

## *Retraction*

# **Retracted: Deep Learning-Based Multitarget Motion Shadow Rejection and Accurate Tracking for Sports Video**

### **Complexity**

Received 15 August 2023; Accepted 15 August 2023; Published 16 August 2023

Copyright © 2023 Complexity. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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] C. Duan, "Deep Learning-Based Multitarget Motion Shadow Rejection and Accurate Tracking for Sports Video," *Complexity*, vol. 2021, Article ID 5973531, 11 pages, 2021.

## Research Article

# Deep Learning-Based Multitarget Motion Shadow Rejection and Accurate Tracking for Sports Video

Chunxia Duan 

Department of Physical Education, Tianjin University of Commerce, Tianjin 300314, China

Correspondence should be addressed to Chunxia Duan; [dcx1977@tjcu.edu.cn](mailto:dcx1977@tjcu.edu.cn)

Received 14 April 2021; Revised 17 May 2021; Accepted 2 June 2021; Published 11 June 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Chunxia Duan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The effect is tested in various specific scenes of sports videos to complete the multitarget motion multitarget tracking detection application applicable to various specific scenes within sports videos. In this paper, deep neural networks are applied to sports video multitarget motion shadow suppression and accurate tracking to improve tracking performance. After the target frame selection is determined, the tracker uses an optical flow method to estimate the limits of the target sports video multitarget motion based on the sports video multitarget motion of the target object between frames. The detector first scans each sports video image frame one by one, observing the previously discovered and learned image frame subregions one by one until the current moment that is highly like the target to be tracked. The preprocessed remote sensing images are converted into grayscale images, the histogram is normalized, and the appropriate height threshold is selected in combination with the regional growth function to realize the rejection of sports video multitarget motion shadow and establish the sports video multitarget network model. The distance and direction of the precise target displacement are determined by frequency-domain vectors and null domain vectors, and the target action judgment mechanism is formed by decision learning. Finally, comparing with other shadow rejection and precision tracking algorithms, the proposed algorithm achieves greater advantages in terms of accuracy and time consumption.

## 1. Introduction

With the rapid development of target tracking technology, target tracking has been successfully applied to many practical systems. In the field of sports video surveillance, public security departments can rely on target tracking technology, using today's cameras all over the world to retrieve surveillance footage to track suspects or help find lost elderly and children, overcoming the disadvantages of the past need to rely on human resources, greatly improving the efficiency of work; in the medical field, the tracking of lesion cells and carrying out interventional treatment can be achieved with the help of target tracking technology, which helps modern medicine to achieve greater breakthroughs [1]. In the military field, the target to be combated is often in the state of sports video multitarget movement, which requires the use of target tracking technology to firmly lock the target to achieve accurate combat; in the field of intelligent transportation, with the continuous advancement of modernization, the number of urban vehicles has increased

sharply, and the probability of traffic accidents has also increased significantly, when there are traffic violations. In the field of unmanned transportation, with the development of artificial intelligence and the arrival of the 5G era, automatic driving is bound to gradually spread to people's daily life, and target tracking, as a key technology of automatic driving, can provide precise identification and track for vehicles through the traffic sign line, vehicles in front and behind, and pedestrians [2]. As a key technology for autonomous driving, target tracking provides a guarantee for vehicle behavior selection and path planning by accurately identifying and tracking traffic signs, vehicles in front and behind, pedestrians, etc. In the field of UAV applications, target tracking technology can also be relied on to track the target to be photographed for a long time to avoid losing focus and achieve better shooting results [3].

The image shadow preprocessing operation mainly includes two aspects: shadow detection and shadow removal. Shadow detection is often used to analyze the geometric properties of target objects in remote sensing images, such as

3D scene reconstruction in aerial images, building recognition, determination of normal direction, and light direction of object surfaces. Shadow removal can enhance the visibility of images and is also helpful to improve the recognition rate of features and image quality [4]. However, the processing of image shadows is still a relatively new research field so far, so this paper summarizes and learns the existing shadow detection and removal algorithms, based on the improvement of some algorithms, and hopes that it can have a positive reference effect on future research in this area [5]. For sports video, multiobjective motion multiobjective tracking can be divided into two types according to the overall difference in the specific number of camera devices it uses [6]. One is the simultaneous operation of multiple cameras and coordinated target tracking, and the other is the use of a single camera to achieve the tracking of the target. However, it should be noted that in the actual use of camera equipment, due to differences in specific scenarios, and mobile and fixed [7]. For many of the current algorithms for real-time simultaneous tracking of multiple targets, its size is relatively small, there is serious obscuration or poor image quality of the target, in the specific tracking effect, but not ideal, and tracking content is not comprehensive, as there is a lack of real time [8].

For each node in the target chain, it is in a continuously updated state, which is often able to reflect the current state of the target and the change of the sports video multitarget movement that has occurred, and according to the actual needs, through appropriate and reasonable deletion or insertion of nodes, it can accurately and real-time reflect the status of the new target, which is used to determine whether the target is leaving or newly entering the tracking area. There are five sections in this paper, and the details of each section are described as follows [9]. Section 1 introduces the significance and background of this paper and outlines the research content and structural arrangement of the paper. Section 2 analyzes the development of domestic and international research on sports video target tracking technology and provides a brief introduction to the related technology areas. Section 3 introduces the relevant theories of several existing shadow removal algorithms and their specific implementation steps, based on which the Wallis filtering algorithm is improved and a shadow removal algorithm based on color constancy theory is proposed. At the same time, the application scenarios and key problems of sports video multitarget motion multitarget tracking are introduced, followed by a deep neural network-based sports video multitarget motion multitarget detection algorithm for this problem, and the design and implementation of the target framing and detection principles, the principle of target center of mass segmentation, and space and frequency domain vector depiction trajectory of multicenter of the mass idea. Section 4 evaluates the accuracy of shadow rejection and accurate tracking results, compares the effect and performance of the deep neural network-based sports video multitarget motion multitarget detection and tracking technology with other detection and tracking technologies, and verifies its adaptability and anti-interference for a variety of special sports videos in several standard sports video test sets. Section 5 summarizes the work of this paper, discusses its shortcomings, and looks forward to future work.

