLO algorithm. First, the YOLO algorithm is improved using deformable convolution; the edge weights of the front and back views within the image are collated, and the motion image is segmented using the improved YOLO algorithm. Second, the structure formed by the initial convolution is used as the initial detection frame structure, the parallel cross-ratio value is set, the overlap generated by the detection frame is controlled, the parameters of the detection frame compression processing are output, the threshold trigger value relationship is constructed, and finally, the detection of the motion image target is realized. The results show that the target false detection rate of the proposed method is only about 15%. The detection a priori frame height value is 80 pixels, and the average detection time consumed is 6.8ms, which proves that the proposed algorithm can be widely used in motion image target detection to improve the detection level.

KEYWORDS

Improved YOLO Algorithm, Motion Image, Parallel Cross-Ratio Value, Target Detection

INTRODUCTION

Target detection is a very popular research topic in the field of computer vision. It mainly deals with the matching and segmentation of objects of interest in motion video images, and finally gets the target information and motion trajectory (Zaghari et al., 2021). Motion image target detection can provide a lot of useful data support for computer analysis. Therefore, it is widely used in intelligent transportation, video surveillance, military, and other fields.

After the application of a self-convolutional neural network to motion image target detection, the detection idea of researchers was broadened for target detection in video motion images. The positioning information of the detected image target is processed as a linear regression problem, and a detection process of numerical processing is constructed (Sri & Esther, 2020). As a part of computer
vision processing, a standardized processing flow has been formed in target detection. Taking the extracted feature point attributes as the basis of constructing the sliding window can improve the performance of the existing image detection methods (Zibang, 2019). However, due to the poor performance of traditional camera equipment and the interference of light, occlusion, and other external factors in the camera process, it is more difficult to detect high-precision targets. Therefore, a more efficient and accurate moving image target detection method is needed.

To address this problem, Kou et al. (2019) proposed a target detection method based on multi-scale uniform features in multispectral images. The local features of infrared point targets are extracted from the uniformity of gray difference distribution, and the target detection results are output through image fusion detection. The detection efficiency of this method is good. Therefore, the detection accuracy needs to be improved. Deng et al. (2018) proposed a motion point target detection method to detect targets from infrared sequence images. An improved spatiotemporal all-variable model is designed for background prediction, and the predicted background is subtracted from the corresponding sequence of images to obtain the segmented image. Finally, the target is detected according to the product of the segmented image and the time comparison filter. The detection accuracy of this method is good. However, the detection efficiency needs to be improved.

Based on the existing related literature, in this paper, the improved YOLO algorithm will be applied to motion image target detection. Using deformable convolution network processing to improve YOLO algorithm, motion image target detection is finally realized through image segmentation and detection frame construction. Experimental results show that the proposed algorithm has better detection performance. The main contributions of this paper are as follows: (1) Using deformable convolution network to improve YOLO algorithm, the improved YOLO algorithm can better adapt to objects with different scales or deformation degrees, and improve the detection effect. (2) The detection frame of the motion image is constructed, the detection range of the detection frame is controlled, and the motion image objects are classified. (3) The image pixels in the detection frame are compressed into cell structure, the gradient and amplitude parameters of the image are calculated, and the target detection results are determined according to the sliding triggered numerical interval. (4) The results of different data sets show that the proposed motion image target detection algorithm based on the improved YOLO algorithm can detect motion image targets accurately and efficiently.

