Hindawi
Computational Intelligence and Neuroscience
Volume 2023, Article ID 9851764, 1 page
https://doi.org/10.1155/2023/9851764

Retraction

Retracted: Optimization Algorithm of Moving Object Detection Using Multiscale Pyramid Convolutional Neural Networks

Computational Intelligence and Neuroscience

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Z. Yang, Z. Bu, and Y. Pan, "Optimization Algorithm of Moving Object Detection Using Multiscale Pyramid Convolutional Neural Networks," *Computational Intelligence and Neuroscience*, vol. 2023, Article ID 3320547, 11 pages, 2023.

Hindawi
Computational Intelligence and Neuroscience
Volume 2023, Article ID 3320547, 11 pages
https://doi.org/10.1155/2023/3320547

Research Article

Optimization Algorithm of Moving Object Detection Using Multiscale Pyramid Convolutional Neural Networks

Zhe Yang,[1,2] Ziyu Bu,[1,2] and Yexin Pan[1,2]

[1]*School of Computer Science and Technology, Soochow University, Suzhou 215006, China*
[2]*Provincial Key Laboratory for Computer Information Processing Technology, Suzhou 215006, China*

Correspondence should be addressed to Zhe Yang; yangzhe@suda.edu.cn

Received 21 July 2022; Revised 10 August 2022; Accepted 12 August 2022; Published 11 March 2023

Academic Editor: Zhao kaifa

Object detection and recognition is a very important topic with significant research value. This research develops an optimised model of moving target identification based on CNN to address the issues of insufficient positioning information and low target detection accuracy (convolutional neural network). In this article, the target classification information and semantic location information are obtained through the fusion of the target detection model and the depth semantic segmentation model. The classification and position portion of the target detection model is provided by the simultaneous fusion of the image features carrying various information and a pyramid structure of multiscale image features so that the matched image fusion characteristics can be used by the target detection model to detect targets of various sizes and shapes. According to experimental findings, this method's accuracy rate is 0.941, which is 0.189 higher than that of the LSTM-NMS algorithm. Through the migration of CNN and the learning of context information, this technique has great robustness and enhances the scene adaptability of feature extraction as well as the accuracy of moving target position detection.

1. Introduction

Images and videos have gradually grown importance as a means of information transmission and acquisition due to the rapid advancement of computer technology. The manual analysis and retrieval method is currently ineffective in the face of the massive amounts of video data that are being created on a daily basis. It is also prone to visual fatigue and judgement errors from prolonged monotonous work. Moving object detection is regarded as one of the most fundamental pieces of work in the application of video analysis and is a very practical and difficult topic in the field of computer vision research. The science and technology of artificial intelligence [1], pattern recognition [2, 3], and computer vision [4] are all used extensively in its technical principle. Military applications, system monitoring, intelligent transportation [5], commodity inspection, and other scenarios primarily reflect its commercial value and application prospects. Traditional detection models typically employ difficult artificial feature extraction techniques, such as scale-invariant feature transformation and directional gradient histogram, to obtain the expression information pertinent to the target in the original input, and then learn the classifier from the extracted feature information pertinent to the target. Due to the complex changes in natural scenes and the interference of man-made noise, traditional methods still have some issues even though the target detection algorithm [6] has produced good results. For example, the algorithm only extracts a single feature, which cannot reflect the influence of other elements. Some features are computationally expensive; because manual features are not universal, different scenes need to design different features, which requires a lot of work and high innovation. As a result, there is no single highly accurate detection method, and ongoing research is still required in the field of object detection technology in complex environments. The target detection algorithm based on CNN has currently taken over the mainstream after years of development. Given this context, it is unquestionably important to investigate in this article the optimization algorithm of moving target detection based on CNN.

CNN is a brand-new network that combines convolutional operation with multilayer artificial neural networks (NNs) [7, 8]. It automatically instructs the computer to perform a convolution operation to extract the desired features from the image, resulting in more universal and natural-looking features. Additionally, it is robust to a certain amount of distortion. The CNN-based target detection algorithm has recently undergone significant development, greatly enhancing both its speed and detection accuracy. Convolution and pool layers were alternately set in the CNNLeNet-5 model, which can transform the input image into a collection of feature maps through a number of nonlinear transformations, and then classify the feature maps using fully connected NN to complete image recognition. The back propagation algorithm was used by CNN's training algorithm to train the network under supervision. Target detection technology primarily addresses the difficult task of classifying, identifying, and localising targets in images or videos. The target detection in video not only completes the target identification and location, but also has the task of target tracking. Applying CNN to the target detection technology, using CNN to train and learn the image features with stronger representation ability to replace the artificially designed features created by human prior knowledge and intelligent design in the past, can make the target detection technology have great progress. However, CNN needs a lot of specimen data for training, and high-quality data sets are scarce. Therefore, when the data set is scarce, the model trained by CNN is often ineffective.

Additionally, the target recognition accuracy is not very high due to the early traditional CNN's shallow hierarchy. In order to recognise moving targets, this article employs a convolution network model with a more complex structure. The following are its innovations:

(1) The algorithm architecture in this article can reduce the parameters of the algorithm as much as possible, while preventing overfitting. At the same time, the model also randomly samples around the pixels to obtain the pixel slice features of neighboring pixels, so that the network can get the constraints of spatial context information, thus improving the classification performance of the model.

(2) This article proposes a prediction structure that fuses the semantic information of multilevel feature maps to address the issue of how to improve the expressive ability of features by fusing the detailed semantic information of low-level convolution features with the abstract information of high-level convolution features. The moving target detection model suggested in this article is not constrained by the simple pixel model and the artificially designed features, so it can better adapt to the complex application scenarios in the real world. This is in contrast to the traditional target motion detection methods.

