WILEY | Hindawi

Research Article

The Role of 5G Network Image Information Based on Deep Learning in UAV Prediction Target Trajectory Tracking

Cundong Tang [iD],[1] Li Chen [iD],[1] Yi Wang,[1] Wusi Yang,[2] Rui Chen,[3] and Zhiping Wang[4]

[1]*College of Information Science and Technology, Northwest University, Xi'an, 710127 Shaanxi, China*
[2]*College of Computer Science, Xianyang Normal University, Xi'an, 712099 Shaanxi, China*
[3]*School of Software, Zhengzhou University of Light Industry, Zhengzhou, 123456 Henan, China*
[4]*School of Computer and Software, Nanyang Institute of Technology, Nanyang, 473004 Henan, China*

Correspondence should be addressed to Cundong Tang; 3021054@nyist.edu.cn

With the development of information technology in the network era and the popularization of the 5G era, UAV-related applications are becoming more and more widely used, which is one of the essential basic technologies. Therefore, the technology has great research value and practical significance, a multiobjective detector based on support vector machine (SVM) is designed based on directional gradient histogram (HOG), and the startup method used with cross-validation methods can improve detector performance. It makes the detector accuracy above 98% and has good resistance to the target scale. A real-time target tracker is designed with its rotation variation and with an improved average displacement algorithm. The algorithm must manually select the target model and suggest the target model to achieve automatic acquisition of the target model. Due to the ambiguity of the target tracking state, several judgment conditions are set to determine whether the tracking has failed and whether the tracker state is correctly verified, with several similar target tracking algorithms. When the system is started, the system detects targets frame by frame. And it will locate a possible target by color segmentation and specify the target to be tracked to recommend the relevant model during the tracking process and open the tracker to determine the target tracking state frame by frame and perform target detection at each frame. Then it will find possible goals and will follow them to achieve a balance of stable and real-time system performance, using the results of the TPD-KCF method. The percentage of correctly tracking images can reach 98%, and the efficiency is significantly improved.

1. Introduction

Since the beginning of the new century, the world economy has basically shown a stable development trend, which has greatly promoted the development of various technologies, the most important of which is multimedia processing technology. Regarding intelligent video surveillance technology and computer communication technology, especially after the emergence of 5G technology, people have made more significant progress in the field of network image data. Unmanned air vehicles originated in the 1920s and have developed significantly after the 1950s. After entering the 21st century, air robotics technology has generally developed. Its efficiency has gradually improved, and it has rapidly developed in the direction of miniaturization, intelligence, and stealth. The concept of "air robot" was first proposed by Robert Michelson of Georgia Institute of Technology in the United States. Tracking several targets and predicting the trajectory of a target in the same way in computer vision and artificial intelligence are also some of the main ideas. At present, the research topics in the field of patrol inspection include traffic violations control, prison inspection, and testing in sensitive areas. Another breakthrough is the application of computer vision target detection and tracking technology to drones. Thanks to the flexible rotor drone, it can monitor and track targets through visibility in the air, for example, by monitoring traffic. For disaster relief, more information can be found here. Therefore, the use of UAV

platforms to detect and track targets is the center of today's research and application. It is also the inevitable trend of future development.

UAV is a remote controlling aircraft, also known as aerial robot or drone. It is mainly composed of three parts: the ground control system. Aircraft platform systems and weighing systems have gradually expanded with the continuous development of UAV technology. (1) In the military, UAVs can be used to detect targets and observe battlefields. The most famous ones are the American "global hawk," "fire scout," and" predator"; (2) in agriculture, drones are used for crop protection, planting, fertilization, pesticide spraying, production assessment, disaster warning, etc.; and (3) in terms of forest fire prevention, drones are used to monitor remote forest fires and track and predict fires in fire areas. The aerial view allows you to travel and play videos, photos, etc. One of the tasks of the UAV is to track and predict the trajectories of several similar target robots on the ground to provide a basis for autonomous decision-making. The UAV provides information support for the system, including moving target detection and classification tracking, which are widely used in aerospace and other fields. Therefore, similar multiground target tracking and quadrotor UAV trajectory prediction are essential to accomplish this task. And this is one of the essential basic technologies. Therefore, this technology has great research value and practical significance.

With the arrival of 5G network data, UAV forecasting and tracking hotspots will once again become popular. Kuechle et al. relies on the behind-the-scenes task of the 7th International Aerial Robotics Competition (IARC). This problem was studied on the quadrotor UAV as the platform, which is the basis for predicting the target trajectory. But this algorithm is difficult to apply to events [1]. Xu et al. address the problem of deviation in the target tracking process. A moving target tracking algorithm combining moving average and trajectory prediction is proposed. First, the algorithm uses the least square method to adjust the trajectory according to the known target position data to obtain the projection position, then uses the average offset algorithm to get the final position of the target. And this principle is limited to competition and theory [2]. However, it is impossible to predict and judge the target only based on the theory of the least square method, and it needs practice and algorithm optimization. Long uses the Kalman algorithm to study the tracking of moving targets and proposes a tracking algorithm based on the Kalman prediction. First, the Kalman prediction is used to narrow the search range. Match and link detection results create reliable short path segments and then use Kalman to predict each fragment of repeated links to create a set of target tracking paths, but the algorithm is complex and difficult to use [3]. Li et al. focus on the trajectory planning of UAV tracking targets and propose an improved A* algorithm with a dual evaluation function. Fuel consumption runway length and aircraft maneuverability are used to design intermediate target point evaluation functions and tracking segment evaluation functions, but this algorithm is difficult to follow in the 5G era [4]. Liu et al. only address the tracking control problem of relative distance data between targets and the wearer. And there is no speed data and the expected viewing angle requirements of the quadrotor UAV under mission conditions. The design purpose of the position and attitude controller is that the effect will not be too bright in practice [5]. Liang mainly researches UAV target recognition and tracking technology. It will analyze image integration technology and its advantages. It also explains the different algorithms of image recognition and positioning systems, using the principles of data sharing and recording algorithms and merging algorithms based on wavelet transform. This principle is of little significance in the prediction and monitoring of UAVs [6]. Araniti et al. solve the problem of measuring the speed of unmanned aerial vehicles on ground moving targets. Therefore, a method using visible light is proposed to measure the speed of UAV ground targets. A target positioning model is provided, which details the principle of ground target speed measurement. Although the speed of the continuous shooting target speed correction algorithm is improved, the stability is greatly reduced [7].

In this paper, the multitarget detector based on support vector machine (SVM) is designed based on the histogram of directional gradient (HOG), and the startup method used together with the cross-validation method can improve the performance of the detector. The accuracy of the detector is more than 98%, and it has good resistance to the target scale. A real-time target tracker is designed by using its rotation change and an improved average displacement algorithm. The algorithm must manually select the target model and suggest the target model to achieve automatic acquisition of the target model. However, due to the ambiguity of the target tracking state, several judgment conditions are set to determine whether the tracking fails and whether the tracker state is correctly verified. When the system starts, the system will detect the target frame by frame, find a possible target through color segmentation, and specify the target to be tracked. During the tracking process, the relevant model is recommended, and the tracker is turned on to determine the target tracking status frame by frame, and target detection is performed in each frame.