RELATED WORKS

Relevant scholars at home and abroad have researched motion image target detection technology. With feature matching as the technical support of target detection, dynamic images with various sizes and running speeds are simulated, and the numerical tracking process is constructed to obtain the motion target detection process. Huan et al. (2020) proposed a Synthetic aperture radar (SAR) multi-target interactive motion recognition method based on a convolutional neural network. This method carries out wavelet threshold de-noising on SAR target images, uses a convolutional neural network to realize target type recognition, constructs a motion feature matrix and motion simulation dataset, and realizes motion target detection. This method has high recognition accuracy for multi-target interactive motion, but the efficiency of target recognition is poor. Xie et al. (2022) proposed a spatiotemporal feature detection method for small and low contrast targets based on a data-driven support vector machine. In addition, a new pixel-level feature, the spatiotemporal contour, is designed to describe the discontinuity of each pixel in the space and time domains. The labeled spatiotemporal contour is used to train the SVM classifier to automatically learn the spatiotemporal feature fusion mechanism, and the final target detection result is generated through a parallel convolution operation. This method has high detection accuracy for small low-contrast targets but poor detection effect for large high-contrast targets. Chen et al. (2022) proposed a general target detection learning scheme, and this paper proposed “motion quality”, which is a new concept. It can mine video frames from “buffered” test video streams to build a fine-tuning set. By using this method, all frames in this set can be detected by “target SOTA model”.
However, there is still a lack of detection speed; Chen et al. (2021) proposed a new detection method. In the process of similar pixel selection using context covariance matrix similarity test, a similar pixel number index that can effectively identify salient targets was derived, and then a detection method based on similar pixel numbers was established to obtain more accurate detection results. However, the analysis of the detection frame is insufficient, and the reliability needs further analysis and verification; Qin et al. (2022) proposed a target tracking method based on interference detection considering similar interference, local occlusion, and scale change in the target tracking process, which uses the edge distribution of the feature map to determine whether there is interference. When there is interference in the scene, the motion vector composed of the predicted values obtained by the Kalman filter is used as the basis for target prediction. This method has achieved good anti-interference effect. However, the detection takes a long time. Cui et al. (2020) used the three-frame difference method to mark the blurred motion region of the motion image and reconstruct the super-resolution of the blurred region. After the preselected box parameters are determined, the detection box generated by target detection is allocated to realize target detection. The detection accuracy of this method is good, but the detection efficiency is poor.

Through the application test, it is found that the above-designed image target detection method has the disadvantages of low reliability of detection frames and poor accuracy of target detection results. Therefore, a motion image target detection algorithm using improved YOLO algorithm is proposed. The results show that the target false detection rate of the proposed algorithm is about 15%, which is relatively low compared with other algorithms, and the detection accuracy is improved. Moreover, the proposed algorithm has fast detection speed, short time consumption, high accuracy, and the average reliability of detection frame is up to 94%, effectively improving the detection level.

**METHODOLOGY**

**Design of Motion Image Target Detection Framework**

The detection of image objects in video motion is an important research content in the field of computer vision. YOLO algorithm is an end-to-end deep neural network model and a relatively mature image object detection tool at present. To effectively improve the detection effect, YOLO algorithm is improved in this paper. Motion image target detection is studied based on the improved YOLO algorithm, and the following frame structure is constructed, as shown in Figure 1.

As can be seen from Figure 1, after the motion image is input, this paper first uses the modified YOLO algorithm of deformable convolution to divide the image grid and segment the motion image
target. Then the motion detection frame is established for the segmented image structure, and the attribute category of the motion image is determined through analysis to complete the target detection.

**Introducing Deformable Convolution to Improve YOLO Algorithm**

All the convolution layers of YOLO algorithm use convolution networks to complete the operation. In motion image target detection, objects with different scales or deformation degrees are often encountered. Because the convolution network can only sample the fixed position of objects according to the feature information, the convolution network has modeling defects in this case. To improve the detection performance of YOLO algorithm, this paper introduces deformable convolution to improve the algorithm, so that the convolution kernel can shift on the position of image feature points, to achieve random sampling of the region of interest of the motion object image and avoid the limitation of sampling.

When YOLO algorithm is running, first divide the input moving image into $S \times S$ grids. If the center of the target to be detected is in a grid, the grid can process the target to be detected. After introducing deformable convolution to improve YOLO algorithm, an offset $\Delta O_n$ is added to the regular grid. The corresponding weight value will also change to $\Delta \omega_n$, and the output value at the same position is

$$Y(O) = \sum \beta(O_n) \cdot X(O_0 + O_n + \Delta O_n) \cdot \Delta \omega_n$$  \hspace{0.5cm} (1)

where $O_0$ represents the initial position of the grid, $X$ is the input value of moving image target, $\beta$ is the convolutional kernel.