2. Related Work

Researchers have put forth a number of target detection techniques after years of investigation. Target detection techniques can be roughly categorised into two groups based on different technical approaches: target detection techniques based on template matching and target detection techniques based on image classification. A face detection technique based on CNN network structure was proposed by Lei et al. This technique has gained popularity as a top face detection technique in the face recognition system due to its excellent robustness from all facial directions. Additionally, the detection and identification system is used in real-world situations [9]. A brand-newCNN-based target detection model was put forth by Roth et al. This model adopts the component detection module, which reduces the amount of calculation by breaking the complex target into multiple distinct components for detection [10]. According to Tran and Hoang, because CNN typically and alternately performs convolution and pooling operations on the input image, the features gradually abstract from low-level to high-level features after numerous nonlinear transformations, giving the features a strong capacity for expressiveness toward the target [11]. Han et al. proposed using SPP-Net to improve the detection speed and accuracy of RCNN. The problem that the whole connection layer will limit the input size is solved by pooling the spatial pyramid. The input image does not need to be cut or scaled, which improves the detection accuracy. SPP-Net only needs to extract the forward CNN feature of an image once, which significantly improves the detection speed [12]. Long et al. put forward DenseNet classification model, in which multiple blocks are stacked in the network, and each convolution layer in each block has a part of feature map that propagates to the next convolution layer, so that the characteristic information of the image can flow better in the network [13]. Ruff et al. proposed to apply CNN to pedestrian detection. In this method, the training samples are used to fine-tune the whole CNN in a supervised way [14]. Tang et al. used a recursive equation to dynamically update the weight of each Gaussian function, and the number of Gaussian functions used in each pixel can be adjusted as needed [15]. Andrearczyk and Whelan proposed to combine the task of pedestrian detection with the task of learning context to optimize CNN by multitask training. The learned context information contains the attributes of pedestrians and scenes, which effectively assists CNN in pedestrian detection and reduces false positives [16]. Kumam and Bhatnagar proposed that multilayer feature maps should be fused before the target candidate frame is generated to obtain multiscale superfeatures, and then the feature expression of the target can be enhanced. HyperNet greatly improves the accuracy of target detection through more accurate candidate frames and proves the importance of candidate frames to target detection [17]. Li et al. proposed another region-based method. This method detects motion by using statistical cyclic shift moments in image areas [18].

Large gaps in target detection were simple to appear in the past due to the poor anti-interference ability of algorithmic detection results. In this article, a brand-newCNN-based optimization algorithm for moving target detection is put forth. The component detection module is used in this model, which breaks the complex target down into various components and detects each one separately to minimise the amount of calculation. The training model learns the classification rules of the targets after identifying the labels of hidden variables in unlabeled samples. Additionally, to hasten training convergence and raise target recognition accuracy, this article combines two strategies for fine-tuning training simultaneously. The network is aided in finding the target object by extending the candidate border into the search area and employing a variety of different input areas, which improves the accuracy of the detection result's location. The models proposed in this article have excellent performance in moving target detection when compared to the current mainstream deep learning and traditional methods, according to extensive experiments on common data sets.

3. Methodology

3.1. Application of NN in Moving Target Detection. By examining the geometrical or statistical properties of the target, one can precisely identify and segment it from the image or video. Researchers have gradually expanded their study of deep learning using computers for data acquisition and training as a result of the ongoing development of computer hardware and Internet technology. Additionally, CNN-based structure has emerged as the primary detection and identification technique being thought of by many researchers [19]. In the fields of computer vision and digital image processing, CNN, an effective artificial neural network built on the foundation of conventional artificial neural networks, is of great importance. Traditional computer systems have difficulty describing nonlinearity and creating corresponding models in logical terms, whereas NN has inherent advantages in this area. It has a good overall approximation to complex nonlinear functions by simulating the structure of human brain neurons and having excellent learning capacity. A multilayer perceptron network that can perform input-output mapping is what deep neural networks, or "deep NN," are essentially. CNN is a type of deep NN. Its primary feature is that it uses local connections and weight sharing, which on the one hand reduces the parameters of network training and makes network optimization easier; on the other hand, it prevents overfitting to some extent. The processing mechanism of the visual system in a mammalian cat served as the model for CNN's multilayer network structure. The entire visual field is covered by the local receptive field after it has been tiled for convolution operation. Convolution operation is performed on the input image by CNN to extract image features while at the same time using a convolution kernel with shared values based on the local receptive field, which can further reduce the number of weights in the CNN model [20]. To automatically extract the target features from the input image, the deep learning theory is applied, and the convolution kernel is calculated by the computer. However, CNN is fairly resilient to a certain amount of distortion and transformation. Various detection techniques have had excellent success in identifying and detecting targets based on these kinds of points. Both forward propagation and backward propagation are used in convolutional network models. The output of the network is computed using the forward propagation algorithm, not adjusted. Each neuron calculates its layer's output and transmits it to the layer below it until it reaches the output layer, where it calculates the network's output result. The set loss function is minimised using the gradient descent method, the weight parameters of the network are adjusted using the back propagation error, and iterative training is used to continuously raise the network's accuracy [21]. Typically, depth is the basis for the classification model flow. Following the processing of the input data by multiple stacked convolution layers, activation function layers, and pool layer modules, CNN primarily extracts the image's features. Next, it feeds the predicted loss into the full connection layer and the objective function, and finally updates the network parameters in accordance with the objective function's loss back propagation. Figure 1 displays the structure of NN.

The CNN model can input the original image directly, which can reduce the number of labor-intensive preprocessing operations on the original image, streamline the workflow, and increase productivity. Weight sharing is a characteristic of convolution neurons, which lowers the number of nodes and parameters. The range is controlled by the local receptive field, which also reduces the complexity of the model. The image is output using the down sampling layer, and the features are extracted to reduce the dimension. With CNN, which was specifically created for images, the local features of the previous layer are used to derive the features of the convolution layer via weight sharing. When CNN's structure is properly adjusted, CNN can be made to have the regression capabilities required by various applications. For feature extraction and compression of the input image, the CNN model typically decides to alternately place the convolution layer and sampling layer in the lower layer of the network. At the top level of CNN model, the full connection layer will be selected to perform a series of processing such as regression classification on the image features extracted from the previous layer-by-layer transformation and mapping, and at the same time, the extracted image features will be aggregated into feature vectors as a new representation of the input image. Because the convolution operation and down sampling operation are added to the basic structure of artificial NN, the extracted features of CNN have some invariable properties in space, and it is more suitable for image detection than traditional NN. At present, in the mainstream CNN image classification models, multiple convolution layers are stacked in each module to extract image features. CNN introduces convolution calculation and sampling operation, so that the computer can automatically learn the target features from the input image, and it has good recognition effect for different targets, and also has good robustness to some degree of distortion and other changing factors. These advantages are based on its three main features: sparse connection, weight sharing, and sampling.