2. Theoretical Analysis and Methods

2.1. Introduction to UAV System. Unmanned aircraft is abbreviated as "unmanned aerial vehicle" and abbreviated as "UAV" in English. It is an unmanned aircraft operated by radio remote control equipment and self-provided program control device, or it is completely or intermittently operated by onboard computer autonomously. Before introducing the visual system, it is necessary to understand the overall structure of the UAV, mainly composed of power supply, motors and blades, sensors, translation/tilt control, remote control, and ground stations. The physical diagram of the UAV and its overall mechanical function diagram are shown in Figures 1 and 2.

Definitions of terms are as follows: *Power* mainly refers to the battery, used to power drones. *Motor and propeller*: the motor is driven by the power supply to drive the propeller to rotate and provide traction for the UAV. *Sensors*
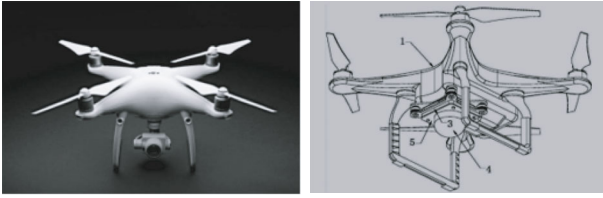
FIGURE 1: Legend of UAV.

mostly include accelerometers, gyroscopes, GPS, and other sensors, obtaining information controller such as air pressure data and geographic location of the aircraft: the central processing unit (MCU) of the UAV is mainly responsible for analyzing the data from each sensor. For example, the engine speed can be obtained by analyzing engine data. And the geographic location can be obtained by analyzing GPS data.

Definitions of terms are as follows: *PTZ*: install the camera, reduce vibration, and reduce camera shake during shooting. *Remote control*: control drone flight through wireless signal. *Ground station* monitors and receives various UAV data for data analysis. The above is the overall control module of the quadrotor UAV used in the machine vision system, which is mainly composed of PTZ, image acquisition card, camera, etc. When using a machine vision system to track a moving target, the flight direction of the gimbal and drone is adjusted on the screen to achieve the tracking function. Figure 3 shows a block diagram of the UAV machine vision system [8–10].

(1) PTZ motor: the main function of the PTZ motor is to expand the field of view of the camera. Using the fixed structure of the PTZ to drive, the motor will rotate within the specified range, which will increase the field of view of the camera

(2) Cloud platform structure: cloud platform operation has two points: one is to install the CMOS camera. Secondly, the physical structure of the cloud platform is used to reduce camera shake and improve the picture quality. Of course, there are various effects of auxiliary cameras, varying from platform to platform. The DJI Genie UAV uses a three-axis cloud platform to reduce vibration and multiple angles. The axial cloud platform used in this paper achieves experimental purposes

(3) The CMOS (complementary metal-oxide semiconductor) camera uses the currents generated by two complementary effects, interprets them in the image, and displays it on the chip to obtain the image. Compared with CCD cameras, CMOS has the advantages of low production cost, high speed, and high plasticity. It is difficult for CCD cameras to increase pixels. Under normal circumstances, they can only achieve a resolution of 6 million through video target tracking technology. The image contains various target data, which affects the extraction of target attributes.

And to a certain extent, it affects the performance of the tracking algorithm

(4) This item selects a macrocrystalline IMX179 image sensor as the camera chip. Excellent performance in dark capacity power consumption and shooting speed, the photosensitive area is 1/3.2 inches, the pixel area is 1.4 m, and the preview frame rate is 30 FPS. The camera's field of view (FOV) is 100° horizontally, 78.5° vertically, and 120° diagonally, indicating that the camera has a wider field of view. The TV distortion is -13.6%, the relative brightness is greater than 47.4%, and the main beam angle is less than 31° [11–13]. So, after introducing the operating principles of UAVs, the algorithm and principles of UAV target prediction will be introduced next.

2.2. Target Detection Algorithm Based on Deep Learning. This is because the video clip consists of a single frame. From the perspective of the accuracy and processing speed of the detection results, this is achieved by detecting targets in a single image, such as supervised or semisupervised targets. To check, we try to use efficient algorithms to ensure calculation speed. In the field of computer vision, one of the most classic methods of this type of detection algorithm is the YOLO fast detection algorithm, which directly selects all images to train the model, which no longer use the sliding window method or the separation candidate frame method. This makes the distinction between target and background easier. And the detection speed is also greatly improved. Compared with Faster R-CNN, YOLO has obvious advantages. First of all, Faster R-CNN requires RPN network instead of selective search to find subscription areas, while YOLO uses 49 areas in 7×7 as direct subscription areas. YOLO simplifies the entire target detection process, and the speed is greatly improved, but the learning process is still relatively time-consuming. YOLO has many aspects that can be improved, for example, the $S \times S$ table is a heuristic strategy. If the two targets of the small target fall in the same grid at the same time, another problem is the loss function of YOLO. Even if the square root method is used, the large target error and the small target error contribute to the learning loss function of the near-value network. This error will optimize the network. At the same time, this leads to greater impact and lower detection accuracy, and due to the use of multiple sampling layers, the YOLO network cannot learn the properties of the correct object. It ultimately affects the test results [14]. Convolutional neural networks (CNN) are a type of feedforward neural network that includes convolution calculations and has a deep structure. It is one of the representative algorithms of deep learning.

For example, suppose $\{x_i, t_i\}$, $i = 1, \cdots, n$, is the information set collected through training, and n is the number of simulated samples, $x_i \in R_d$ is the extracted density and features, which is the feature of dense trajectory, and based on the deep learning features of motion information, $t_i \in R_d$ is the real action collection class, where q is the total number of classification calculations for all training samples.
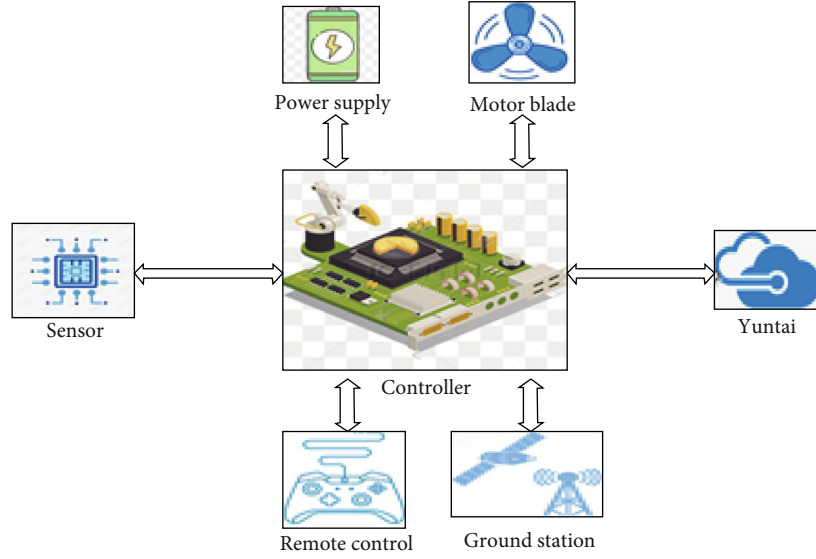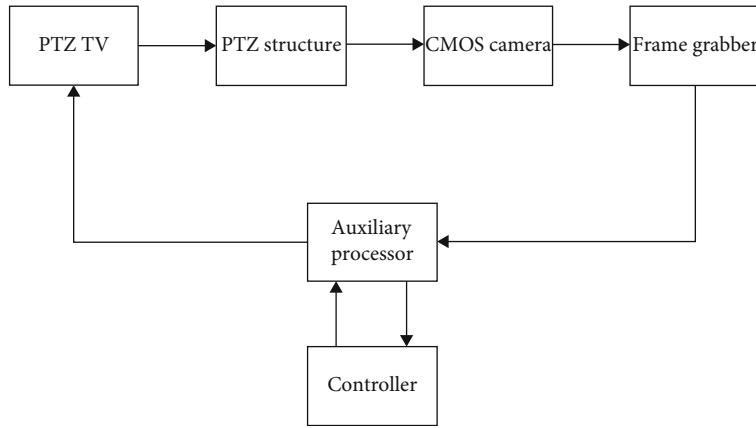
FIGURE 2: UAV structure block diagram.