**Motion Image Segmentation**

The Gibbs energy of the motion image is determined by sorting out the edge values output by the motion image (Jiang et al., 2021) and according to the edge weights of the front and back scenes of the image. The numerical relationship can be expressed as:

$$E(z) = U(k) + r(a)$$  \hspace{0.5cm} (2)

where $E(z)$ refers to the Gibbs energy of the motion image, $U(k)$ refers to the probability of smooth separation, and $r(a)$ refers to the edge weight function of the foreground and background of the motion image (Cao et al., 2020; Rong et al., 2020). The operation time of the improved image is determined according to the energy of the motion mask in the specified motion frame. The processing process can be expressed as

$$\min_k E(z) = (\theta_i, \alpha_j)$$  \hspace{0.5cm} (3)

$$\alpha_i = \frac{K}{\theta}$$  \hspace{0.5cm} (4)

$$\theta_i = k_i \times n$$  \hspace{0.5cm} (5)

In the above numerical relationship, $\alpha_i$ refers to the initialization function of the fuzzy mask, $K$ refers to the interaction coefficient of image motion, $\theta$ refers to the updated function of the image, $k_i$ refers to the convergence coefficient, and $n$ refers to the initialization times of the algorithm. The optimized motion image parameters are sorted out, and the segmentation probability of the image structure corresponding
to the parameters is determined by a Gaussian processing process according to the size of the parameters (Moskopp et al., 2019; Zheng & Yao, 2019). The numerical relationship can be expressed as

$$p(f) = \frac{\sum_{n=1}^{K} P(w_n)}{C_n}$$  \hspace{1cm} (6)$$

where $p(f)$ refers to the segmentation probability of the obtained image structure, $P(w_n)$ refers to the Literature Gaussian run function, and $C_n$ refers to the number of attribute probabilities. According to the processing process from large to small output probability (Br et al., 2021), the selected processed motion image is segmented, and the image motion detection frame is constructed by attribute parameters (Wang et al., 2020).

**Construction of Motion Image Detection Frame**

Based on motion image segmentation, to better complete target detection, set the detection range, build an image motion detection box, further control the detection range, and lay a foundation for improving the detection accuracy.

Image structure obtained by improved YOLO algorithm. The default initial convolution kernel structure is the detection frame of the motion image (Ognard et al., 2019), and the anchor frame with an aspect ratio of 2:1 is set to process the pixels of the segmented image (Sandfort et al., 2021). At this time, a pixel candidate region is generated in the convolution regression layer.

$$G(x, y) = \frac{m^2(x^3 + y^3)}{k - \sigma}$$  \hspace{1cm} (7)$$

where $G(x, y)$ refers to the determined pixel candidate region, $(x, y)$ refers to pixel coordinates of the image, $m$ refers to the scale parameter of convolution, $\sigma$ refers to the contour parameter of the detected image, and other parameters remain unchanged. To control the position accuracy of the detection frame (Ma & Li, 2021), the center of the detection frame is determined according to the pixel value. The numerical processing process can be expressed as

$$\Gamma(x) = 1 - \min_{y \in k} \left( \frac{I(x)}{A_y} \right)$$  \hspace{1cm} (8)$$

where $\Gamma(x)$ refers to the constructed detection function, $I(x)$ refers to the scale function of the image pixel point, $A_y$ refers to the gas light component, and the other parameters remain unchanged. During the frame switching process of the motion image, the detection frame is not switched in time, resulting in the overlap of the detection frame of the next frame (Yu et al., 2020). To control the overlap of this part, the values at the boundary of the detection frame are set and compared. The numerical relationship can be expressed as

$$I_o = \frac{a(c) \cap a(G)}{a(c) \cup a(G)}$$  \hspace{1cm} (9)$$
where \( I_e \) refers to the set parallel cross-ratio parameters, \( a(c) \) refers to the predicted target frame area, and \( a(G) \) refers to the area generated by the real detection frame. The parallel cross-ratio value is used as the trigger threshold of the detection frame to control the detection range of the detection frame (Frei & Kruis, 2020). Within the determined detection frame area, formulate the motion target detection process and finally realize the target detection function.

**Realization of Target Detection**

The image pixels in the detection frame are compressed into a cell structure, and the gradient and amplitude parameters of the structure are determined

\[
D(x,y) = \sqrt{I_x(x,y)^2 + I_y(x,y)^2} 
\]

\[
F(x,y) = \arctan\left(\frac{I_x(x,y)}{I_y(x,y)}\right) 
\]

where \( D(x,y) \) refers to the gradient function of the cell structure, \( I_x(x,y) \) and \( I_y(x,y) \) refer to the gradient function of the pixel point at axis \( x \) and axis \( y \), and \( F(x,y) \) refers to the amplitude of the cell structure. The above output parameters are sorted, and the cell structure of the known attribute parameters is equally divided into 9 interval feature blocks (Eppenhof et al., 2020). Set the value range of the sliding trigger. The value relationship can be expressed as

\[
col = \begin{cases} 
W - S, & S > 0 \\
W + S, & S \leq 0 
\end{cases}
\]

where \( col \) is the set sliding trigger value relationship, \( W \) refers to the number of sliding windows, and \( S \) refers to the step size of the sliding window. According to the numerical interval constructed above, the target detection results are finally output.

