Image Recognition Technology of UAV Tracking Navigation Path Based on ResNet

Lulu Liu[1](✉), Degao Li[1], Junqiang Jiang[1], Shibai Jiang[1], Linan Yang[1], and Xinyue Chen[2]

[1] State Grid Xinjiang Electric Power Co., Ltd., Information and Communication Company, Urumqi 830001, China
2429496766@qq.com

[2] The Hong Kong University of Science and Technology, Hong Kong 999077, China

Abstract. The image of UAV tracking navigation path is easily affected by noise, and there are problems of low image recognition accuracy and poor image enhancement effect. Therefore, a ResNet-based image recognition technology for UAV tracking navigation path is proposed. Extract image features and analyze image color and shape. The Laplacian operator is used to process the image to enhance the edge of the image. Using the bilinear interpolation method, the image is scaled and grayscale is processed, and the noise is processed by combining the wavelet transform. Build a ResNet-based recognition model, use a multi-resolution octree hierarchical structure, render each node, and output any node image coordinates. Perform global pooling on the input feature map to improve image degradation. The gradient image is binarized using the binarization method. Fully consider the characteristics of the UAV tracking and navigation path, and use the statistical averaging method to obtain the average interference amplitude and phase, and calculate the interference characteristic distance. The iterative threshold selection method is used to obtain image recognition results. It can be seen from the experimental results that this technology can extract comprehensive image information, and has a high signal-to-noise ratio, which can achieve the purpose of image enhancement, and the highest image recognition accuracy obtained is 0.96, with accurate recognition results.

Keywords: ResNet · UAV Tracking · Navigation Path · Image Recognition

1 Introduction

UAV is a kind of aircraft that can independently complete specific tasks under unmanned conditions. It is the product of continuous development and progress of information technology, mainly involving sensor technology, communication technology, information processing technology, intelligent control technology, aviation propulsion technology and other advanced technologies. At the beginning of the UAV, it was widely used in the military field because of its small size, strong mobility, and the ability to effectively avoid casualties. UAV tracking navigation is a navigation technology based on computer

B. Wang et al. (Eds.): ICMTEL 2023, LNICST 533, pp. 23–39, 2024.
https://doi.org/10.1007/978-3-031-50574-4_3

vision. Due to the advantages of complete autonomy, anti electromagnetic interference, and high navigation accuracy when approaching the target, UAV tracking navigation has received extensive attention and research at home and abroad, and has become a navigation mode with broad prospects. Some scholars have discussed using UAV tracking navigation to achieve autonomous landing or landing of UAV. UAV tracking navigation is to calculate the motion parameters of the carrier according to the real-time images taken by the camera on the carrier, so as to conduct the navigation and control of the carrier. The premise of navigation solution using target image information is to accurately obtain the features of target image, and then quickly use the obtained features to calculate the position and attitude of UAV. This process requires image recognition [1].

At present, reference [2] has proposed an image recognition method of anti UAV system based on convolutional neural network. Using self-made optical system acquisition equipment, pictures of different types of UAVs and birds have been collected, and convolutional neural network and support vector machine for small sample recognition of UAVs have been designed. The designed convolutional neural network is used to recognize MNIST datasets, UAV images and bird images respectively, and support vector machine is also used to recognize UAV and bird images, and comparative experiments are carried out. Reference [3] proposed a research and application of UAV image recognition technology system. Based on the YOLO v3 algorithm framework, a high-precision YOLO v3-SE target detection algorithm is constructed by introducing the attention module SE, forming a UAV image recognition technology system. It has been successfully applied to the detection and recognition of abnormal features in massive UAV images of many water conservancy projects. Reference [4] proposed a parallel feature pyramid network UAV aerial image target detection algorithm based on Cascade R-CNN. An improved algorithm based on feature pyramid network is studied. Parallel branches are added on the original basis, and feature information of two different sampling methods is used to enhance the expression ability of small target features. At the same time, increase Cascade R-CNN to strengthen small target positioning capability. Reference [5] proposed to apply machine learning algorithm to UAV to obtain high-precision aerial images. The image is divided into training image and experimental image. The trained image is input into different models, and the output of each model is evaluated using independent test data set. Experiments show that the image obtained by this method has high accuracy. Reference [6] proposed an improved UAV identification and detection method based on the integrated learning method of hierarchical concept. The method consists of four classifiers and works in a hierarchical manner. The first classifier is used to check the practicability of the UAV, the second classifier is used to detect the UAV type, and the last two classifiers are used to process the relevant samples. The public data set is used to evaluate the proposed method, which shows good recognition and detection efficiency.