FIGURE 3: Block diagram of UAV vision system.

If the activation function of a hidden layer is defined as $g(x)$, there are generally l neurons in the hidden layer. The weight and deviation of the jth Tibetan layer generated randomly are represented as a_j and b_j, respectively, and the weight vector connecting the node of the jth hidden layer and outputting a node is represented as M. One of the main learning goals of the extreme learning machine is to minimize the training error as much as possible and to reduce the output weight as much as possible (1).

$$\min \|M\alpha - T\|^2 \& \|\alpha\|. \tag{1}$$

In the equation, $M = \{MIJ\} = \{G(a_j, b_j, x_j)\}$, $i = 1, \cdots, n$, $j = 1, \cdots, L$, H is the output matrix attached to the middle of the airtight layer, HIJ is the output of the jth confined layer node, x_i represents a confined layer node. In Equation (1), $\alpha = [\alpha_1, \alpha_2, \alpha_L] T$, $M(x_i) = [m_1(x_i), m_2(x_i), m_L(x_i)]$, and $T = [t_1, t_2, t_n] T$.

According to the literature [15], formula (1) can be solved by the following formula:

$$\alpha = M^+ T. \tag{2}$$

In the equation: M^+ is the generalized inverse of matrix M. The earliest learning machine principle was to deal with the faults of the feedback neural network of a single airtight layer, but in the later stage, a large number of work-related personnel extended the principle of extreme learning to problems that are not network neural, which also verified the limit. The applicable conditions of the learning machine are lower than the vector mechanism and the least squares mechanism [16], and this is the case for the extreme learning machine in this article.

The main constraint optimization problem of the extreme learning machine is defined as the following

formula:

$$\min LPELM = \frac{1}{2}\|\alpha\|^2 + C\frac{1}{2}\sum_{i=1}^{n}\|\gamma_i\|^2. \tag{3}$$

The constraints are

$$h(x_i)\alpha = t_i^T - \gamma_i^T, i = 1, \cdots n. \tag{4}$$

In the equation: $i = [\gamma_i, 1, \gamma_i, 2, \cdots \gamma_i, q]\, T$ is the vector error of the action collection of the q output nodes for the sample x_i, and c is the regularization variable. According to the conditions of Kaiduoyili, the optimization problem encountered can be transformed into the following equation:

$$\alpha = H^T\varphi, \varphi_i = C\gamma_I, h(x_i)\alpha - t_i^T + \gamma_i^T = 0, i = 1, \cdots n, \tag{5}$$

where T is the Lagrange multiplier matrix. The final output weight α is calculated as the following formula:

$$\alpha = H^T\left(\frac{I}{C} + HH^T\right)^{-1} T. \tag{6}$$

Therefore, the output function of the extreme learning machine can be defined as the following formula:

$$f(x_j) = h(x_i)H^T\left(\frac{I}{C} + HH^T\right)^{-1} T, j = 1, \cdots n. \tag{7}$$

The extreme learning mechanism and the vector support principle are highly similar; you can convert the above kernel function into the extreme learning machine and then limit the range of conditions for its function. The limit condition Merece theorem can transform the output into the following equation:

$$f(x_j) = h(x_i)\alpha = \left[K(x_j, x_i) \cdots K(x_j, x_n)\right]^T\left(\frac{T}{C} + K\right)^{-1} T. \tag{8}$$

In the equation, $j = 1, \cdots, n$. After the forwarding simplified processing, the program can classify and quantify the output of the video.

The backing of deep learning requires countless data and a large number of complex calculations. In order to improve the training effect of neural networks, it is necessary to carry out tens of thousands of iterations for each training process and adjust millions of parameters. Because this subject needs to process a large amount of data and train and test the deep network, Amax's XG-48201G model server is selected as the deep learning computing platform [17]. The specific configuration of the platform is shown in Table 1.

In the process of multitarget optimization by the classic detection-tracking-self-learning tracking method, the method of dynamically adjusting the detection area is used to optimize the detector. The multitarget tracking module

TABLE 1: Configuration information of deep learning computing platform.

Part name	Selection type
CPU	E5-2650 V4
GPU	NVIDIA Tesla V100 × 8
RAM	DDR4 16 G × 2
Hard disk	SSD 480 G × 2
System	Ubuntu 16.04
BIOS version	2.0a
CUDA version	9.0
Deep learning framework	Caffe/TensorFlow/PyTorch

is constructed with multiple multicentroid methods, and the multitemplate library is used. The online learning method constructs a multiobjective self-learning module. Because the structure mechanism of detection-tracking-self-learning has irreplaceable advantages, the idea of this module is still used in the process of multiobjective optimization. However, the internal implementation has been greatly improved.

2.3. Image Information Preprocessing. SVM (support vector machine) is a very classical machine learning approach. It has great advantages in the recognition of small sample size and nonlinear shapes, for example, global optimization. Strict application of simple structure, learning and short prediction time, originated from today. It is as alive as neural networks, and at this point, it remains the mainstream of machine learning. It has attracted a research boom from many institutions and scholars. And it produced many improved SVM methods; moreover, the SVM algorithm has many applications in face detection, language recognition, text classification, etc. This topic is to detect multiple targets using SVM methods. The key to detection lies in the training of the SVM classifier, and the proof of training is the selection of sample properties. And it selects the HOG function to describe the example [18–20].

The HOG function is the most widely used function for detecting image targets. It is very suitable for extracting preview attributes with rich edge information. And it is widely used in target detection. Especially combining SVM algorithm in pedestrian detection, the effect is very good. First, divide the image into smaller parts. Then, get the histogram statistics of the pixel gradient direction in the cell, and finally connect them to create high-dimensional attributes. The extraction process is shown in Figure 4.