## 2. Related Work

A larger portion of traditional tracking methods uses manually extracted target features or features that fuse several simple features to achieve target tracking. Since the features extracted by traditional target tracking methods are not comprehensive enough, they greatly limit the performance of target tracking methods [10]. Yuan et al. proposed an asymmetric twin network, which inputs the target image and the test image into two convolutional neural networks separately and extracts the feature information of the input image and then searches for the target location in the image [11]. The full convolutional twin network built by Cai et al. is to input the first image of the sports video into one neural network and then input the image of the target to be tracked into another neural network to obtain the possibility that the target is at a certain location and then get the specific location of the target [12]. Razzaq et al. used a set of image sequences with time-varying illumination and constant scene reflection coefficients, first performed median filtering on the image sequences according to the time variation, and integrated the results of the median filtering to obtain an Eigen reflected an image with the shadow removal effect [13]. Lu et al. automatically detected and removed shadows from a single image to obtain a good removal effect. However, this method has many limitations on the light source and imaging equipment, and if the texture in the image is complex, the image after shadow removal is likely to become blurred [14].

The research of shadow removal is not very deep, and the existing algorithms are not very effective in removing shadows, which need to be improved to some extent, or new removal algorithms are proposed to remove shadows efficiently and accurately. Accordingly, deep neural networks can be applied to sports video analysis techniques to extract target information, classify and frame tracking targets according to target features, and give markers [15]. Lin et al. have improved the Retina algorithm in the form of a center-surround, which is expected to achieve shadow removal. However, this algorithm needs to iterate on each pixel on the image, determine their filter scales, and compute the anisotropic spread of complex equations after principal component analysis, avoiding the selection of parameter thresholds in the elimination process [16]. Sun et al. take the Gaussian model as the basic object according to the practical application needs and incorporate some more typical prior probability parameters more reasonably [17]. Yang et al. proposed a parallax estimation network, which only generates depth maps for target candidate regions, learned the shape prior of feature categories to reduce the generation error of depth maps, and finally used depth maps to solve the target localization problem, which improves the detection. Sun proposed a parallax estimation network to generate depth maps only for the target candidate regions and learned the shape prior of the feature class to reduce the generation error [18].

The online multitarget tracking algorithm is proposed based on the constructed 3D target detection network and combined with the optimal recursive algorithm Kalman

filtering to address the problem of false and missed detection. The tracking strategy and similarity calculation method are optimized to reduce the algorithm time consumption. By generating the target trajectory, the false detection rate and missed detection rate of the detection algorithm are reduced, and the stability of the algorithm is improved. In general, this paper focuses on two main techniques: one is multitarget motion shadow rejection for multitarget sports video, and the other is multitarget motion precision tracking for multitarget sports video. The detection analysis of fixed camera includes the fast acquisition of data in sports video, collection of event information of target groups in sports video, sports video event refinement, and extraction of those targets that are more interested in tracking target features and are in sports video multitarget motion state; then according to the actual situation and needs, the relevant targets are systematically divided and each target to be tracked is targeted tracking [19, 20].

### 3. Deep Learning Multitarget Motion Shadow Rejection and Accurate Tracking for Sports Video

*3.1. Sports Video Multiobjective Network Model Construction.* This paper presents a 3D target detection algorithm based on binocular vision and Faster R-CNN. For the problem of detecting small targets at long distances, the VGG16 network is replaced by ResNet, and a multiscale feature extraction network is constructed; for the problem of poor localization accuracy of spatial targets, the projection centroid of spatial targets on the image plane is taken as the key point, and the projection centroid regression branch is established; furthermore, a 3D target frame correction algorithm based on photometric correction is proposed to complete the correction of target parameters. The improved network is named the centroid network, and its overall structure is shown in Figure 1, which consists of four parts, including the multiscale feature extraction network, the region recommendation network, the parameter regression branch, and the 3D target frame correction module.

The left and right images are input into the shared multiscale feature extraction network to get the left and right feature maps of different sizes; the left and right feature maps of each size are overlaid by the number of channels and input into the region recommendation network to generate the candidate regions where the target may exist; the features are extracted from the specified feature maps according to the region size and input the target category, border regression, projection center point, orientation angle, and physical size branch to get the target parameters, calculate the parallax from the target projection center point to get the target depth and then get the spatial coordinates according to the camera parameters, and finally, extract the valid pixel points in the 3D target frame to correct the 3D target frame according to the photometric error to get the final 3D target frame. The depth of the network is crucial for learning features with stronger expression ability, but the deeper the network is, the more serious the gradient disappearance

problem is. For this reason, the ResNet residual model is introduced instead of the VGG16 network model to solve the gradient disappearance problem when the network is deepened and to improve the network robustness and feature extraction ability.

The main path of both residual units is composed of two layers of  $1 \times 1$  convolution and one layer of  $3 \times 3$  convolution. Each layer of convolution has the effect of increasing the nonlinear expression capability of the network, and besides, controlling the number of channels of the output by  $1 \times 1$  convolution can reduce the computational effort of the network [21–23]. The difference between the two is whether the input  $3 \times 3$  of the main and bypass connections is downsampled by the convolution operation. The downsampling is done by setting the  $3 \times 3$  convolution with a step size of 2, which replaces the pooling layer required for downsampling in previous convolutional neural networks and makes the overall network more compact and regular. After improving the depth of the original network, we further improve the single-scale output of the Faster R-CNN feature extraction network, which only uses the last feature layer as the output after feeding the image into the feature extraction network. This process will lead to a significant reduction in the detection accuracy of small target objects because as the convolutional layer downsamples the input image, the size of the image will become smaller and smaller, the pixels occupied by the objects in the image will be reduced, and the proportion of small objects in the final feature map will be very small or even zero.

The network in Table 1 contains five convolutional layers with decreasing convolutional kernel size, and two GPUs are used for training to improve efficiency. The network makes the nonlinear activation function ReLU can be network training to achieve faster convergence and improve the training speed. It breaks through the limitations of traditional neural networks, the first pooling layer uses maximum pooling to avoid fuzzification effect, improve the richness of features, and use Dropout to modify the network structure, effectively preventing the neural network overfitting.