Generally, the low-level structure of CNN consists of convolution layer and pool sampling layer alternately, while the high-level structure consists of full connection layer and

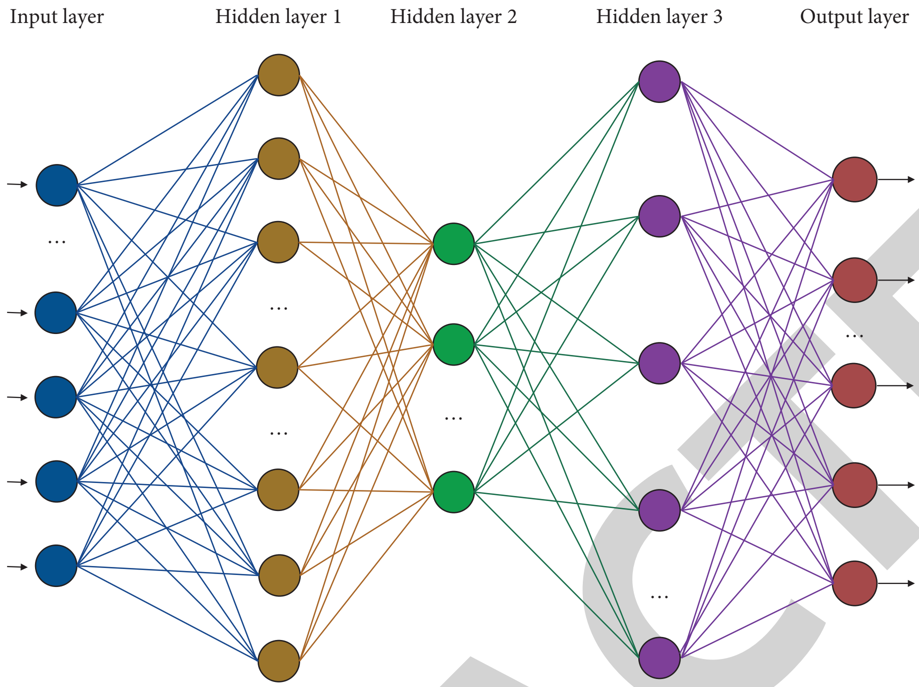

Figure 1: NN structure.

corresponding classifier. The input of the first fully connected layer is the feature image of the lower structure after feature extraction, and the last layer uses the classifier as the output layer. Among them, logistic regression, Softmax regression, and support vector machine can be used to classify the images finally. In the model structure, the characteristic weight sharing is used to perform convolution kernel operation on the whole image with the same convolution kernel, so as to generate a complete feature graph. This operation will greatly reduce the scale of weight parameters required by the network structure. Under the same circumstances, the number of weight parameters is greatly reduced after adopting the weight sharing operation, which is one of the most significant evolutions of CNN. CNN adopts down sampling operation, which is located behind the convolution layer. It down samples the features of the convolution output layer and integrates the information. It is possible to reduce the computational complexity of the network by greatly reducing the hidden layer, including the number of units between the input layer and the output layer. Additionally, the model exhibits some aspects of spatial invariance. We can gradually classify images from easy to difficult by cascading multiple CNNs, which will lessen the complexity of training.

3.2. CNN Structure. Three components make up the convolution layer, sampling layer, and full connection layer in a conventional CNN model. Additionally, the central components of feature extraction are the convolution layer and pool layer. This section will introduce the CNN model's convolution layer, pool layer, activation function layer, full connection layer, and target loss function, in that order. Convolution layer is crucial to the success of CNN network model because it is the primary structure for extracting image features. The number of parameters and the amount of calculation in the convolution layer can be kept within a reasonable range by the sparse connection mechanism and the shared weight scheme. In reality, CNN uses a two-dimensional discrete convolution operation for convolution. Each convolution kernel conducts a convolution operation on the local receptive field of the feature map of the preceding layer as it connects each layer to the one before it. Convolution kernel parameters are connected to local pixels in the corresponding feature graph and can be thought of as traditional NN weight parameters. Each convolution layer in the entire CNN classification model is made up of convolution kernels with multiple channels, so as convolution kernel parameters increase, those of large convolution kernels will increase significantly relative to those of small convolution kernels. The process of calculating each feature map in the convolution layer can be broken down into three steps: first, various convolution kernels are convolved with the feature map corresponding to the previous layer; and second, the result is multiplied by the number of feature maps in the convolution layer.The associated convolution offsets and results are then added up. Finally, a nonlinear activation function is used to combine the results, and the result is a convolution layer characteristic map. The input image is mapped to the hidden layer feature space following network convolution. The fully connected layer then performs a classifier by mapping the hidden layer's feature representation into the sample label space.

After the input has passed through the convolution layer, the obtained result needs to be transformed nonlinearly by the activation function, which is equivalent to convolving the input and then connecting a nonlinear connection layer. Therefore, the activation function plays a vital role in deep CNN. Sigmoid function is mathematically defined as follows:

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{x}}. \tag{1}$$

Because the function curve is smooth, continuous, and monotonic and because the value is between 0 and 1, it can be viewed as a probability distribution function for classification problems, it is frequently used as the threshold function of NNs. The nonzero centre of the Sigmoid function is a problem that is resolved by the emergence of the Tan*h* function, whose convergence rate is faster than that of the Sigmoid function. Tan*h*'s output and input can maintain a nonlinear monotonic rising and falling relationship, which is consistent with the BP network's gradient solution. It is described as follows:

$$F(x) = \tanh(z) = \frac{e^{z} - e^{-z}}{e^{x} + e^{-z}}. \tag{2}$$

The ReLU function is defined as follows:

$$\varnothing(x) = \max\ (x, 0). \tag{3}$$

When the input signal is less than 0, the output of the ReLU activation function is entirely 0. When the signal is greater than 0, the input and output are linearly equal.