The implementation process of the HOG feature extraction algorithm is as follows: Reduce the influence of light factors. The collected image will first become gray. Then, it is corrected by the gamma method and normalized to obtain the normalization of the color space. In order to obtain the corresponding dimension of the HOG attribute, the generated target image is scaled to the same size. The pixel
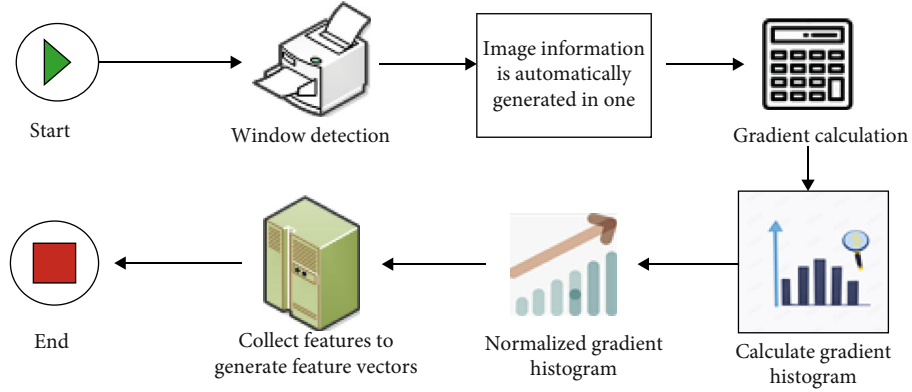
FIGURE 4: HOG feature extraction process.

gradient (x, y) is

$$G_x(x, y) = H(x + 1, y) - H(x - 1, y),$$
$$G_y(x, y) = H(x, y + 1) - H(x, y - 1), \tag{9}$$

where $H(x, y)$, $G_x(x, y)$, and $G_y(x, y)$ represent the pixel value of each pixel. The horizontal gradient and vertical direction, as well as the size and direction of the gradient, can be determined by the following formula:

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2},$$
$$\theta(x, y) = \tan^{-1}\left(\frac{G_y(x, y)}{G_x(x, y)}\right). \tag{10}$$

Separate the scaled image from the ground target in the same 16×16 pixel unit, and create a histogram of 9 boxes for each unit (dividing the gradient from 0 to 180 degrees into 9 regions. Each channel corresponds to the gradient direction within the range of 20 degrees). Each pixel in the cell is a "vote" for certain bin fields in the histogram, and votes have different weights to make the "voting" more efficient. The linear interpolation method of the gradient direction is used for the pixels, and the gradient amplitude is weighted, namely

$$k(x, y, \theta_1) \longleftarrow k(x, y, \theta_1) + |\nabla f(x, y)| \left[\frac{\theta_2 - \theta(x, y)}{\theta_2 - \theta_1}\right],$$
$$k(x, y, \theta_2) \longleftarrow k(x, y, \theta_2) + |\nabla f(x, y)| \left[1 - \frac{\theta_2 - \theta(x, y)}{\theta_2 - \theta_1}\right], \tag{11}$$

where (x, y) is the gradient direction of the pixel (x, y) in the cell. Date is the center of two adjacent boxes, $0(x, y)$, $k(x, y, 0)$; instead, corresponding to the weight in the tank, the method used to combine the honeycomb unit into a large area (block) involves combining single cellular units that are combined into large, spatially connected blocks. In order to concatenate the feature vectors of all cell units in a block to obtain the HOG features of the time period, these inter-

vals overlap, that is, individual cell features appear multiple times in the feature vector, and the results are different. We call the standard HOG descriptor vector attributes and finally collect the HOG attributes of all overlapping intervals in the detection window. And the final feature vector is synthesized for use by the classifier, which is an example of SVM classifier [21, 22].

Suppose there is a set of training samples x of size N, which is a vector of dimension d and the label of the sample. It shows two different types: positive samples and negative samples. The main idea of SVM is to find the most suitable hyperplane for classification.

$$w \times x + b = 0, \tag{12}$$

which makes

$$w \times x_i + b \geq 1, x_i \in \text{class1 } y_i = 1,$$
$$w \times x_i + b \leq -1, x_i \in \text{class2 } y_i = -1, \tag{13}$$

combined into a unified form:

$$y_i[(w \times x_i + b)] - 1 \geq 0, i = 1, 2, \cdots, N. \tag{14}$$

The inner products of vectors w and x and class1 and class2 are two hyperplanes. A suitable hyperplane will correctly separate the positive and negative samples from the training samples. And the corresponding maximum classification range (margin) is shown in Figure 5.

This results in the following classification function:

$$f(x) = \text{sign} (w \times x + b). \tag{15}$$

The classification results now have better generalization capabilities. The symbol is a function of the symbol. It can be seen from Figure 5 that the appropriate hyperplane parameter solution should be maximized by 2, which is equivalent to 1/2 reduce, which is transformed into a scaling problem. We write the following secondary program:

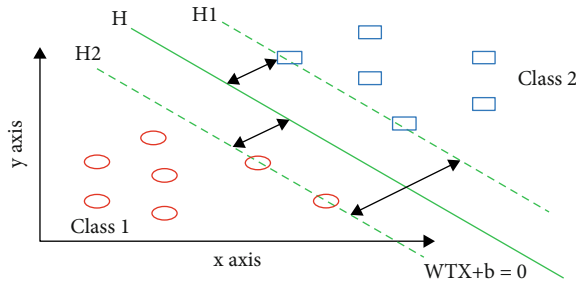$$\min \frac{1}{2} \|w\|^2. \tag{16}$$

Figure 5: Optimal classification hyperplane.

The SVM algorithm is derived from the optimal classification function. Combined with the HOG function introduced in the previous section, the classification of positive and negative samples can be achieved by training the classifier, and the next step is the training process of the classifier.

3. Experimental Model and Framework

3.1. YOLO Model. YOLO (you only look once) is a unique neural network-based target detection system proposed by Joseph Redmon and Ali Farhadi in 2015. Based on the end-to-end concept, YOLO treats the target detection task as a regression problem and directly obtains the object confidence of the bounding box and category probability of all pixels of the bounding box coordinates. The YOLO forward object detection method essentially creates a large number of possible bounding boxes, which can contain objects to be detected through the selected area, then uses the classifier to determine whether each bounding box contains an object. And the probability or certainty of a class of objects, such as R-CNN, Fast R-CNN, and Faster R-CNN. The first step for YOLO is to correspond. First, the captured image is divided into table cells, and each cell is explored to find out which cell is closest to the center of the ground truth range area. The second part defines the trust level. It uses the specific cells detected in the first step to determine whether the boundary line overlaps with the real situation (IoU) and then reduces the confidence of the boundary line with less overlap. It decreases the confidence of all bounding boxes that have no objects in each cell. The structure of the YOLO network consists of 24 convolutional layers and the last 2 fully connected layers [19].

Compared with other deep learning target detection algorithms, YOLO has the following advantages: (1) fast detection speed, YOLO solves the target detection problem as a regression problem. And each frame only predicts $7 \times 7 \times 2 = 98$ cells, so the speed is extremely fast. The standard version of YOLO can reach 45 FPS; (2) low background error detection rate on Titan X GPU of all images during training. However, detection algorithms such as R-CNN use sliding windows or region suggestions. The classifier can only receive local data. And it is easy to detect the background as an object; (3) YOLO can learn the general characteristics of the object. It is also suitable for object detection in this field. And the detection rate is higher than the R-CNN detection method, but the early version of YOLO also has

some shortcomings, such as easy to miss detection. The positioning accuracy is low, and the detection effect of small objects is poor. In order to overcome the above problems, researchers have proposed improved versions of YOLOv2 and YOLOv3 [23, 24]. The comparison is shown in Table 2.