*3.2. Multitarget Sports Video Multitarget Motion Shadow Rejection and Accurate Tracking.* The basic idea of region growth segmentation is to start from a set of “seed points”, merge the neighboring pixels over the region with similar properties to these seed points to form new “seed points,” and repeat this operation until no new “seed points” meet the conditions [24–26]. The operation is repeated until no new “seed points” are generated. The similarity between “seed points” and neighboring pixels or regions can be based on image information such as grayscale value, color, or texture. In MATLAB, the region grow function is used to perform region growing segmentation with the following syntax.

$$h\{I, N, S, T\} = \text{region grow}(g, M, N), \quad (1)$$

where  $g$  is the segmented image, which converts the original RGB image into a grayscale image;  $M$  is an array or scalar of the same size as  $g$ .  $I$  is the region segmentation result image,  $N$  denotes the number of regions segmented, and  $S$  and  $T$

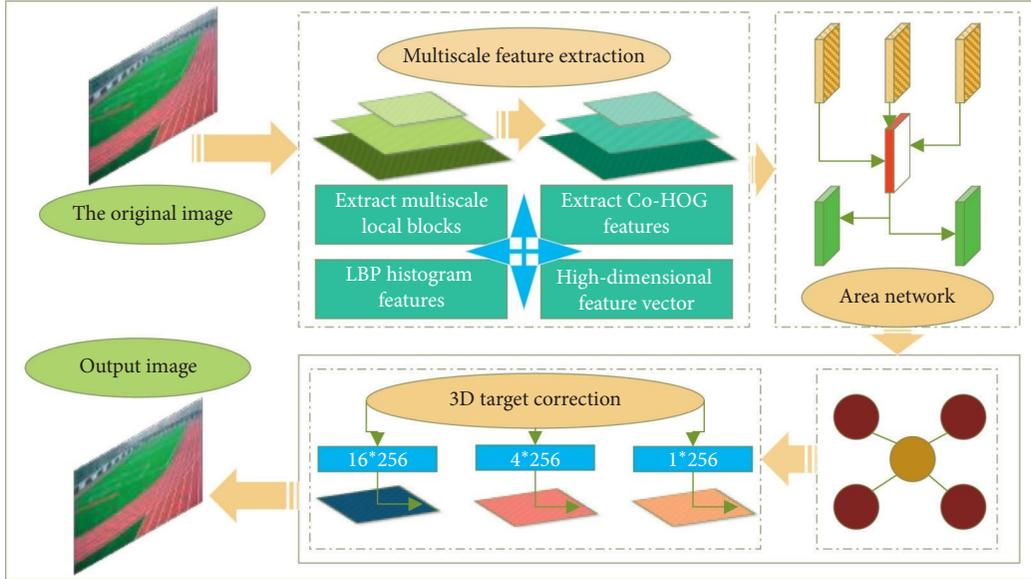


FIGURE 1: The general structure of the 3D target detection network.

TABLE 1: Network structure.

Number of layers	Convolution kernel size	Quantity	Step size
1	1*1	96	3
2	2*2	128	2
3	3*3	256	4
4	4*4	256	2
5	5*5	384	5
6	6*6	384	4

denote the images with seed points and the images that pass the connectivity processing and thresholding tests, respectively, both of which should be of the same size as  $g$ .

The grayscale image is segmented by region growth according to equation (1), and the shadow regions of the features in the image are extracted, and they are overlaid with the original image. The implementation flow is shown in Figure 2.

For deep neural networks, there are multilayer nonlinear computing units, and the output of the lower layers is input to those relatively higher layers. The output of the lower layers is input to the higher layers, and the effective features are systematically learned and represented from the various data input according to reality. For each convolutional layer, maps are the basic features of the upper layer; it should be noted that the features are convolved by a corresponding convolutional kernel capable of learning, and the corresponding output maps are obtained by activation of the relevant functions after the process is completed. The computational cost of your method is roughly 20% lower than the cost of other studies compared to other methods.

$$f(x) = h \frac{\sum_{i,j}^{\infty} (x_i^j \cdot M_{ij} + x_j^i \cdot N_{ij})}{\sum_{i,j}^{\infty} (x_i^j \cdot M_{ij} - x_j^i \cdot N_{ij})} \quad (2)$$

If the output feature mapping sets map  $i$  and map  $j$  are obtained as the sum of the convolutions in the input

mapping, then they correspond to different convolution kernels. As shown in equation (3),  $G()$  represents a downsampling function. In most cases, the basic operation is to use the input image as a basic object and then to sum all the pixels corresponding to each block.

$$x_i^j = \frac{f(\psi) \cdot G(x_i^{j-1}) + f(\phi) \cdot G(x_j^{i-1})}{2} \quad (3)$$

The detection module is the most time-consuming and important, and the multiobjective optimization of the detection module is of utmost importance. It is important to ensure the accuracy of the detection module for target detection and to reduce the time required for the detection module as much as possible [27–29]. Firstly, each target to be tracked is determined based on the sports video multitarget motion of each target using the detection region optimization algorithm from the previous chapter and corrected for prediction separately. Then, the subimage blocks in the subwindow of the intersection part of the detection region with each target are classified and detected using pixel difference classifier, integrated classifier, and nearest neighbor classifier.

During the tracking process, the target will change to different degrees, and the correlation filter, as an online tracking model, needs to continuously introduce new samples for model learning according to the target changes to ensure adapting to the target changes and accurately