The CNN network model's sampling layer, also referred to as the pool layer, is crucial for lowering the computational complexity of the model and the resolution of image features. Features that are useful in one area of an image may also be applicable to other areas of the image because images are static. The specific operation process of the pool layer is to aggregate the features of small areas on the previous feature map for statistics in order to describe large image information. Maximum pool refers to the result of taking the maximum value as the input and average pool refers to the calculation of the average value of pixels in small areas as the pooled value. The small area currently corresponding to the previous layer is then replaced by this value on the feature map of the subsequent layer after the result value has been used to replace the current small area. The sampling layer typically comes after the convolution layer and operates statistically on the limited set of image features that the previous convolution layer extracted. The sampling layer is static, and its parameters don't affect how back propagation is changed. The pool layer's primary purpose is down sampling and further reducing the number of parameters by removing pointless samples from the feature graph. The two most popular pooling techniques in the CNN model are maximum pooling and mean pooling. The CNN classification model, in general, primarily employs the maximum pool operation. Through numerous convolution layers and pool layers, CNN gradually transitions from low-level features to high-level features in order to extract the features of the input image. The output layer and full connection layer classify the high-level features, and a vector is created to represent the category of the input image. Figure 2 depicts the CNN organisational structure.

The whole connection layer is usually set at the tail of CNN, and all neurons between adjacent layers are connected with weights, which is the same as the connection way of neurons in traditional NN. The classifier for the entire model is implemented throughout the connection layer. The image features extracted from the previous convolution layer and pool layer are mapped to the mark space of the sample, and the input of each node in the entire connection layer is connected with each output node in the previous layer. After feature extraction, the entire connected layer is located, and every neuron in the previous layer is connected to every other neuron in the entire connected layer. All of the connection layer's high-level features can be mapped in accordance with the specific output layer tasks. The primary purpose of the entire connection layer is to combine the two-dimensional feature map's feature information into a one-dimensional feature vector, which is then used by the classifier to classify the image using the extracted one-dimensional feature vector. The output layer is the final output structure, which will be classified as the output of the deep learning network structure, and the image classification category will be judged by using the regression function. The traditional CNN parameters are huge in scale, and the training process is complex and redundant, which invisibly increases the computational overhead and affects the network performance. This article will improve it. The Softmax classifier is shown as follows:

$$y(x_i) = \frac{\exp(x_i)}{\sum_{i=1}^{M} (\exp(x_i))}. \tag{4}$$

In the formula, x is the feature vector output by the fully connected layer and M is the number of categories of the classification.

3.3. Construction of Moving Target Detection Model Based on CNN.

The artificial feature modelling method involves extracting the features from the original image using a feature extraction technique that was artificially created, putting those features into a classifier so it can learn the rules of classification, and then letting the trained classifier find the new target. The main steps are as follows: information collection, preprocessing, feature extraction, feature selection, and classification training. In the process of video sequence production, transmission, and recording, these images are easily mixed with other noises, resulting in data loss. The image quality will be degraded, and even compression or transmission errors will occur. Therefore, in the process of target detection, pretreatment must be carried out to reduce the interference of noise. The preprocessing process is to denoise the detected image, enhance the image and transform the color space. The process of window sliding is to slide a window with a fixed size in the detected image, and take the subimage in the window as the candidate area. CNN model can directly input the original image, which can avoid a large number of complicated preprocessing operations on the original image, simplify the working steps, and improve the working efficiency. At the same time, CNN can completely recover the required information from incomplete or noisy input signals. In this article, through preprocessing, the final detection model can obtain features more in line with the characteristics of the target, and improve the accuracy of target detection.

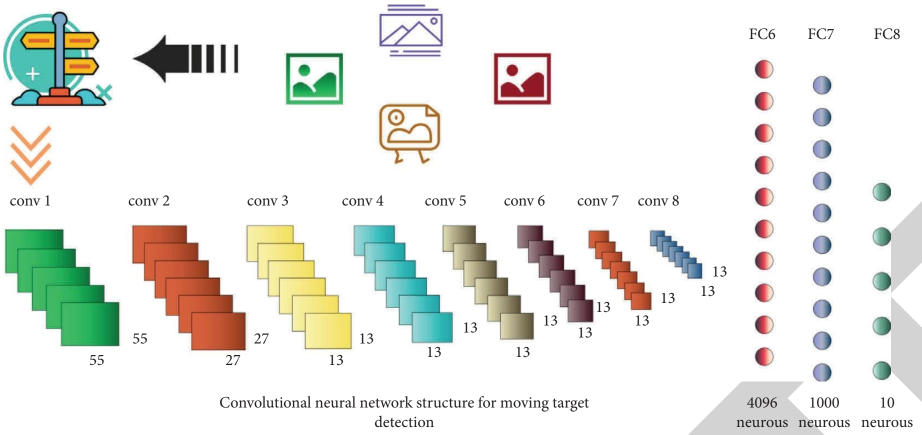

Figure 2: The CNN structure of this article.

Median filtering is a common postprocessing method in motion detection, which is mainly used to eliminate outliers and smooth noise in foreground images. The output of the network is filtered by spatial median, and then each pixel value is processed by global threshold. Mean filtering is a linear filtering algorithm in spatial domain. The theoretical principle uses smooth template to convolute the image and replace the target pixel value with the neighborhood average around the central pixel of the template, so as to remove the noise existing in the image. By averaging the pixels in the template, the pixel (x, y) we need to process is replaced by the average value obtained, and the gray value of the pixel to be processed is $g(x, y)$, that is,

$$g(x, y) = \frac{1}{N} \sum_{f \in I} f(x, y), \tag{5}$$

where the total number of pixels in the template is noted because, typically, we compute the weighted average of the template to ensure that the processed pixels' grey values are consistent with the original pixels. In general, the median filter's processing effect is superior to that of the mean filters because it preserves more of the image's feature information while maintaining the sharp inner edge of the image. However, using median filtering may produce a very rough background image when there are many moving objects in the video, which will have a significant negative effect on the detection model.