YOLOv3 uses the latest classification network, which is better than other classification networks. Compared with YOLOv1 and v2, the improvements are mainly in the following aspects: YOLOv3 uses logistic regression to predict the unfair score of each bounding box. If the currently projected bounding box is more in line with the ground truth object than the previous box, if the current prediction is not the best, the score is 1. Instead, it coincides with a real object on the ground above a certain threshold, and the neural network ignores this prediction [25]. The standard used in the test is 0.5. Unlike YOLOv2, YOLOv3 assigns a limit frame to a simple real object. For each item only, if no previous bounding box is assigned to the corresponding object, only invalid objects will be detected without affecting the coordinates or classification prediction. The boundary block diagram with a priori dimension and position prediction is shown in Figure 6.

First, compare the reference network with the fully rotating twin network of the reference network using this algorithm. Both use only n-layer networks to separate functions, which means that it ends after the nth warp layer (if this is the editing layer, it is merged, if it is the aggregation layer). It is not included when $n = 2$, as shown in the figure. As shown in Figure 7(a), this paper is significantly better than the reference network when $n = 5$. As shown in Figure 7(b), the gap between DCFSNN and the reference network is not large. The results show that as the network depth increases, DCFSNN's the progress will not be as great as possible. There are more benefits to using shallow network features to check whether the DCFSNN algorithm is most suitable for shallow networks. In the theoretical part of the DCFSNN algorithm, after the neuron sums the weights, it needs to perform a nonlinear transformation, which is to pass in the activation function as a parameter. This activation function is a DCFSNN algorithm.

The SRDCF algorithm is a generalized correlation filtering algorithm. And this article is used as a reference for the filter part. Due to the different network frameworks used, it is not suitable for direct testing and comparison. Theoretically, the improvement effect is very good. The distortion factorization leads to significant performance improvement and significant reduction in complexity. The filter part goes from D to C [26]. Updating the training set can improve performance and simplify calculations, with samples ranging from M to L. The new update strategy reduces the number of unplanned updates required and further increases the speed.

3.2. UAV Target Tracking System Framework. Several similar targets include five identical red targets and similar RGB colors and five similar green targets. Continuous tracking means that in an infinitely long video clip, when the tracked target is stuck or the tracking fails, the system can redetect the target and continue tracking [27].

Table 2: Improved comparison between YOLOv2 and YOLOv1.

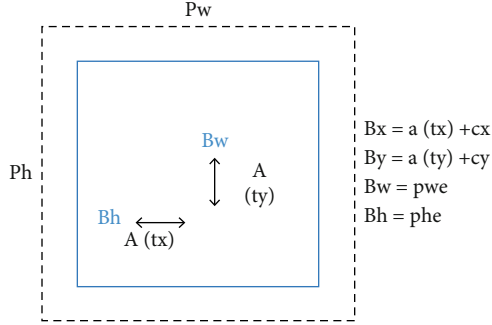|	YOLO				YOLO$_{V2}$
Batch norm?		✓	✓	✓	✓
Hi-res classifier?			✓	✓	✓
Pass through?				✓	✓
Anchor boxes?		✓	✓		
New network?		✓		✓	✓
VOC2007 mAP	61.5	65.3	68.4	74.2	73.5

Figure 6: Bounding box with dimensionality prior and location prediction.

The main function of the designed fusion device is to identify the candidate targets obtained from the detector, and the color segmentation method will be used to solve this problem. Color is the most intuitive attribute of human vision. Hundreds of color spaces have been provided, many of which are used for specific applications. The most commonly used color spaces today are RGB, HSV, YUV, CMY, etc. The color space is more frequent. Therefore, it uses the HSV color space to segment the target color. The OpenCV image processing library is used to segment the target color. In the library, colors are stored as H, S, and V in an 8-bit char format: the hue, saturation, and brightness range of the actual model h, s, and v. It is the value range of the corresponding OpenCV image library. Common color ranges are shown in Table 3.

The 8×8 pixel block in the image is taken as a cell, the gradient histogram is counted in each cell, and the gradient direction (0-360) is divided into nine, as the horizontal axis of the histogram, the corresponding gradient in the angle range. The accumulated value is used as the vertical axis of the histogram. For an image, first change the width and height to a value divisible by 8, and then divide it into non-overlapping cells according to 8×8, each 2×2 cell as a block; there is overlap between different blocks. Next, normalize the block, and combine all the block features to form a relatively large feature vector. This is the TPD-KCF algorithm. The detector trained in Section 2 can obtain the detection frame of the target. However, due to factors such as the environment and the detection angle, false detection frames will inevitably appear, because the target is divided into red and green. We need to specify a goal, that is, the possible goals must be divided into three types: red goals.

It can be seen from the green target and target interference that TLD shows considerable advantages in terms of accuracy. In the long-term tracking process, the accuracy rate reaches 100%, the tracking-learning-detect mechanism works well, and the average CLE is the smallest. That is, the accuracy of the tracking result frame is very high. But it also wastes a lot of time. The particle filter algorithm also shows better results in tracking short scenes or single scenes, although the processing speed cannot meet the real-time requirements. However, the calculation process is relatively simple and not suitable for complex scene tracking. The overall performance of TPD-KCF tracking based on this article is the best, and it can accurately track long-term results in real time. The tracking performance results of the TPD-KCF algorithm in this paper are shown in Table 4.

When using the TPD-KCF algorithm of this article for tracking, Video 2 is not lost, Video 3 is lost 3 times, and Video 1 is lost 2 times, which can be parsed as 25 frames. The actual situation of surveillance and control of small drones can be effectively tracked. Figure 8 shows the distance accuracy curve. The distance accuracy curve can be used to indicate the percentage of images that are tracked correctly against a set of central error standards. This reflects the positioning accuracy of the tracking method to the target center. For foreground prediction, the main purpose of this operation is to solve the target loss caused by problems such as occlusion. However, due to the combination of multicentroids and the correction of decision-making learning, this problem is better solved, the state strategy is integrated into the trajectory tracking of each centroid, and excellent results are obtained.

If the distance between the target center of the tracking result and the actual target center is less than or equal to the threshold, the frame is considered tracked correctly. When planning distance standards, accuracy parameters are not required. This makes the curve clear and easy to interpret.