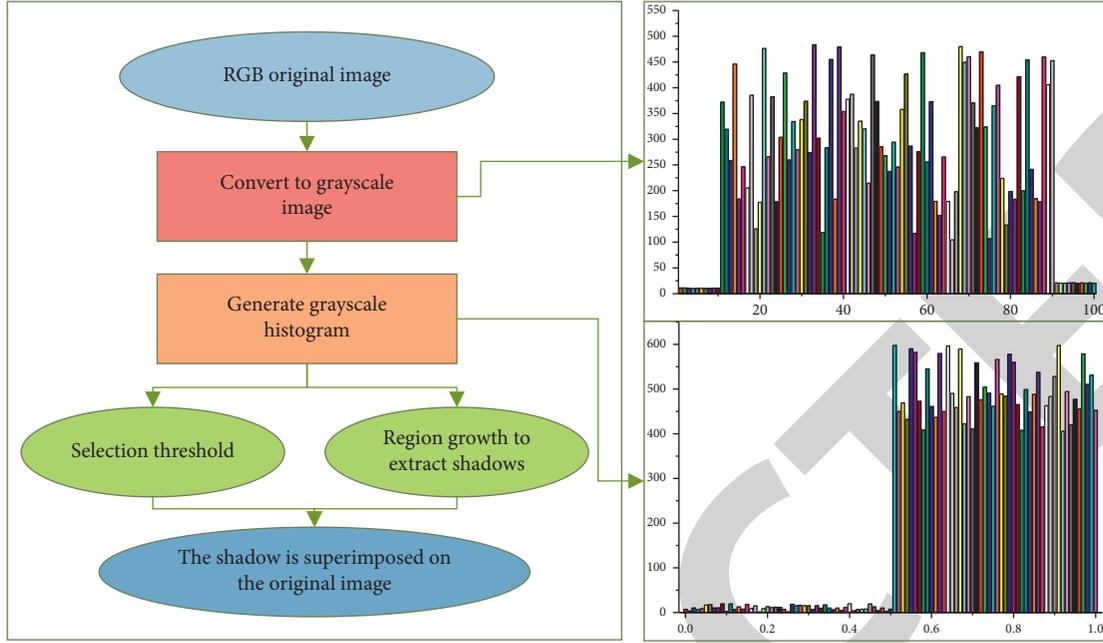


FIGURE 2: Area growth method segmentation.

tracking the target. If the model is updated every frame, it is easy to cause model overfitting and too much computation. Therefore, this chapter chooses to update the correlation filter model every two frames, and the model is updated as in equation (4), where  $\alpha$  is the frame sequence,  $\beta$  is the number of feature map channels, and  $i$  is the learning rate, which indicates the learning ability of the target appearance to new frame images.

$$\begin{cases} M_a = (1 + \alpha\beta) \cdot M_{a-1} + \frac{1 + \beta}{1 - \beta} \cdot X_a + (1 + \alpha) \cdot X_{a-2}, \\ N_a = (1 + \alpha\beta) \cdot N_{a-1} + \frac{1 + \beta}{1 - \beta} \cdot X_a + (1 + \alpha) \cdot X_{a-2}, \\ F_a = \sum_i^j \frac{M_a + x_i}{N_a + x_i} * (1 + \alpha) * (1 + \beta). \end{cases} \quad (4)$$

The optimization objectives and constraints of SVM are as follows:

$$\text{SVM} = \sum_{i=1}^M \psi_i + \frac{1 + \beta}{1 - \beta} \sum_{j=1}^N \varphi_j. \quad (5)$$

In the discrete-time linear state space, the state noise and observation noise of the system are both Gaussian white noises, and the state noise and observation noise of the system do not change when the system state changes, as shown in equation (6).  $M$  and  $N$  denote the state transfer matrix and observation matrix, respectively,  $A_t$  denotes the state vector,  $m$  is the state mean,  $M_t$  is the covariance, and  $B_t$  denotes the target observation vector.

$$\begin{cases} A_t = XA_{t-1} + M_t, \\ B_t = Xb_{t-1} + N_t. \end{cases} \quad (6)$$

To realize the cyclic calculation of the filter, the covariance of the system also needs to be updated, as shown in equation (7), and  $Q(t)$  denotes the updated result of the state covariance at time  $t$ , the covariance of the state optimum. Conventional Kalman filtering requires constant target state estimation and updates when target tracking is performed.

$$Q(t) = \frac{(1 + \alpha\beta)Q(t + 1)}{(1 - \alpha\beta)Q(t - 1)}. \quad (7)$$

**3.3. Data Set and Evaluation Indexes.** To evaluate the performance of tracking algorithms under the influence of various factors, several important datasets have been formed in the field of target tracking, which is applied to test various types of tracking algorithms, namely, OTB dataset, VOT dataset, Tracking Net dataset, and Temple Color 128 datasets. In this paper, the OTB dataset is chosen for the study, and Figure 3 shows the tracking sequences used for evaluation in the OTB.

The OTB dataset is mainly evaluated for tracking algorithms in terms of Center Location Error (CLE), Overlap Precision (OP), and Distance Precision (DP). The Center Location Error is the Euclidean distance between the predicted target center and the true target center before the center of the target in pixels, such that  $(m_x, m_y)$  denotes the coordinates of the predicted location of the target and  $(n_x, n_y)$  denotes the true coordinates of the target, calculated as follows:

$$\text{OTB}(m, n) = \iint \sqrt{\frac{(m_x - m_y)^3 + (n_x - n_y)^3}{(m_x - m_y) - (n_x - n_y)}} dx dy. \quad (8)$$

The distance accuracy refers to the percentage of frames whose center position error is below a certain threshold



FIGURE 3: Trace sequence for evaluation in OTB.

(usually set to 20 pixels) to the total number of frames in the sports video sequence; the overlap accuracy refers to the percentage of frames whose overlap ratio between the predicted target bounding box and the real target bounding box is higher than a given threshold (usually set to 0.5) to the total number of frames in the sports video sequence, and the overlap ratio is calculated by equation (9), where  $S_x$  and  $S_y$  denote the target prediction bounding box and the true bounding box, respectively. When performing algorithm evaluation, the larger values of distance accuracy and overlap accuracy indicate the more frames being effectively tracked and the better tracking results.

$$\text{rate}(x, y) = \frac{S_x \otimes S_y}{S_x \odot S_y} + \frac{\lambda}{1 - \lambda} \text{OTB}(m, n). \quad (9)$$

Under the OTB evaluation criteria, robustness evaluation of the algorithm is usually done; i.e., the tracking algorithm is initialized based on the target position given in the first frame of the sports video sequence, and then, the algorithm is run in the test sequence to derive the tracking accuracy or success rate, called one-time pass evaluation. Besides, to address the problem that tracking algorithms may be sensitive to initialization, two methods are proposed to analyze the robustness of tracking algorithms to initialization, namely, temporal robustness evaluation and spatial robustness evaluation to evaluate the performance of tracking algorithms by giving different initializations at different initial frames for better or worse performance. Temporal robustness evaluation refers to dividing the sports video sequence into several segments, initializing them from different frames, evaluating the tracking algorithm at each segment, and computing the overall information.