In this article, the filtering algorithm in the convolution model adopts the following formula to describe the model structure risk minimization function for training:

$$\beta = \min \sum_{i=1} (f(x) - y)^2 + \alpha \|\theta\|^2. \tag{6}$$

The linear transformation function input by the image is shown in the following formula:

$$\partial = \sum_{i} \beta_i \omega(x_i). \tag{7}$$

The coefficient ω vector can be used to represent the weight ∂ value. In the target detection task using the correlation filter function, the image information feature sample is usually used as the input variable, and the involved kernel function is shown in the following formula:

$$\theta = \theta\left(x_i, x_j\right). \tag{8}$$

The input image is detected by the filter β generated after training, and its image information is calculated and displayed in response to the calculation, as shown in the following formula:

$$f(y) = \sum \beta_i \theta(y, x_i). \tag{9}$$

After calculating the response value of its image feature, the filter update can be performed on the matched image sample feature. As the gradient of information transmission, specifically, it is the gradient of the error to the weight parameter W and the gradient of the bias parameter $\partial E/\partial W$. The weight parameter b and the bias parameter $\partial E/\partial H$ are adjusted according to these two gradients. In the simplest case, the updated formulas for both are as follows:

$$\begin{aligned} \Delta W &= -\eta \frac{\partial E}{\partial W}, \\ \Delta b &= -\eta \frac{\partial E}{\partial b}. \end{aligned} \tag{10}$$

In the formula, ΔW is the update amount of the weight parameter W, Δb is the update amount of the bias parameter b, and η is the learning rate. The model proposed in this article uses the momentum formula to calculate the update of the parameters as follows:

$$\begin{aligned} \Delta W^l(t+1) &= \mu \Delta W^l(t) - \eta \frac{\partial E}{\partial W^l}, \\ \Delta b^l(t+1) &= \mu \Delta b^l(t) - \eta \frac{\partial E}{\partial b^l}, \end{aligned} \tag{11}$$

Table 1: Model learning rate.

Iterative period	Learning rate
1	50
2	50
3	20
4	20
5	10
6	10
7	5
8	5

where μ is the momentum coefficient. This added momentum term reduces the stepping in the direction of high curvature, thereby indirectly increasing the effective learning rate in the direction of low curvature, improving the speed of convergence in the learning algorithm.

Three steps make up the CNN's training process: forward calculation, backward calculation, and parameter updating. The correct gradient calculation is the key to the back propagation algorithm, and it is necessary to calculate the gradient of each layer input and parameter with respect to the objective function. To make the calculation of the gradient of the previous layer parameter about the objective function by applying the chain rule easier, the gradient of each layer input about the objective function is computed. The parameters of the network are updated by back propagation during training, and the loss between the predicted category and the actual category is calculated according to the loss function. The detection boxes are categorised during testing using the Softmax function. In this article, by increasing the step size of convolution kernel in some convolution layers to replace the original down sampling function of the maximum pool layer in the network model, the complexity and calculation amount of the model are reduced; so as to improve the running efficiency of the model based on depth separable convolution. In addition, this model adopts an adaptive learning rate reduction method. This method uses a larger initial value to accelerate convergence, and in the subsequent learning process, the learning rate will be judged according to the difference between the current cost function and the previous cost function. If the difference is not obvious, the current learning rate is halved as the new learning rate. As shown in Table 1.

4. Result Analysis and Discussion

In this article, INRIA human database was used to evaluate the multicomponent detection effect of complex targets. This database contains all kinds of standing pedestrian images with different human postures and background environments collected from GRAZ01 data set, personal digital images, network images, etc. Each image has a pedestrian height of at least 100 pixels. Among them, 85% of the samples are used as training sets; 15% of the samples are used as the test set. Image graphics card can provide parallel accelerated computing function for CNN network model on a reasonable software platform. Graphics card in this article is NVIDIAGeForceGTX1080GPU. Operating system is Ubuntu, GPU graphics card nvidiadrok6000, memory is 64 G, based on Tensorflow deep learning framework. The software platforms used are Caffe 2 and OpenCV 3.2, and the programming languages used are C++ and Python. The mini-batch size of the improved CNN model structure training stage is 256; The ratio of positive and negative samples is 1 : 3; The initial learning rate is 0.0001; Train 10 k rounds. The improved CNN model uses Softmax function to classify the categories of detection boxes in the testing phase. Table 2 shows the test results of insulator data set.

Table 2: Detection results of insulator dataset.

Method	Detection accuracy (%)
LSTM-NMS	91.51
RPN + conv3	88.32
STN-RPN	89.54
Faster RCNN	90.36
Improve CNN	94.57

In reality, there are usually special target detection and scene detection tasks, but due to the limited source of data sets, the detection tasks cannot be completed because of insufficient data sources. Therefore, in this article, data enhancement is used to avoid the problem of insufficient data sources, which can effectively improve the detection target recognition rate and can be well applied to CNN model training. The comparison before and after data enhancement is shown in Figure 3.

ROC curve was used to visually represent the experimental results of each foreground detection method. Calculate the true positive rate (TPR) and false positive rate (FPR), and then draw the ROC curve. ROC curves of each algorithm are shown in Figure 4. The closer the curve is to the longitudinal axis of TPR, the lower the false detection rate of experimental results. The closer to the horizontal axis of FPR, the higher the false detection rate of experimental results.