According to the above processing target detection process, calibrate the target information specification of the motion image corresponding to the attribute categories of different motion images, and finally output the target detection result according to the above processing process (Jayme & Guilherme, 2019). Based on the above analysis process, the design of motion image target detection technology based on the improved YOLO algorithm is finally completed.

The process of improving YOLO algorithm to achieve motion object image detection can be described as follows:

**Input:** The edge value of the motion image output is sorted, and the Gibbs energy of the motion image is determined according to the edge weight of the front and back scenes of the image.

**Output:** detection result of the motion image target.

1. Using deformable convolution to improve YOLO algorithm;
2. Sort out the edge value of the motion image output, take its edge weight, and determine the Gibbs energy of the motion image \( E(z) \) to complete image segmentation;
3. In the framework of the improved YOLO algorithm, use the mask structure to optimize the coefficients in motion images, and the image motion detection frame is constructed by using attribute parameters;
4. In the detection box, determine the gradient and amplitude parameters \( D(x,y) \) and \( F(x,y) \) of the structure;
5. Set the value range of the sliding trigger, determine the attribute category of motion image target detection, and output the target detection result.
6. End

The processing process of motion image target detection technology based on improved YOLO algorithm is shown in Figure 2.

**Figure 2.**
Processing of motion image target detection technology based on improved YOLO algorithm
EXPERIMENTAL ANALYSIS AND RESULTS

Data Sets

Pascal VOC data set: It contains tasks such as object classification, target detection, and image segmentation. A specific object is recognized from it. It contains 20 classes of objects to be recognized, with a total of 11,530 images containing 27,450 ROI annotated objects and 6,929 segments. BDD100K data set: The BDD100K dataset contains 100,000 HD videos, and each video is approximately 40 seconds, 720p, and 30 fps. The keyframes are sampled at the 10th second of each video, and 100,000 images (image size: 1280x720) are obtained and annotated. Set the image acquisition command, select five groups of videos with type attributes in the above two datasets, intercept 100 frames of motion images as test objects, and organize the basic parameters of motion image data according to the object attributes of visual tracking, as shown in Table 1.

After collation of the collected motion image data attributes, Label visual annotation is used to extract the commands for the image compilation. According to the data format generated by the data mark, the text type configuration file is used to convert the script format of the picture, the script format of the text attribute is more displayed in the configuration file in the upper computer, and the running network state is set to fine tune. According to the processing process of layer-by-layer analysis, the network structure corresponding to the state is maintained, and the tensor of input and output is adjusted. According to the directory structure of the YOLOv3 source code, convert the configuration file carrying the image, call the cfg file in the network structure, and replace the anchor frame file with the YOLOv3 cfg type file. After the split code configuration is output, YOLOv3.0 of the script includes a cfg file to generate labels that can be used for direct detection. After taking the RELU function as the default function processing value, the learning rate of the function algorithm was set to 0.001. After the loss value output after iteration tends to be stable, the above processed image dataset is taken as the processing object of target detection. Taking the algorithm in SARMIM (Huan et al., 2020), the algorithm in SLCTD (Xie et al., 2022), the algorithm in NVSODM (Chen et al., 2022), the algorithm in SFSISD (Chen et al., 2021) and the algorithm in TTMID (Qin et al., 2022) as experimental comparison algorithms, the performance indexes shown in the target detection process are selected to compare the performance of different detection algorithms.