3.3. Target Detection and Tracking. Moving target tracking involves finding moving targets in each frame of a continuous video segment, and matching targets between adjacent video images, thereby establishing connections between objects in the video sequence and determining the trajectory of the target. In recent years, researchers have developed many tracking algorithms. At present, the most common target tracking algorithm is a feature-based tracking algorithm. The feature-based tracking method includes two steps: separating attributes and matching attributes. It can also be used when the object is partially occluded. But there are several drawbacks to note:

(1) When the POI is closed while the target is moving, it is hard to distract

(2) It is difficult to use uniform functions in all occasions. The appropriate attributes often depend on the actual detection target

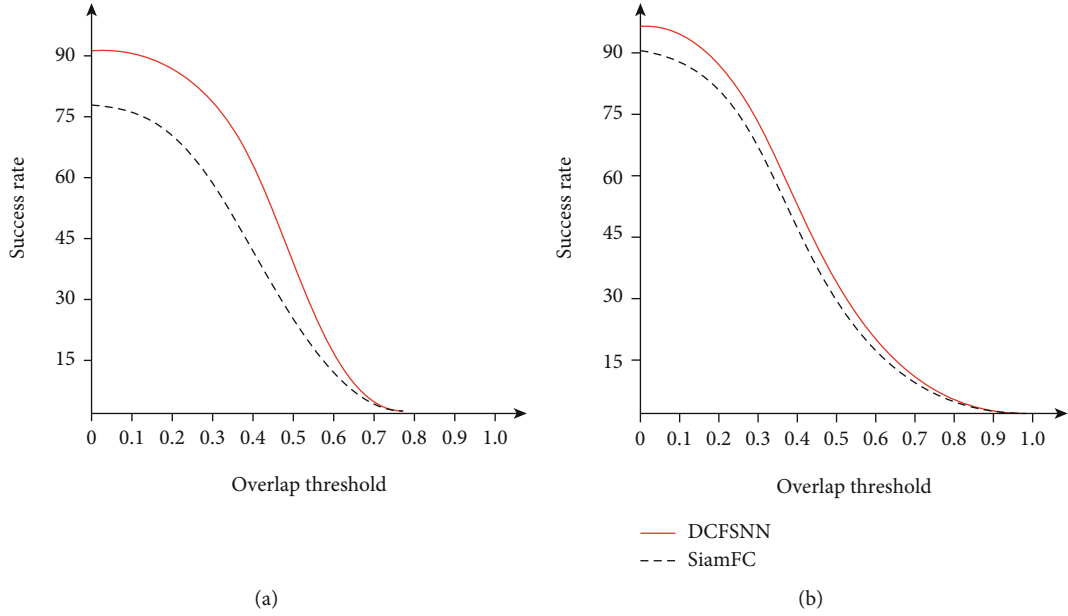(3) For a specific object, many features are often required. In actual testing, missed testing and

(a)

(b)

— DCFSNN
--- SiamFC

FIGURE 7: The success rate of rectangle overlap of the tracker on the verification set.

TABLE 3: Common color HSV space value range.

	Black	Gray	White	Red		Orange	Yellow	Green	Blue	Purple
h_{min}	0	0	0	0	154	11	25	36	45	135
h_{max}	185	185	185	15	185	40	10	185	115	175
s_{min}	0	0	0	48		35	95	235	257	180
s_{max}	115	125	145	255		255	255	255	255	255
v_{min}	10	145	135	255		255	255	255	255	255
v_{max}	185	125	145	255		255	255	255	255	255

TABLE 4: Algorithm tracking performance.

Video	TPD-KCF		
	Video 1	Video 2	Video 3
Average CLE	1.359	2.689	13.587
Average OP	99.85%	90.69%	93.451%
Average frame loss rate	0.158%	15.236%	6.012%
Average FPS	26.78 FPS	29.68 FPS	30.25 FPS

enhanced functions may occur. This makes it difficult to identify objects.

The use of correlation filters in the tracking field was first mentioned in a CVPR article in 2010. Subsequent studies have shown that this type of algorithm can greatly improve the tracking speed. Therefore, it has become a hot topic in the current tracking field. After the improvement of the academic circle, the tracking algorithm based on the correlation filter has made significant progress in both the algorithm efficiency and the tracking effect. Table 5 lists the algorithms based on the monitoring filter provided by the academic circle in recent years.

Taking street motion multitarget motion prediction (including target motion direction prediction and target number recognition) as an example, MOT benchmark is selected as the test and performance evaluation benchmark to conduct multiscene detection and tracking effect testing and comparative analysis. On the basis of horizontal and horizontal movement, a longitudinal tracking video sequence from far to near is added to test the detection and tracking effect of the system when processing target size and turning angle changes. Experiments show that the overall efficiency of the GMR algorithm is the best, and the improved GMR algorithm has a better detection effect. Not only the target is highlighted while blurring the background but also the outline of the key target is continuous. Basically, the detection results for different targets are the same. The performance of the algorithm is also superior, because the direction of movement of the tracking target will have an angle every 5 s. By obtaining the system time t, the trajectory of the target can be predicted within an offset interval of 5 s, that is, the collected duration adjustment data. The trajectory is the remainder from t to 5. As shown in Figure 9, the target tracking superframe comparison line chart in 5 test flights.
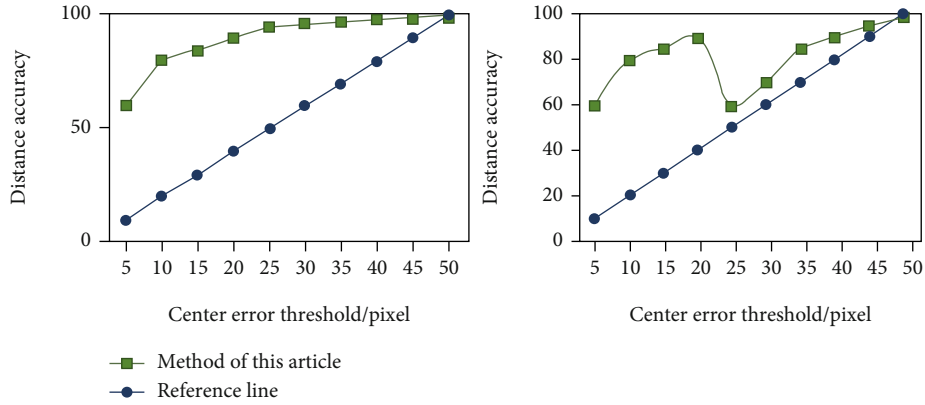
Figure 8: Distance accuracy curve.

Table 5: Some related filter algorithms.