The interframe difference method has the worst detection effect in the data set, and there are a lot of false detections in the complex scene with global illumination shadow, as can be seen from Figure 4. Vibe algorithm has better detection results and a lower false detection rate when compared to experimental results of the interframe difference method. The algorithm described in this article has the best checking effect of all of them. The method used in this article to address the issue of identifying small targets means that background-filled images not only do not contain the target information itself but also have an impact on training results. The detection effect of the images on small targets will be improved if the small targets are processed centrally. As a result, the images are joined together, and the joining is performed in an unfixed mode. Precision rate and recall rate are typically used in target detection to assess the effectiveness of the detection model. Simply put, the recall rate shows the proportion of correctly identified objects to all the objects that needed to be detected. The accuracy rate represents the percentage of targets that were correctly predicted across all suggestion boxes by the target detection

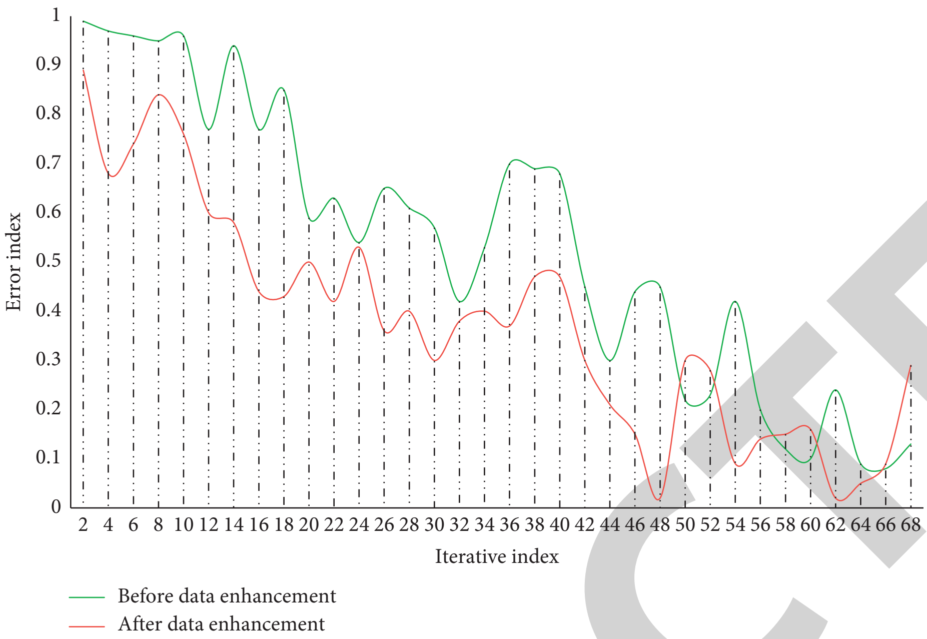

Figure 3: Comparison before and after data enhancement.

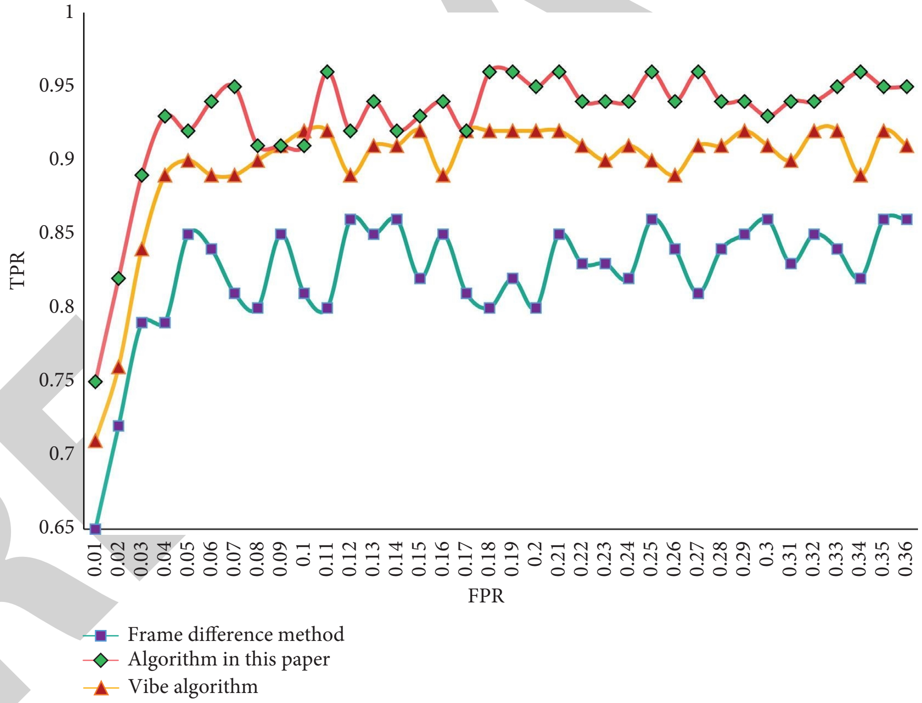

Figure 4: ROC curve of each algorithm.

model. This section has been put to the test, and Figure 5 displays the accuracy test results for various algorithms. Figure 6 displays the recall test results for various algorithms.

In addition, this article uses a curve called DET (Detection Error Trade-off) to show the global detection effect of the model. DET curve uses the logarithm of missed detection rate relative to the first error rate of unit window to evaluate the detection performance of the model. Figure 7 shows the DET curves of different models.

The DET graph illustrates how the detection effect of the model on the target improves with decreasing curve position. Figure 7 illustrates how the model suggested in this article outperforms other comparison models in terms of overall detection performance. The experimental results conclusively demonstrate that the model presented in this

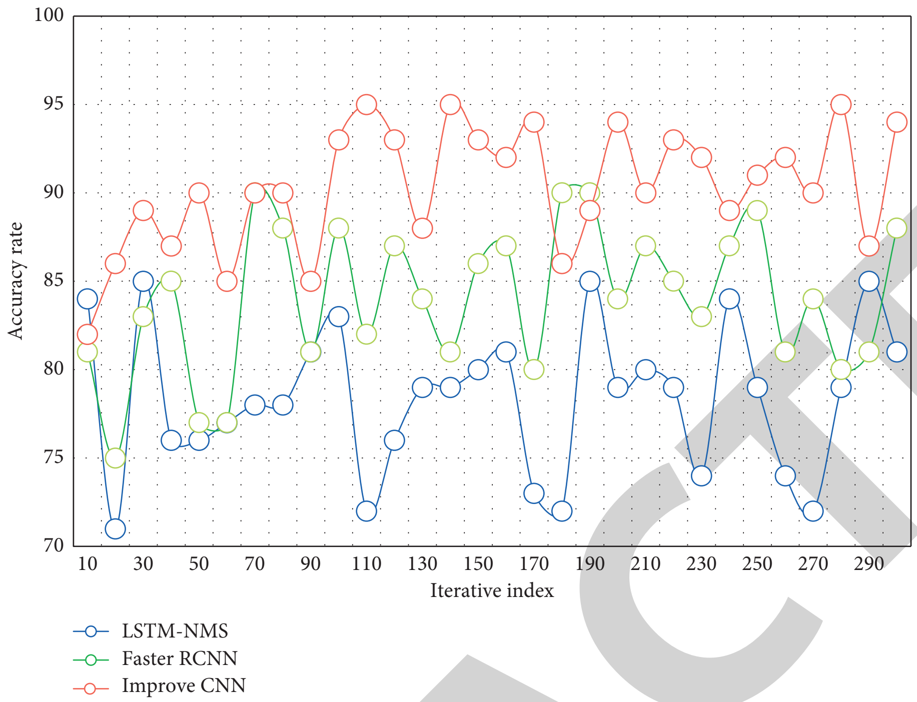

FIGURE 5: Accuracy test results of different algorithms.