<table>
<thead>
<tr>
<th>Dataset Type</th>
<th>Dataset Serial Number</th>
<th>Resolution</th>
<th>Acquisition Speed/fps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pascal VOC data set</td>
<td>SP-01</td>
<td>1280x720</td>
<td>13.52</td>
</tr>
<tr>
<td></td>
<td>SP-02</td>
<td>1280x720</td>
<td>41.35</td>
</tr>
<tr>
<td></td>
<td>SP-03</td>
<td>640x360</td>
<td>24.32</td>
</tr>
<tr>
<td></td>
<td>SP-04</td>
<td>640x360</td>
<td>9.82</td>
</tr>
<tr>
<td></td>
<td>SP-05</td>
<td>1280x720</td>
<td>13.44</td>
</tr>
<tr>
<td>BDD100K data set</td>
<td>SP-06</td>
<td>640x360</td>
<td>16.42</td>
</tr>
<tr>
<td></td>
<td>SP-07</td>
<td>1280x720</td>
<td>17.63</td>
</tr>
<tr>
<td></td>
<td>SP-08</td>
<td>640x360</td>
<td>8.64</td>
</tr>
<tr>
<td></td>
<td>SP-09</td>
<td>1280x720</td>
<td>7.64</td>
</tr>
<tr>
<td></td>
<td>SP-10</td>
<td>640x360</td>
<td>14.52</td>
</tr>
</tbody>
</table>
Evaluation Criteria

Target false detection rate: the motion amplitude of the target to be detected is small. According to the numerical relationship of the recall rate of the image output, the scale parameters of the image to be detected are calibrated, and the false detection rate generated by different detection algorithms is defined

\[ R = \frac{s_i \times f}{M} \times 100\% \]  

(13)

where \( R \) refers to the number of false detections calculated, \( s_i \) refers to the calibrated motion image scale function, \( f \) is the frequency of target detection, and \( M \) refers to the spatial bandwidth function.

A priori frame height: the detection range of the scale parameters of the motion image.

\[ L = \frac{N_i - R \times N_i}{c} \]  

(14)

where \( L \) refers to the detected output of the a priori box height parameter, \( N_i \) refers to the number of motion images in the image set, and \( c \) refers to the number of frames generated during target detection overlaps.

Detection time of the interference environment: process the detection images of two attributes in the image set into a detection group and count the target detection time-consuming results in the interference environment.

Target detection speed: The calculation equation is as follows:

\[ v_c = \frac{m \times L \times N_{a_i}}{A_p \times T} \]  

(15)

where \( v_c \) refers to the calculated target detection speed. \( m \) refers to the average precision (MAP) value of the image output from visual tracking. \( N_{a_i} \) refers to the number of images in the image set. \( A_p \) refers to the image detection scale parameter, and \( T \) refers to the time for image target detection.

Confidence of the image detection frame: The equation is calculated as follows:

\[ h = \frac{v_c e^u}{\sigma_i} + l_0 \]  

(16)

where \( h \) is the constructed grayscale processing function. \( u \) denotes the number of HOG extracted features. \( \sigma_i \) denotes the affiliation function, and \( l_0 \) denotes the fixed pixel value of the image.

Detection accuracy: accuracy is an important indicator to measure the performance of the algorithm. Therefore, the detection accuracy is selected as the indicator to compare the proposed algorithm with the other algorithms.
Results and Discussion

The called video types are the SP-01 and SP-06 motion image sets. By sorting out the motion pictures, it can be seen that there are cone targets in the images in the dataset, and there is no large occlusion in the image range. Therefore, the target imaging is relatively clear. Sort out the false detection results generated by different target detection algorithms in the motion image group and calculate the average false detection rate of motion image sets of different target detection technologies in SP-01 and SP-06. The results of the false detection rate are shown in Figure 3.

According to the data in Figure 3, we sort out the false detections generated by different image target detection technologies in the same motion image. From the error detection rate results are shown in Figure 3. On the SP-01 motion image set and the SP-06 moving image set, the false detection rate of the proposed algorithm is low and has significant advantages, especially on the SP-06 motion image set, the false detection rate is only about 15%. Other algorithms have high error reduction rate. The algorithm of SARMIM (Huan et al., 2020) and NVSODM (Chen et al., 2022) has the highest error detection rate on the two image sets. SLCTD (Xie et al., 2022) is relatively low. However, it is also about 38%. SFSISD (Chen et al., 2021) and TTMID (Qin et al., 2022) both have more than 40%. Compared with the target detection technology compared with other algorithms, the target detection method in this paper has the lowest false detection rate.