| Algorithm name	Year of publication	Main contribution
MOSSE	2010	Proposed to apply correlation filters to virtual tracking
CSK	2012	Introduce the use of circulant matrix to solve the ridge regression problem in the kernel function
STC	2013	Spatiotemporal contextual information
CN	2014	Introduced color features as feature extraction
DSST	2014	Use pyramid features and three-dimensional correlation filters to solve scale problems
RPAC	2014	Local tracking strategy
KCF	2014	Use multichannel HOG features as feature extraction
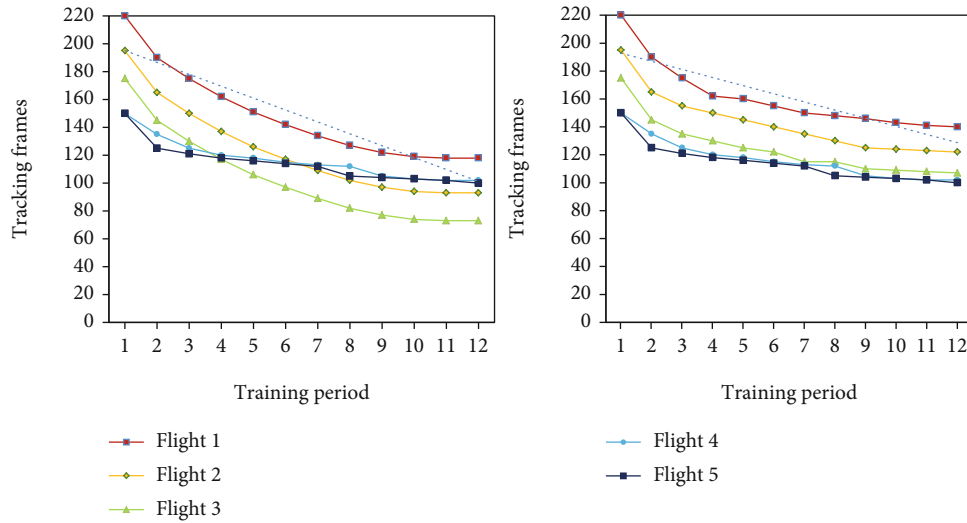
Figure 9: Line chart of flight tracking comparison.

For the data set, it mainly contains two parts, one is the training set, and the other is the measurement set; and for the table information, it mainly includes the actual number of labeled target frames, the basic size of each frame of image, the length of time, and the average actual number of pedestrians appearing in each frame. At this time, the trajectory of the moving target can be regarded as linear, so the fitting function can be simplified as $F(x) = ax + b$, which is 20 FPS, so the exact length of the data $h \in (0.100)$ can be determined. Given the length of the data, the trajectory with $h < 10$ is rejected, and the trajectory with $h > 20$ is intercepted with 60 valid data, which speeds up the speed of straight-line adjustment and the real-time adjustment. When the abnormal deviation of the target angle causes
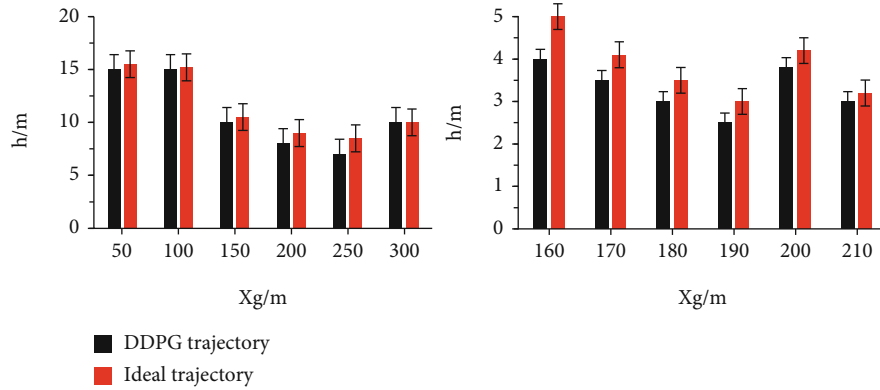
Figure 10: Comparison of partial magnification effects.

the target tracking to fail, the long-term path, the prediction has no practical meaning. And at this time, the target path disappears only once.

4. Analysis of Experimental Results

4.1. Controller Accuracy Analysis. Compared with the traditional PID controller with a maximum height error of 0.4361 m, the DDPG control method has a maximum height error of only 0.2491 m. The control accuracy is better than the PID control method. It does not involve the DDPG actor network in the previous actions, so the continuity of the output operation is very high. This translates into huge fluctuations in the middle position. Check the efficiency of the Kalman filtering and least squares trajectory prediction methods. It is determined by predicting the variability and complexity of targets outside the global field of view. The experiment is to predict the target trajectory appearing in the four-leaf field of view in the ROS RVIZ experiment under the auxiliary environment of the visual interface. The grid size in the figure is 11 mm. The big green square is a four-rotor motion component. The dots on the rectangle represent the small target node of the four-rotor. The black line represents the quadrotor flight trajectory, and the black dot represents the quadrotor flight trajectory. For the rotorcraft, the white line represents the ground target trajectory. The red dot represents the impact comparison between the ground mobile robot target and the PID controller, as shown in Figure 10.

The above experimental results show that when starting the target manually, the tracking period does not match the tracking target for various reasons. At this time, because the scale selected by the tracking window is too large, there is background interference data, which affects the learning of the filter algorithm and uses prominence to isolate the tracking target. And the target has been recalibrated in the gray-scale image. This eliminates a lot of background in the original trace frame. The target tracking of frame 350 in the tracking scene 3 is highlighted and optimized without a hit detection algorithm. Follow the target as green, and mark red as the following target. A highlight detection algorithm can be used. For the tracking of nonlateral movement of the target, the tracking effect of the TC-ODAL method is

not good. There is a multiframe missed detection and false detection, and the continuous tracking processing of the target size deformation also has an error. In the target tracking number and target tracking frame number, there are problems in target tracking speed and anti-interference ability.

4.2. UAV Predictive Tracking Analysis Results. In this experiment, the robot bulb is spherical. Regardless of the position of the ball, any direction will be circled in the image. The image of the center of the sphere on the image is the center of the circle on the image. Therefore, this method is more accurate than traditional smart CCTV detection algorithms. According to experimental measurements, the maximum error of the robot position data detected by this method is no more than 3 cm, which is nearly 10 times more accurate than the traditional detection algorithm (encoded book algorithm). This algorithm saves a lot of memory space and improves real-time performance. The average detection time of this algorithm on the host is about 20 milliseconds per frame. The codebook algorithm has a detection time of more than 30 ms per frame.

For the training sequence KITTI-17, the tracker in PETS09-S2L1 achieved a more accurate tracking effect than KITTI-13. In summary, the algorithm of this subject has a good performance in the test sequence of each scene. Therefore, the detection algorithm based on local features is efficient and is suitable for this experimental platform. It can be seen that TLD shows a great advantage in accuracy. In the process of long-term tracking, 100% accuracy and learning tracking-the efficiency of the detection engine are very good; the average CLE is extremely small, that is, the accuracy of the frame tracking result is very high. But it also wastes a lot of time. The particle filter algorithm also shows better results in tracking short scenes or single scenes, although the processing speed cannot meet the real-time requirements. But the calculation process is very simple and cannot adapt to the situation of complex scene tracking. The overall performance of TPD-KCF tracking based on this article is the best. Here, we have chosen a representative accuracy score, which is a distance standard of 20 pixels. The value in the description represents the distance accuracy value of the method in the article. This is the same as the KCF method. When the center error value is 20 pixels, it

can be seen that when the center error value is 20 pixels, the result of the TPD-KCF method is tracked correctly, and the frame rate can reach 98%. Compared with the use of KCF tracking, the performance is significantly improved.