## 4. Analysis of Results

*4.1. Simulation Analysis of the Multiobjective Motion of Sports Video.* We extract some data from the relevant dataset as the

validation set to ensure that the validation set is convincing with a wide range of sources and diversity. The experimental code uses the Python framework and an NVIDIA GTX 2020TI GPU. The base learning rate of the feature branching experiment training phase is 0.006. During the experiment, the total epoch number is 50, the learning rate impulse is 0.95, and the weight decay is 0.0002. As shown in Figure 4, since the detection module of our model is the same as the process of the traditional Faster R-CNN algorithm, with improvements only in the feature fusion module and parameter settings, the performance of the detection module in the comprehensive analysis should be close to that of Faster R-CNN, and it can be seen that our algorithm is the highest in the value of AP = 0.8, which is close to Fast R-CNN and Yolov3-based JDE. The performance of the detector is still at a high level. The entry accuracy of the model is guaranteed, and of course, our framework is specifically trained with parameters obtained from the pedestrian detection dataset, and the performance is more qualitative than the detector for multiclass targets. Our results show an approximate 5% increase in accuracy and a 10% increase in efficiency compared to other studies.

Figure 5 lists the more challenging segments in the MOT dataset to show the FPS runtime speed of various algorithms. Comparing MOT16-11 and MOT20-08, we find that the speed of many algorithms drops abruptly, only our method and JDE performance does not drop much because MOT20-08 is a test sequence of ultradense crowd shortly after public where the number of people in a single frame exceeds 150, resulting in a precipitous drop in performance for many algorithms, while our algorithm can maintain good stability. The detection model is first used to locate the location of the target's bounding box in the image, and then, the association model is used to extract Reidentification (Re-ID) features for each bounding box and associate the bounding box with an existing tracking result based on a specific metric defined by these features. Where the target detection in the detection model is to discover all the targets in the current frame, the

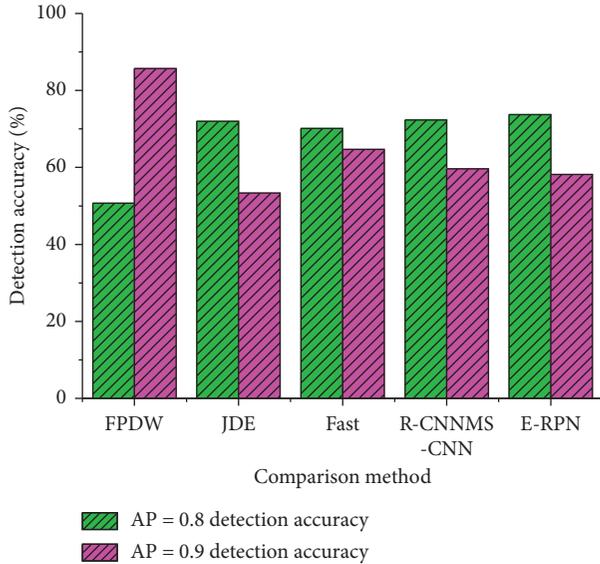


FIGURE 4: Experimental comparison of multiobjective tracking algorithms.

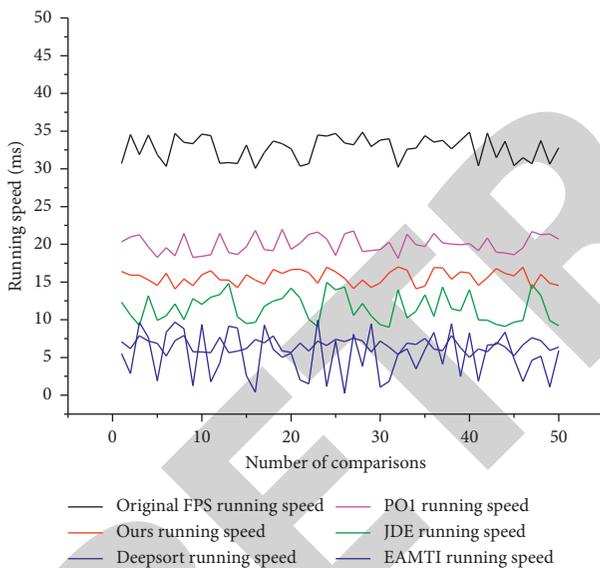


FIGURE 5: Running speed comparison results.

Re-ID is to associate all the current targets with the targets of the previous frames, which can then be associated by distance comparison of the Re-ID feature vectors and the target area intersection ratio (IOU) by using the Kalman filter and the Hungarian algorithm. The advantage of the two-step approach is that they can use the most appropriate model for each task separately without making trade-offs. In addition, they can crop the image patches according to the detected bounding boxes and resize them to the same size before predicting the Re-ID function, which helps to handle the scale changes of the objects.

Using the above metrics and datasets, we conduct comparative experiments with related multitarget tracking algorithms to highlight the advantages of the E-RPN network, which reduces the inference time and makes real-time

multitarget tracking more possible without compromising accuracy. Through the work in this chapter, we also found the problem of E-RPN, where there is a misalignment between the anchor box and the target area, specifically, the perceptual field of each pixel point on the feature map is the same, but the anchor box size is different, and it is difficult to predict the features of different targets using the same feature vector.

*4.2. Performance Analysis of Shadow Rejection and Accurate Tracking Algorithm.* In this section, a high-resolution sports video with obvious shadow features is selected, and all operations such as detection and extraction of multiple sports video multitarget motion target shadows in the image and removal of sports video multitarget motion target shadows are completely realized. The statistical results of the shadow removal algorithm for HSI color space analysis of the experimental images are shown in Figure 6. The shadow removal algorithm is greatly affected by the shadow detection accuracy, and the nonshadowed areas that are mistakenly detected as shadow areas will participate in the algorithm and be removed, so they must be satisfied simultaneously and combined to realize the complete shadow rejection of high-resolution remote sensing images. We validated VOT2018 and found that our results can be applied as well.

The speedup between the detection of the deep neural network-based sports video multitarget motion multitarget tracking algorithm and the time used by the detection-tracking-self-learning tracking algorithm is shown in Figure 7. From Figure 7, the detection time of the deep neural network-based sports video multitarget motion multitarget detection algorithm in the six sports videos is much shorter than the time used by the detection-tracking-self-learning tracking algorithm alone. The detection speed of the deep neural network-based sports video multitarget motion multitarget detection algorithm is much faster than that of the detection-tracking-self-learning tracking algorithm.