Call the SP-01 ~ SP-10 image datasets according to the error detection rate parameters determined above. After the image corresponding to false detection is removed, the scale parameters of the
Motion image are processed quantitatively. According to the frame number overlap generated by adaptive segmentation processing, the frame number overlap caused by scale change is determined. The detection windows of different target detection technologies are fixed to the same size, and the numerical relationship of the a priori frame height is output. Collating the a priori frame heights obtained from different target detection algorithms, the results are shown in Figure 4.

According to the defined value relationship of the detection a priori frame, take the average of the above detection a priori frame height value as the final detection a priori frame test result. It can be seen from the change of the value of the detection prior frame shown in Figure 4 that the mean value of the detection prior frame in SLCTD (Xie et al., 2022) is only 25 pixels, the mean value of the detection prior frame in SFSISD (Chen et al., 2021) is 38 pixels, and the mean value of the detection prior frames in SARMIM (Huan et al., 2020), NVSODM (Chen et al., 2022) and TTMID (Qin et al., 2022) is more than 50 pixels. The height of the detection prior frame of the proposed target detection algorithm is 80 pixels. Compared with the detection algorithms participating in the comparative test, the detection prior frame height of this target detection algorithm is the largest, and the actual detection range is the largest.

Keep the above test environment unchanged, call SP-05 and SP-06 in the image dataset, call MATLAB of the upper computer to simulate a sparse window in the test image, and take the sparse window as the interference in the image scene. After unifying the image resolution, different target detection algorithms are run in the host computer. In the process of target detection and processing,
the task manager that hosts the host computer is invoked. The detection image of the two attributes in the image set is processed as a detection group, and the time-consuming results of different target detection algorithms are counted in the above interference environment, as shown in Table 2.

According to the above sparse processing motion image detection interference environment, when calling the host computer for target detection, the detection time consumed by different target detection algorithms is counted. According to the results shown in Table 2. The method in SARMIM (Huan et al., 2020) consumes an average of 22.5 ms when detecting a group of images, which is the longest detection time. When the methods of SLCTD (Xie et al., 2022), NVSODM (Chen et al., 2022), SFSISD (Chen et al., 2021) and TTMID (Qin et al., 2022) are used to detect image groups of the same specification and content, the average detection time consumed is 15.6 ms, 16.78 ms, 15.0 ms and 18.35 ms, respectively. In the same detection interference environment, the average detection time consumed by the proposed target detection method is only 6.8 ms, which is compared with the target detection technology involved. The proposed target detection algorithm consumes the shortest detection time under interference conditions.

The image sets with attributes SP-07 and SP-08 are called, the two attribute image sets are connected to the same recognition scale range, and the AP values of the recognition scale in the unified image set are 0.49, 0.55, 0.58 and 0.70. Output the MAP value of the calibrated image set of the visual tracking system. After taking this value as the fixed influence parameter, the detection speed generated by different target detection algorithms is determined. Under the same recognition scale AP value, the detection speed results of different target detection algorithms are sorted. The results of the detection speed are shown in Figure 5.

According to the mean MAP value of the image output by the visual processing system, the detection speed of different detection algorithms is defined. According to the speed results shown in Figure 5. The speed curve of the proposed algorithm is at the highest position in Figure 5 and is higher than other algorithms at different recognition scales, up to about 20 fps. The speed curves of SARMIM (Huan et al., 2020), SLCTD (Xie et al., 2022), NVSODM (Chen et al., 2022) and SFSISD (Chen et al., 2021) intersect. The average speed is about 10 fps, and the maximum speed is not more than 15 fps, which is lower than the proposed algorithm. The detection speed of TTMID (Qin et al., 2022) is lower, and the maximum speed is only about 7 fps. Compared with the comparison algorithm,
the proposed target detection algorithm has the highest target detection speed. Call the attribute image sets sp-09 and SP-10, process the image set into a sequence, and use hog to extract the features in the image sequence. The neural network is used to test the features repeatedly, sort out the number of frames in the image set, and extract the features by gray processing. The numerical relationship of gray processing can be expressed as

\[ m(T) = \frac{h + \lambda}{p(s)} \times 100\% \]  

(17)

where \( m(T) \) refers to the determined resulting confidence value, \( \lambda \) refers to normalized parameters, \( p(s) \) refers to functions processed in batch, and all the other parameters remain unchanged. According to the reliability value determined by the above numerical relationship, the reliability results of the detection frames of different target detection technologies are statistically obtained, as shown in Figure 6.