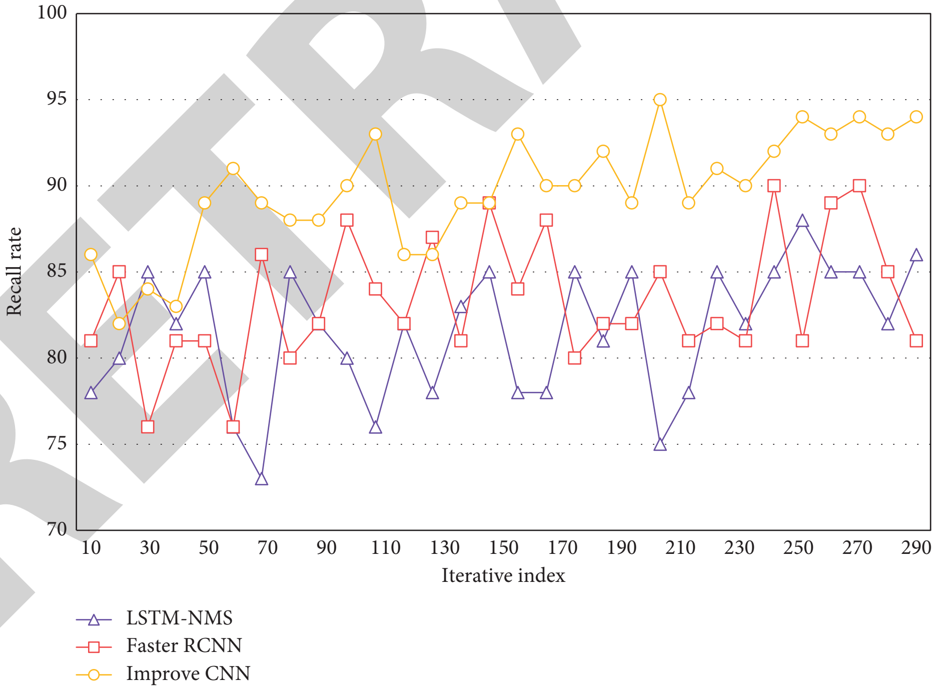

FIGURE 6: Recall test results of different algorithms.

article is more effective at detecting complex targets than some of the main detection models currently in use. The dynamic background is tested experimentally using the dynamic background data set in change detection. Table 3 displays the performance evaluation findings for each algorithm.

Table 3 shows that the results of data statistics and experimental detection agree with each other. The algorithm suggested in this article has a false detection rate that is lower than that of the comparison algorithm and a higher accuracy than the comparison algorithm. Because it can extract moving objects better, the improved method's experimental

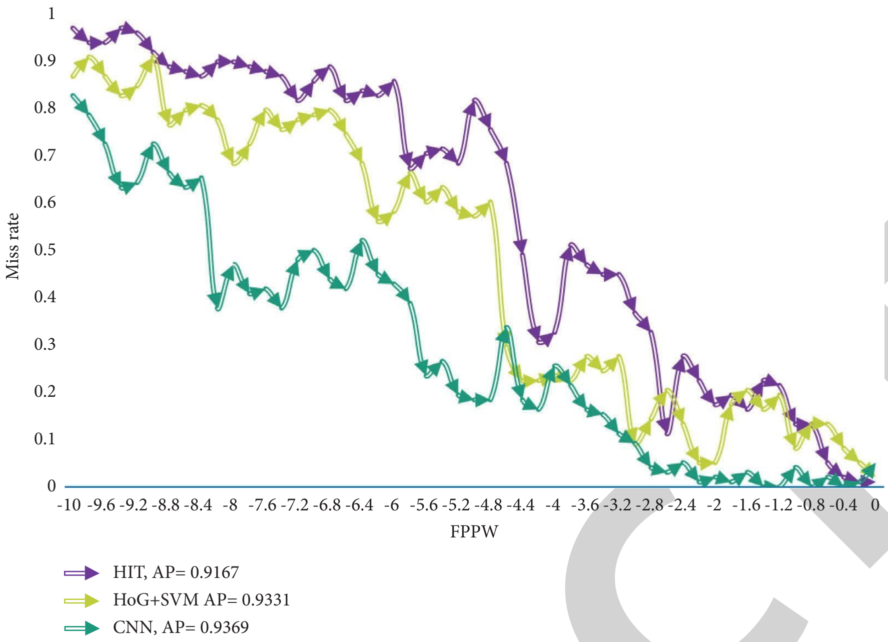

Figure 7: DET curves of different models.

Table 3: Performance evaluation results of each algorithm.

Algorithm	Accuracy rate	False detection rate
LSTM-NMS algorithm	0.752	0.0219
Vibe parameter optimization algorithm	0.736	0.0311
YUV_vibe algorithm	0.841	0.0254
Fasercnn algorithm	0.918	0.0271
Improve CNN algorithm	0.941	0.0197

results in this article outperform those of the comparison algorithm. The experimental results in this section show that the accuracy of this algorithm is 0.941, which is higher than that of LSTM-NMS algorithm by 0.189. This method improves the scene adaptability of feature extraction and the accuracy of moving target location detection by migrating CNN and learning context information. It accelerates the convergence speed of training and has strong robustness.