As far as the current development situation is concerned, it is still in the initial stage. The image recognition algorithms used have not been well applied to image recognition related fields. Some applied algorithms still have many problems, such as low recognition accuracy, poor generalization performance, low processing speed, etc. Faced with the above problems, the image recognition technology of UAV tracking and navigation path based on ResNet is proposed. The color and shape features of the image are extracted, and the image is processed by edge enhancement, scaling gray scale and denoising. Build a recognition model based on ResNet, input the preprocessed image, render the image with multi-level detail rendering, and perform global pooling processing to improve image degradation. The gradient image is binarized by the binarization method. According to the path characteristics of UAV, the image recognition results are obtained by combining the statistical average method and iterative threshold selection method.

2 Image Preprocessing of UAV Tracking and Navigation Path

Before the UAV tracking navigation path image recognition, it is necessary to preprocess the image to improve the image quality, which can effectively reduce the randomness of the UAV tracking process. The tracking navigation path image can be obtained through the acquisition card, and the noise interference factors it receives directly affect the accuracy of the recognition and analysis results. Therefore, the image preprocessing can remove the background and redundant parts that are not recognized targets. Image preprocessing includes image graying, image scaling, image filtering, image edge detection and image binarization.

2.1 Image Feature Extraction and State Analysis

Image recognition is the main technology of UAV tracking navigation path image, and feature extraction is the basis of recognition. Image features include color and shape, among which shape features are the most intuitive and can describe basic attributes. In the image shape features, the image matrix has rotation and movement invariant attributes, which can be used as a recognition feature. In the binary image $g(m, n)$, the expression of the image area $(a + b)$ phase matrix and the center matrix is as follows:

$$E_{ab}=\sum m^b n^a g(m, n) \tag{1}$$

$$O_{ab}=\sum \left(m - m'\right)^n \left(n - n'\right) g(m, n) \tag{2}$$

In the formula, $\left(m', n'\right)$ represents barycentric coordinates respectively. The central matrix can be further defined by Formula (2). In the discrete case, for the preprocessed image, in the analysis process, it is necessary to first obtain the target recognition invariant moment feature vector, which is insensitive to noise and has good stability. It is a feature vector, which can effectively identify and analyze the UAV tracking navigation path image.

In the field of target recognition, the background of UAV tracking and navigation path image recognition is extremely complex, and the illumination will change at any time, so the recognition accuracy cannot reach a satisfactory level. Therefore, the use of support vector machine has large operation, strong generalization and good real-time performance [7]. When the support vector machine is used for recognition and classification, the core function needs to be selected first. For the mapping of the sample space, it needs to be converted to the space first, and then the hyperplane classification is carried out; According to the classification results, the appropriate hyperplane can be selected and the support vector can be given; The problem of classification and recognition is transformed into the problem of solving functional relations. Because there are many kinds of UAV tracking and navigation path images, different kinds of classifiers need to be combined into multiple classifiers. Although the training time is extended, the constraints between different samples are fully reflected. Image moment features can be used as the input vector to identify and analyze images and ensure the accuracy of recognition results [8].

In the process of image acquisition, the poor focusing of UAV will introduce noise, which seriously affects the image quality and brings difficulties to image recognition and analysis [9]. The UAV obtains the video image from the received image, and first performs preprocessing. The image after preprocessing can improve the image quality, segment the target area where the substation equipment is located, and obtain the main characteristics of the UAV tracking and navigation path image recognition after binary processing. Image invariant moment features are mainly reflected in different image types. Input the feature vector into the support vector machine, and the UAV tracking and navigation path can be recognized after training [10]. The application of artificial intelligence method can greatly improve the level of automatic recognition. Based on this, an image recognition model is established.