This paper proposes a neural network target tracking algorithm fused with a redetection mechanism. It is improved based on the existing algorithm and adds a redetection module. When the current tracking effect is judged by analyzing the response graph, the detector is called to reidentify targets to increase the accuracy of tracking. At the same time, it is different from the existing algorithm that only uses the first frame target as a template to track the target offline and adopts a high-confidence model update strategy to avoid model pollution. The improved algorithm largely solves the problem of poor robustness and easy target loss of existing algorithms when faced with occlusion and interference from similar objects. The comparison between the OTB2016 data set and several representative algorithms proves that the improved algorithm can significantly improve the tracking effect while meeting real-time requirements.

*4.3. Analysis of Shadow Rejection and Precise Tracking Evaluation.* After analyzing the validation set, this paper uses all training sequences for training and tests all trained

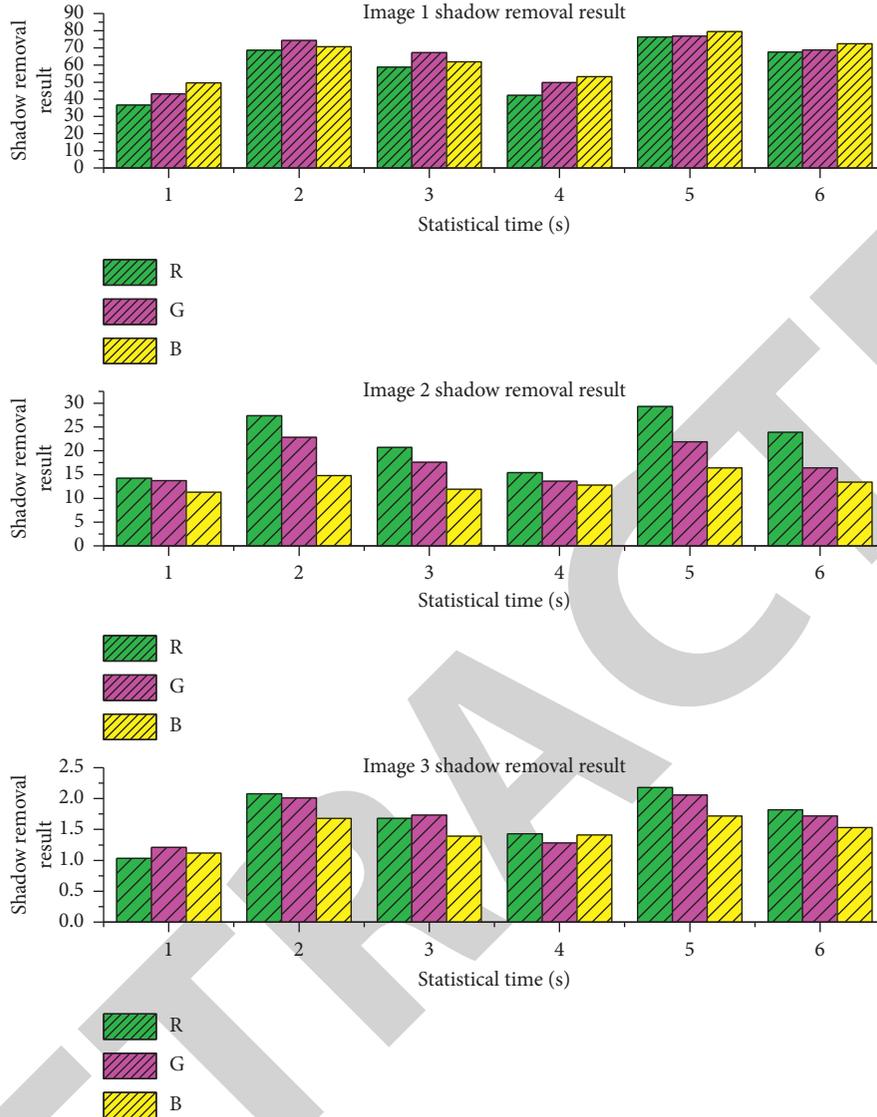


FIGURE 6: Statistical results before and after removal of shadows.

trackers in the test set, where 20 data-associated templates are used. The results are submitted to the multiobjective tracking Benchmark for evaluation, and Figure 8 shows the tracking performance on the test set. In this paper, we compare our algorithm with several of the current mainstream multitarget tracking algorithms. As can be seen in Figure 8, our algorithm improves 7.5% over the second-ranked algorithm in the MOTA metric and achieves the best results in metrics MT and ML. Such excellent results demonstrate the advantages of deep learning-based multi-objective tracking algorithms.

Real-time detection and tracking of motion targets are a fundamental step in intelligent surveillance and video activity recognition applications. Motion target detection segments the scene into foreground targets and background regions, but the shadows cast by motion targets are easily misclassified as foreground targets in this process, and this misclassification can cause the merging of multiple targets or the change of target shapes. To improve the effect of motion

target segmentation, an algorithm for real-time shadow detection and elimination based on the light intensity, chromaticity, and reflectance is proposed, which does not require a priori knowledge of the target's features, illumination conditions of the scene, etc. Simulation results show that this algorithm has better performance than other methods. The test evaluation results of the three algorithms are shown in Figure 9. From Figure 9, MOTA, MOTP, MT, and ML indicate that this paper's algorithm outperforms the other two algorithms. The results of target sports video multitarget motion direction prediction and quantity recognition are better than the other two algorithms.

This section compares the multimass center of this algorithm for each mass center with TC-ODAL and RMOT methods, showing higher tracking accuracy and tracking accuracy. The reason for this is that the corresponding grayscale characteristics of the target in the foreground region during the specific deformation phase are mastered; for this characteristic, it is a typical feature that differentiates

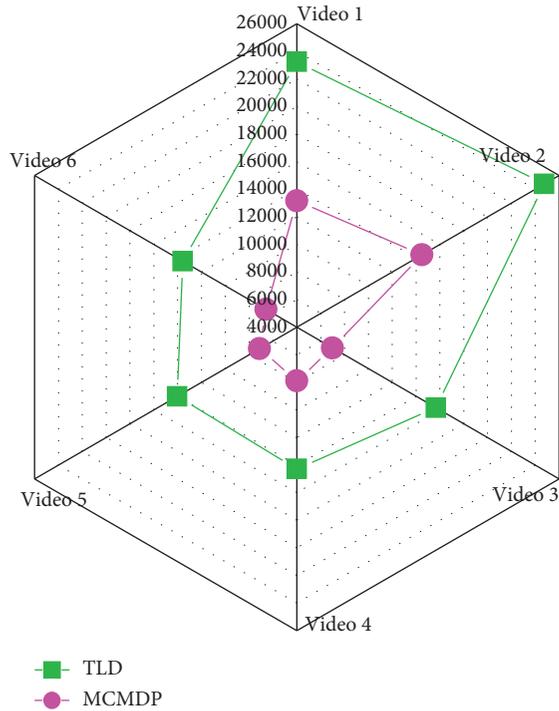


FIGURE 7: Tracking time comparison.