5. Conclusion

Although this article considers that the experimental conditions have a slight influence on the target trajectory detection and tracking algorithms, a more suitable detection algorithm and an improved shadow appearing algorithm are selected for the experimental conditions. Solve the problem of poor target positioning accuracy caused by traditional detection algorithms, which cannot meet the robot's construction positioning and rounding requirements. According to the actual situation of the robot in the experiment, the feature-based local detection algorithm proposed is better, which solves the problem of positioning accuracy and has better real-time performance. Image acquisition cycle is 36 ms, and the speed of the robot in the experiment will be the fastest, 1/ms. The image size is 600×800, and the vertical direction of the ground where the image can be taken is about 10 meters. Therefore, the robot can move up to 2 pixels vertically between adjacent images. The exposure time is short, and the robot moving distance is small. Therefore, according to the detection results, this paper adopts the feature-based tracking method and the Kalman filter-based target tracking method. It can be seen that using the results of the TPD-KCF method, when the overall error criterion is 20 pixels, the percentage of correct tracking images can reach 98%. Compared with KCF tracking, the performance has been significantly improved, indicating this article. The algorithm is efficient and performs well in real time, but the research content and application background of moving target detection and tracking are very extensive, and there are many factors that need to be considered. Due to personal energy and time constraints, this integrated UAV system that predicts path following from image data still has a long way to go. The system's response time and UAV tracking stability need to be further improved. Looking forward to the following algorithms, it can solve the problem that the frame rate of the drone when tracking the target may reach a higher frame rate.

Data Availability

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

Conflicts of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgments

References

[1] G. Kuechle, B. Boulu-Reshef, and S. D. Carr, "Prediction- and control-based strategies in entrepreneurship: the role of information," *Strategic Entrepreneurship Journal*, vol. 10, no. 1, pp. 43–64, 2016.

[2] W. Xu, Z. Zhu, F. Ge, Z. Han, and J. Li, "Analysis of behavior trajectory based on deep learning in ammonia environment for fish," *Sensors*, vol. 20, no. 16, p. 4425, 2020.

[3] T. Long, "Research on application of athlete gesture tracking algorithms based on deep learning," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 9, pp. 3649–3657, 2020.

[4] J. Li, W. Zhan, Y. Hu, and M. Tomizuka, "Generic tracking and probabilistic prediction framework and its application in autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3634–3649, 2020.

[5] Y. Liu, H. Wang, J. Fan, J. Wu, and T. Wu, "Control-oriented UAV highly feasible trajectory planning: a deep learning method," *Aerospace Science and Technology*, vol. 110, no. 3, article 106435, 2021.

[6] G. Araniti, A. Orsino, L. Militano, L. Wang, and A. Iera, "Context-aware information diffusion for alerting messages in 5G mobile social networks," *IEEE Internet of Things Journal*, vol. 4, no. 2, pp. 427–436, 2017.

[7] J. Wang, "Research on data mining and prediction of large-scale competitions based on online data migration model," *Journal of Intelligent and Fuzzy Systems*, vol. 2, pp. 1–12, 2021.

[8] Z. Lv, A. K. Singh, and J. Li, "Deep learning for security problems in 5G heterogeneous networks," *IEEE Network*, vol. 35, no. 2, pp. 67–73, 2021.

[9] W. Gai, J. Liu, C. Qu, and J. Zhang, "Trajectory tracking control for a quadrotor UAV via extended state observer," *Systems Science & Control Engineering*, vol. 6, no. 3, pp. 126–135, 2018.

[10] D. Popescu, F. Stoican, and L. Ichim, "Control and optimization of UAV trajectory for aerial coverage in photogrammetry applications," *Advances in Electrical & Computer Engineering*, vol. 16, no. 3, pp. 99–106, 2016.

[11] L. H. Zhou, J. X. Dou, J. Q. Zhang, and B. C. Wen, "Trajectory tracking control for a quadrotor UAV based on improved backstepping," *Journal of Northeastern University*, vol. 39, no. 1, pp. 66–70, 2018.

[12] R. Lou, Z. Lv, S. Dang, T. Su, and X. Li, "Application of machine learning in ocean data," *Multimedia Systems*, no. - article 733, 2021.

[13] J. Li-Chee-Ming and C. Armenakis, "UAV navigation system using line-based sensor pose estimation," *Geo Spatial Information Science*, vol. 21, no. 1, pp. 2–11, 2018.

[14] W. Wang, X. Li, L. Xie, H. Lv, and Z. Lv, "Unmanned aircraft system airspace structure and safety measures based on spatial

digital twins," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10, 2021.

[15] X. Hu, K. K. Wong, K. Yang, and Z. Zheng, "UAV-assisted relaying and edge computing: scheduling and trajectory optimization," *IEEE Transactions on Wireless Communications*, vol. 18, no. 10, pp. 4738–4752, 2019.

[16] Y. He, Y. Peng, S. Wang, and D. Liu, "ADMOST: UAV flight data anomaly detection and mitigation via online subspace tracking," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 4, pp. 1035–1044, 2019.

[17] X. Yu and L. Liu, "Target enclosing and trajectory tracking for a mobile robot with input disturbances," *IEEE Control Systems Letters*, vol. 1, no. 2, pp. 221–226, 2017.

[18] X. Lu, Z. Jia, X. Hu, and W. Wang, "Double position sensitive detectors (PSDs) based measurement system of trajectory tracking of a moving target," *Engineering Computations*, vol. 34, no. 3, pp. 781–799, 2017.

[19] S. Zhou, Y. Song, and K. Zhao, "Accelerated neuroadaptive tracking control of strict-feedback nonlinear systems without precise knowledge of target trajectory," *International Journal of Adaptive Control & Signal Processing*, vol. 32, no. 1, pp. 27–49, 2018.

[20] Z. Wan, Y. Dong, Z. Yu, H. Lv, and Z. Lv, "Semi-supervised support vector machine for digital twins based brain image fusion," *Frontiers in Neuroscience*, vol. 15, article 705323, 2021.

[21] A. Milan, K. Schindler, and S. Roth, "Multi-target tracking by discrete-continuous energy minimization," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 38, no. 10, pp. 2054–2068, 2016.

[22] S. Memon, H. Son, K. H. Memon, and A. Ansari, "Multi-scan smoothing for tracking manoeuvering target trajectory in heavy cluttered environment," *Iet Radar Sonar & Navigation*, vol. 11, no. 12, pp. 1815–1821, 2017.

[23] L. Xi, Z. Peng, L. Jiao, and B. M. Chen, "Smooth quadrotor trajectory generation for tracking a moving target in cluttered environments," *SCIENCE CHINA Information Sciences*, vol. 64, no. 7, pp. 1–16, 2021.

[24] S. Vilardaga, X. Prats, P. P. Duan, and M. uijt de Haag, "Conflict-free trajectory optimization with target tracking and conformance monitoring," *Journal of Aircraft*, vol. 55, no. 3, pp. 1252–1260, 2018.

[25] Z. Lv, D. Chen, H. Feng, R. Lou, and H. Wang, "Beyond 5G for digital twins of UAVs," *Computer Networks*, vol. 197, article 108366, 2021.

[26] J. Xiang, N. Sang, J. Hou, R. Huang, and C. Gao, "Multitarget tracking using Hough Forest random field," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 11, pp. 2028–2042, 2016.

[27] M. Anvaripour, M. Saif, and M. Ahmadi, "A novel approach to reliable sensor selection and target tracking in sensor networks," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 171–182, 2020.