2.2 Image Edge Enhancement Processing

The Laplacian operator can be used to process the image to improve the image quality. The binary processing method can make the image extremely simple and facilitate the calculation speed. The Formula (3) can be used to convert image $g(m, n)$ into a binary image, as shown below:

$$g(m, n) = \begin{cases} 0 \; g(m, n) < \alpha \text{ or } g(m, n) > \beta \\ 1 \; \alpha \leq g(m, n) \leq \beta \end{cases} \tag{3}$$

In Formula (3), α and β represent upper and lower thresholds respectively. The upper and lower thresholds α and β are determined according to the image attributes. If the gray image histogram is roughly similar to the distribution form of the operating conditions, it can be explained that the image contains both the target and the background, as well as the determined appropriate threshold. Only in this way can the target to be identified and the background be effectively separated. Use the Laplacian operator to enhance the image edge.

2.3 Image Scaling Grayscale Processing

When video files are used as image sources, the decoded bitmap frame sequence is a true color image. For unified processing, the acquisition format of the image acquisition card is also set to true color 24 bits. Before further preprocessing, the obtained image is grayed. The graying formula is:

$$f = \frac{R}{\lambda_1} + \frac{G}{\lambda_2} + \frac{B}{\lambda_3} \tag{4}$$

In Formula (4), $\lambda_1, \lambda_2, \lambda_3$ represents three color component coefficients respectively; R. G and B are the red, green and blue gray scale components of the original image, both of which range from 0 to 255. In order to avoid occupying CPU time by allocating space each time during image processing, static storage space is used to store image data, which requires that the size of the image is fixed. Therefore, it is necessary to reduce the width and height of the image to a fixed value. The size of the image after scaling is 400 * 300. The method used for image scaling is bilinear interpolation. Because the image noise collected by the video capture card is relatively large, it is necessary to denoise the zoomed gray image. These noises are random noises [11].

2.4 Image Denoising

Image recognition technology firstly monitors the recognized image, and intercepts a frame of image as the input sample for processing [12]. Extract the features of the captured input image, and output the features in digital form. However, due to the noise interference of the extracted features, there are many noise scatters in the feature extraction results. Therefore, image recognition technology needs to be used for denoising.

Because the images transmitted by UAV tracking navigation path have different frequency characteristics, the recognition result is the result of multiple frequency superposition and coupling. In addition, when the UAV is interfered by noise in the process of tracking navigation path image recognition, a small frequency pulse will appear, and these images are easily obscured by background noise and large amplitude natural frequency components. Therefore, a single wavelet can not meet the needs of image decomposition. Therefore, in various cases, we must use multi wavelet to extract the complementary features of the image from multiple angles to improve the recognition accuracy. Different wavelet bases have great differences in the analysis results of images. In order to get complete reconstruction from wavelet decomposition, appropriate wavelet bases must be selected. By selecting the appropriate wavelet base, the image can be processed by wavelet transform to achieve the purpose of noise reduction. Low pass filter is usually used in conventional noise reduction methods.

When wavelet transform is used for noise reduction, since the energy of noise components is mainly concentrated in the wavelet transform area, the real image is all concentrated on the wavelet coefficients. The appropriate wavelet coefficients are determined by the threshold setting method, estimated by the threshold function mapping method, and reconstructed to reduce the noise. For any integrable function, its wavelet transform can be expressed as:

$$\psi(\varphi, L) = |\varphi| \int_{-\infty}^{+\infty} f(t)\psi^*\left(\frac{tL}{\varphi}\right)\mathrm{d}t \tag{5}$$

In Formula (5), φ represents the wavelet scaling factor; L represents the relative displacement factor; ψ^* represents the conjugate wavelet basis; t represents the transformation time.

For discrete noisy image computation, it can be expressed as:

$$g'(t) = r(t) + \varsigma r'(t) \tag{6}$$

In Formula (6), $r(t)$ represents an effective image; $r'(t)$ represents a noise image; ς represents a noise coefficient. Using wavelet analysis method, select an appropriate threshold, when the discrete noise image coefficient is less than this threshold, it can be regarded as 0; When the discrete noise image coefficient is greater than this threshold, it indicates that the image is a noise image and needs to be processed immediately, thereby completing the noise reduction processing of the image.