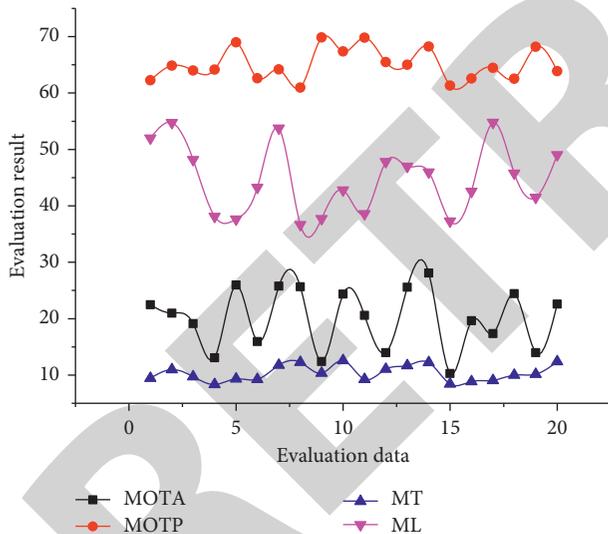


FIGURE 8: Test evaluation results.

and differs between the sports video multitarget motion target and the complex background. In this paper, by establishing the target edge spectral band and the spatial-spectral band as well as the feature constraint model, we realize the effective extraction of the same-frequency amount in the target foreground region. It highlights the difference between the sports video multitarget motion of target heel and background and the important region of the target, suppresses the interference region, establishes the cofrequency amount constraint model, extracts the impact intensity of target sports video multitarget motion, and predicts the multitarget foreground, which solves the

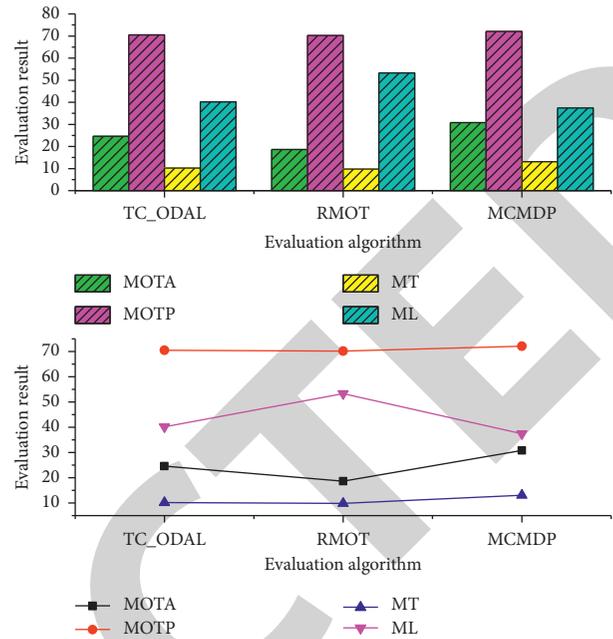


FIGURE 9: Evaluation results of the three methods of testing.

problem of target misfollowing caused by mutual interference and distorted deformation. Meanwhile, compared with the other three methods, this algorithm can be applied to the target foreground by cofrequency amount extraction, target fore-trend. At the same time, compared with the other three methods, this algorithm implements target tracking through four steps: same-frequency extraction, target pretrend prediction, and target localization and tracking, highlighting the strong advantages of practical robustness and spatial robustness.

For the Benchmark performance evaluation system, the sports video data is more objective and comprehensive in testing the target tracking algorithm under the combined influence of various factors (such as light intensity change, shape change magnitude, feature domain rotation, scale expansion intensity, and similar background interference), and the whole evaluation process is also more comprehensive and rigorous. Besides, the whole evaluation process is more comprehensive and rigorous, so that the obtained sports video data and environment will be more suitable and like the real situation. Therefore, the tracking method proposed in this paper has strong advantages and good tracking performance in terms of accuracy, precision, real-time, and robustness.

## 5. Conclusion

In this paper, we propose a correlation filtering tracking algorithm that combines deep features and underlying features of a convolutional neural network to provide better tracking performance than using only single-layer convolutional features, considering that deep features and underlying features of a convolutional neural network each have advantages in the target tracking task. The multilayer convolutional features of the target are extracted by a

pretrained network model, and the response maps are fused to localize the target. A feature selection algorithm is also used to reduce the dimensionality of the extracted convolutional features to reduce the computational effort and improve the real-time performance of the algorithm. The existing shadow removal algorithms are analyzed and summarized, the classical Wallis filtering algorithm is improved, and a shadow removal algorithm combined with color constancy theory is proposed. It is verified through experiments that both algorithms can better achieve the removal of building shadows in high-resolution remote sensing images, and the removal accuracy is higher than the existing algorithms. Simulations are conducted on the sports video training set and test set for the deep neural network-based multitarget motion multitarget tracking algorithm, and the performance of the algorithm in different sports video sequences is analyzed using the positive and negative sample-based methods. The algorithm meets the requirements of accuracy and robustness of the target tracking algorithm in the case of stable sports videos. The algorithm of multitarget motion detection and tracking of sports video based on a deep neural network is further upgraded to continuously improve the ability of target identification and anti-interference of this paper's algorithm, which is a wider and more valuable application area of this topic.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The author declares no conflicts of interest.