After the motion image target is prepared by gray processing, the credibility numerical relationship is constructed, and the credibility generated by 100 frames of images is detected by a statistical detection process. According to the reliability results shown in Figure 6. The proposed algorithm has absolute advantages. The confidence value of detection frame is about 94%, while the confidence value of detection frame of SARMIM (Huan et al., 2020) algorithm, SLCTD (Xie et al., 2022) algorithm, NVSODM (Chen et al., 2022) algorithm, SFSISD (Chen et al., 2021) algorithm and TTMID (Qin et al., 2022) algorithm is below 90% or even lower. From the above analysis. Compared with the comparison algorithm, the target detected by the proposed target detection algorithm meets the detection expectation, and the detection result is the most accurate.
Comparing the motion image target detection accuracy of proposed algorithm with that of SARMIM (Huan et al., 2020), SLCTD (Xie et al., 2022), NVSODM (Chen et al., 2022), SFSISD (Chen et al., 2021) and TTMID (Qin et al., 2022), the results are shown in Table 3.

According to the data in Table 3, in many tests, the motion image target detection accuracy of different algorithms shows relatively good and stable performance as a whole. Except for the low average detection accuracy of the algorithms in SFSISD (Chen et al., 2021) and TTMID (Qin et al.,

**Table 3.**
Comparison results of target detection accuracy

<table>
<thead>
<tr>
<th>Number of Experiments/Time</th>
<th>Proposed Algorithm/%</th>
<th>SARMIM (Huan et al., 2020)/% Algorithm</th>
<th>SLCTD (Xie et al., 2022)/% Algorithm</th>
<th>NVSODM (Chen et al., 2022)/% Algorithm</th>
<th>SFSISD (Chen et al., 2021)/% Algorithm</th>
<th>TTMID (Qin et al., 2022)/% Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>93.5</td>
<td>76.3</td>
<td>69.3</td>
<td>71.2</td>
<td>59.2</td>
<td>61.2</td>
</tr>
<tr>
<td>20</td>
<td>94.6</td>
<td>75.2</td>
<td>68.5</td>
<td>75.3</td>
<td>56.1</td>
<td>63.2</td>
</tr>
<tr>
<td>30</td>
<td>96.5</td>
<td>74.9</td>
<td>70.1</td>
<td>74.4</td>
<td>52.0</td>
<td>66.5</td>
</tr>
<tr>
<td>40</td>
<td>95.2</td>
<td>75.8</td>
<td>71.2</td>
<td>70.1</td>
<td>53.8</td>
<td>66.8</td>
</tr>
<tr>
<td>50</td>
<td>95.1</td>
<td>76.9</td>
<td>74.0</td>
<td>74.2</td>
<td>58.9</td>
<td>64.9</td>
</tr>
<tr>
<td>Average</td>
<td>95.0</td>
<td>75.8</td>
<td>70.6</td>
<td>73.0</td>
<td>56.0</td>
<td>64.5</td>
</tr>
</tbody>
</table>
2022), the accuracy values of other algorithms are relatively high, and the average detection accuracy of the algorithms in SARMIM (Huan et al., 2020), SLCTD (Xie et al., 2022) and NVSODM (Chen et al., 2022) is more than 70%. However, it can be found that the average detection accuracy of proposed algorithm is 95.0%, which is much higher than other algorithms. Therefore, the motion image target detection algorithm designed by improving YOLO algorithm in this paper has good performance, high accuracy and strong applicability in motion image target detection.

CONCLUSION

This paper uses the improved YOLO algorithm to design motion image target detection. According to the frame structure attribute of the motion image, the detection technology suitable for different target detection specifications is constructed. The results show that the designed target detection technology has a small target false detection rate, high detection a priori frame height, short detection time and high reliability of the detection frame. It can improve the low detection accuracy and poor efficiency of traditional target detection technology. In future work, the detection target feature annotation algorithm will be constructed to realize the accurate classification of the detection target features.

COMPETING INTERESTS

The author declares that there is no conflict of interest with any financial organizations regarding the material reported in this manuscript.

FUNDING AGENCY

No funding was used to support this study.
REFERENCES


Tian Zhang received his B.S. degree from Northeast Normal University and received his M.S degree from Jilin University. He is a Lecturer in ChangChun Normal University. His research interests include computer graphics and image, software engineering.