5. Conclusions

Target detection is very difficult, with the two main challenges being: (1) For moving targets, scale differences, local occlusion, attitude changes, and other factors will significantly alter the targets' appearance, leading to false positives in target detection. (2) The more complex the scene, the harder it is to tell the target from the nontarget, leading to false positives and false negatives in target detection. For the scene, the appearance of the target will also be deformed due to factors such as the change of illumination and visual angle. This article develops an optimised model of moving target detection based on CNN to address the issues of insufficient positioning information and low target detection accuracy. In this article, the target classification information and semantic location information are obtained through the fusion of the target detection model and the depth semantic segmentation model. This article suggests a prediction structure that fuses the semantic information of multilevel feature maps in order to address the issue of how to improve the expressive ability of features by fusing the detailed semantic information of low-level convolution features with the abstract information of high-level convolution features. The moving target detection model suggested in this article is not constrained by the simple pixel model and the artificially designed features, so it can better adapt to the complex application scenarios in the real world. This is in contrast to the traditional target motion detection methods. According to experimental findings, this algorithm's accuracy rate is 0.941, which is 0.189 higher than that of the LSTM-NMS algorithm. In this article, the model reduces computation and gets around the issue of insufficient parameter learning. Additionally, it accomplishes the desired result and displays improved robustness to noise and shadow interference. This article is valuable both commercially and practically. In our upcoming research, we will be able to deepen the testing speed of the model by optimising the learning network, the suggestion box information, and the feature information of the detection box.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 62002253), National Science Foundation of the Jiangsu Higher Education Institutions of China (no. 19KJA230001), and project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

[1] Y. Lin, F. Fan, J. Zhang et al., "DHI-GAN: improving dental-based human identification using generative adversarial networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 1, pp. 1–13, 2022.

[2] Z. Zhang, Y. Ding, X. Zhao et al., "Multireceptive field: an adaptive path aggregation graph neural framework for hyperspectral image classification," *Expert Systems with Applications*, vol. 217, Article ID 119508, 2023.

[3] W. Cai, X. Ning, G. Zhou et al., "A novel hyperspectral image classification model using bole convolution with three-direction attention mechanism: small sample and unbalanced learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–17, 2023.

[4] Y. Ding, Z. Zhang, X. Zhao et al., "Unsupervised self-correlated learning smoothy enhanced locality preserving graph convolution embedding clustering for hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.

[5] Y. Lin, L. Deng, Z. Chen, X. Wu, J. Zhang, and B. Yang, "A real-time ATC safety monitoring framework using a deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 11, pp. 4572–4581, 2020.

[6] J. Zhang, Z. I. Ye, X. Jin, J. Wang, and J. Zhang, "Real-time traffic sign detection based on multiscale attention and spatial information aggregator," *Journal of Real-Time Image Processing*, vol. 19, no. 6, pp. 1155–1167, 2022.

[7] X. B. Jin, Z. Y. Wang, J. L. Kong et al., "Deep spatio-temporal graph network with self-optimization for air quality prediction," *Entropy*, vol. 25, no. 2, p. 247, 2023.

[8] X.-B. Jin, Z.-Y. Wang, W.-T. Gong et al., "Variational bayesian network with information interpretability filtering for air quality forecasting," *Mathematics*, vol. 11, no. 4, p. 837, 2023.

[9] C. Lei, D. H. Ma, H. Q. Zhang, and Lm Wang, "Moving target network defense effectiveness evaluation based on change-point detection," *Mathematical Problems in Engineering*, vol. 2016, no. 6, Article ID 6391502, 11 pages, 2016.

[10] H. R. Roth, L. Lu, J. Liu et al., "Improving computer-aided detection using CNNs and random view aggregation," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1170–1181, 2016.

[11] D. P. Tran and V. D. Hoang, "Adaptive learning based on tracking and ReIdentifying objects using convolutional neural network," *Neural Processing Letters*, vol. 50, no. 1, pp. 263–282, 2019.

[12] J. Han, D. Zhang, G. Cheng, N. Liu, and D. Xu, "Advanced deep-learning techniques for salient and category-specific object detection: a survey," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 84–100, 2018.

[13] T. Long, Z. Liang, and Q. Liu, "Advanced technology of high-resolution radar: target detection, tracking, imaging, and recognition[J]," *Science China Information Sciences*, vol. 62, no. 4, pp. 1–26, 2019.

[14] Z. J. Ruff, D. B. Lesmeister, C. L. Appel, and C. M. Sullivan, "Workflow and convolutional neural network for automated identification of animal sounds," *Ecological Indicators*, vol. 124, no. 2, Article ID 107419, 2021.

[15] F. Tang, X. Zhang, S. Hu, and H. Zhang, "Convolutional features selection for visual tracking," *Journal of Electronic Imaging*, vol. 27, no. 3, p. 1, 2018.

[16] V. Andrearczyk and P. F. Whelan, "Using filter banks in Convolutional Neural Networks for texture classification," *Pattern Recognition Letters*, vol. 84, no. 11, pp. 63–69, 2016.

[17] M. Kumar and C. Bhatnagar, "Hybrid tracking model and GSLM based NN for crowd behavior recognition[J]," *Journal of Central South University*, vol. 24, no. 9, pp. 147–157, 2017.

[18] J. Li, G. Li, and H. Fan, "Image reflection removal using end-to-end convolutional neural network," *IET Image Processing*, vol. 14, no. 6, pp. 1047–1058, 2020.

[19] N. Liu, Y. Xu, Y. Tian, H. Ma, and S. Wen, "Background classification method based on deep learning for intelligent automotive radar target detection," *Future Generation Computer Systems*, vol. 94, no. MAY, pp. 524–535, 2019.

[20] J. Wang, X. Wang, K. Zhang, Y. Cai, and Y. Liu, "Small UAV target detection model based on deep neural network," *Xibei Gongye Daxue Xuebao/Journal of Northwestern Polytechnical University*, vol. 36, no. 2, pp. 258–263, 2018.

[21] C. Wang, J. Zheng, and J. Bo, "Deep NN-aided coherent integration method for maneuvering target detection," *Signal Processing*, vol. 182, no. 9, Article ID 107966, 2021.