## References

- [1] X. X. Zhang and H. K. Xu, "An efficient and flexible approach for multiple vehicle tracking in the aerial video sequence," *International Journal of Remote Sensing*, vol. 39, no. 19, pp. 6234–6265, 2018.
- [2] E. U. Haq, X. Huarong, C. Xuhui et al., "A fast hybrid computer vision technique for real-time embedded bus passenger flow calculation through camera," *Multimedia Tools and Applications*, vol. 79, no. 1, pp. 1007–1036, 2020.
- [3] Y. Chen, J. Wang, Y. Bai et al., "Probabilistic semantic retrieval for surveillance videos with activity graphs," *IEEE Transactions on Multimedia*, vol. 21, no. 3, pp. 704–716, 2018.
- [4] Z. Boukhers, K. Shirahama, and M. Grzegorzec, "Less restrictive camera odometry estimation from monocular camera," *Multimedia Tools and Applications*, vol. 77, no. 13, pp. 16199–16222, 2018.
- [5] Y. Hu, "Image segmentation based on velocity feature vector for moving target extraction," *IEEE Sensors Journal*, vol. 20, no. 20, pp. 11983–11991, 2020.
- [6] D. Zhou, S. Qiu, N. Li et al., "Quadruped animal moving target extraction and background reconstruction algorithm," *Revista Científica de la Facultad de Ciencias Veterinarias*, vol. 29, no. 5, pp. 1416–1425, 2019.
- [7] D. Israni and H. Mewada, "Identity retention of multiple objects under extreme occlusion scenarios using feature descriptors," *Journal of Communications Software and Systems*, vol. 14, no. 4, pp. 290–301, 2018.
- [8] O. W. Ibraheem, A. Irwansyah, J. Hagemeyer et al., "FPGA-based vision processing system for automatic online player tracking in indoor sports," *Journal of Signal Processing Systems*, vol. 91, no. 7, pp. 703–729, 2019.
- [9] Y. Zhang, M. Zhang, Y. Cui et al., "Detection and tracking of human track and field motion targets based on deep learning," *Multimedia Tools and Applications*, vol. 79, no. 13, pp. 9543–9563, 2020.
- [10] S. H. Kim and H. L. Choi, "Convolutional neural network for monocular vision-based multi-target tracking," *International Journal of Control, Automation and Systems*, vol. 17, no. 9, pp. 2284–2296, 2019.
- [11] D. Yuan, X. Shu, and Z. He, "TRBACF: learning temporal regularized correlation filters for high performance online visual object tracking," *Journal of Visual Communication and Image Representation*, vol. 72, 102882 pages, 2020.
- [12] C. Cai, K. Wu, and Y. Yan, "Rapid detection and social media supervision of runway incursion based on deep learning," *International Journal of Innovative Computing and Applications*, vol. 9, no. 2, pp. 98–106, 2018.
- [13] M. A. Razzaq, J. M. Quero, I. Cleland et al., "uMoDT: an unobtrusive multi-occupant detection and tracking using robust Kalman filter for real-time activity recognition," *Multimedia Systems*, vol. 26, no. 5, pp. 553–569, 2020.
- [14] W. Lu, Z. Zhou, L. Zhang et al., "Multi-target tracking by non-linear motion patterns based on hierarchical network flows," *Multimedia Systems*, vol. 25, no. 4, pp. 383–394, 2019.
- [15] H. Yu, G. Li, W. Zhang et al., "The unmanned aerial vehicle benchmark: object detection, tracking and baseline," *International Journal of Computer Vision*, vol. 128, no. 5, pp. 1141–1159, 2020.
- [16] X. Lin, Y. Ding, X. Xu et al., "A multi-target detection algorithm using high-order differential equation," *IEEE Sensors Journal*, vol. 19, no. 13, pp. 5062–5069, 2019.
- [17] Y. Sun, J. Cheng, G. Zhang et al., "Mapless motion planning system for an autonomous underwater vehicle using policy gradient-based deep reinforcement learning," *Journal of Intelligent & Robotic Systems*, vol. 96, no. 3, pp. 591–601, 2019.
- [18] J. Yang, W. Zhao, Y. Han et al., "Aircraft tracking based on fully conventional network and Kalman filter," *IET Image Processing*, vol. 13, no. 8, pp. 1259–1265, 2019.
- [19] Y. Liu, L. Yao, W. Xiong et al., "Ship target tracking based on a low-resolution optical satellite in geostationary orbit," *International Journal of Remote Sensing*, vol. 39, no. 9, pp. 2991–3009, 2018.
- [20] M. C. Chen, S. Q. Lu, and Q. L. Liu, "Uniqueness of weak solutions to a Keller-Segel-Navier-Stokes Model with a logistic source," *Applications of Mathematics*, vol. 121, 2021.
- [21] S. A. Vahora and N. C. Chauhan, "Deep neural network model for group activity recognition using contextual relationship," *Engineering Science and Technology, an International Journal*, vol. 22, no. 1, pp. 47–54, 2019.
- [22] Y. Wang, H. Mao, and Z. Yi, "Stem cell motion-tracking by using deep neural networks with multi-output," *Neural Computing and Applications*, vol. 31, no. 8, pp. 3455–3467, 2019.
- [23] L. He, G. Liu, G. Tian et al., "Efficient multi-view multi-target tracking using a distributed camera network," *IEEE Sensors Journal*, vol. 20, no. 4, pp. 2056–2063, 2019.
- [24] B. Huang, T. Xu, B. Liu, and B. Yuan, "Context constraint and pattern memory for long-term correlation tracking," *Neurocomputing*, vol. 377, pp. 1–15, 2019.

- [25] B. Wang, L. Zhang, H. Ma, H. Wang, and S. Wan, "Parallel LSTM-based regional integrated energy system multienergy source-load information interactive energy prediction," *Complexity*, vol. 2019, Article ID 7414318, 13 pages, 2019.
- [26] W. Wei, Q. Ke, J. Nowak et al., "Accurate and fast URL phishing detector: a convolutional neural network approach," *Computer Networks*, vol. 178, 107275 pages, 2020.
- [27] S. Yang, T. Gao, J. Wang et al., "Efficient spike-driven learning with dendritic event-based processing," *Frontiers in Neuroscience*, vol. 15, 97 pages, 2021.
- [28] J. Yang, T. Liu, B. Jiang, W. Lu, and Q. Meng, "Panoramic video quality assessment based on non-local spherical CNN," *IEEE Transactions on Multimedia*, vol. 23, pp. 797–809, 2020.
- [29] A. Zielonka, A. Sikora, M. Woźniak, W. Wei, Q. Ke, and Z. Bai, "Intelligent Internet-of-Things system for smart home optimal convection," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 4308–4317, 2021.