3 Path Image Recognition Technology Based on ResNet

The deep residual network Res-Net is a convolution-based network model structure that improves the accuracy by deepening the network layer. The ResNet model parameters are reduced by replacing a large convolution kernel with several small convolution kernels, while increasing the number of nonlinear activation functions of the ResNet network model, which reduces the computational complexity of the ResNet model and improves computational efficiency. If the size of the input and output feature maps in the convolutional layer is the same, the number of filters remains unchanged; if the size of the feature map becomes half the original size, the number of filters will be doubled, and the pooling layer step size of the feature map is set into 2. The main difference between ResNet models with different depths is the number of layers of convolutional layers. Some image recognition classification projects can achieve high recognition accuracy because their network models have a deep network structure. The network structure extracts richer features by increasing the number of convolutional layers, so that the recognition effect will be better. However, if it is just a simple stacking network, the gradient will disappear. To solve this problem, the deep residual network adopts standard initialization and regularization to retain all feature variables, so that the network model will not lose accuracy. When the network structure is deeper, the phenomenon of overfitting will occur, which will lead to poor recognition effect. For the above problems, the ResNet network structure is combined with the residual learning method and optimizes the deep network to improve the recognition accuracy and learning speed.

3.1 Construction of Recognition Model Based on ResNet

The characteristics of small differences between classes and large differences within images determine the importance of detailed features in the recognition process. In order to solve the problem that the recognition accuracy is difficult to improve due to this feature, an image recognition model that can enhance the detailed features is proposed. According to the characteristics of the image, it aims to enhance the ability to extract features from local to global, which is composed of deep shallow shared attention modules and enhancement blocks. The deep shallow shared attention modules improve the feature extraction effect by learning the weights of different channels of the image, and embed the network from shallow to deep to maximize the enhancement of global features; The enhancement block is used to improve the ability to extract the local details of the image. It uses asymmetric convolution to construct the enhancement core to replace the ordinary convolution core. At the same time, it establishes the connection between different layers of networks through the idea of jump connection, and realizes local feature enhancement while reducing degradation. Based on this, the recognition model based on ResNet is constructed as shown in Fig. 1.

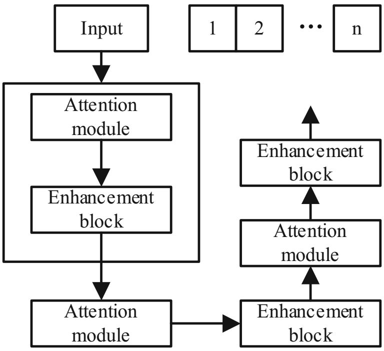

Fig. 1. Recognition model based on ResNet

When the initial optimal and user mapping becomes the optimal mapping by integrating the nonlinear layer with other mappings, if the original mapping is not as good as the residual mapping, then under special circumstances, the residual may be set to 0, which is easier to approximate the mapping to another mapping. Add an object that can realize identity mapping in the feedforward network. In this way, the number of network parameters will not change, and the calculation will not be cumbersome. The network model can still use back propagation for end-to-end training.

3.2 Detail Visualization Rendering

Due to the huge number of point clouds obtained by the UAV tracking the navigation path, it is not easy to achieve image rendering by simply loading all at one time. Therefore, a multi-level detail rendering method based on the ResNet algorithm is proposed. The ResNet algorithm regards each group of networks mapped by the residual network as a building block, which is defined as:

$$D = g(v, t) \tag{7}$$

In Formula (7), v represents the input vector, D represents the output vector, and $g(v, t)$, as the residual mapping to be trained, has certain requirements for the dimensions of v and g in the formula, which need to be the same, otherwise, linear projection needs to be added to modify the configuration dimensions. There are some differences between convolutional neural network and ResNet algorithm. ResNet algorithm directly connects the network input with other layers in order to make the network layer learn residuals better. The convolution of convolution neural network will cause data loss during training, which will lead to the decline of recognition accuracy; In this case, ResNet algorithm bypasses the input information to the output in the convolution layer to ensure data integrity, reducing some tedious work and improving the accuracy of the algorithm. The ResNet algorithm has the following network design principles: ① Layers with the same feature map size have the same convolution kernel; ② When entering the pooling layer, the feature map size becomes half of the original size. In the residual network, the connection of dotted and solid lines is used to distinguish whether the network dimensions match. If there is a dimensional mismatch in the network, there are two options: the first is to directly use zero padding to add dimensions, and the second is to multiply the i matrix and project it into a new space.

In the multi-level detail modeling, the hierarchical structure of the multi-resolution octree is used as shown in Fig. 2.

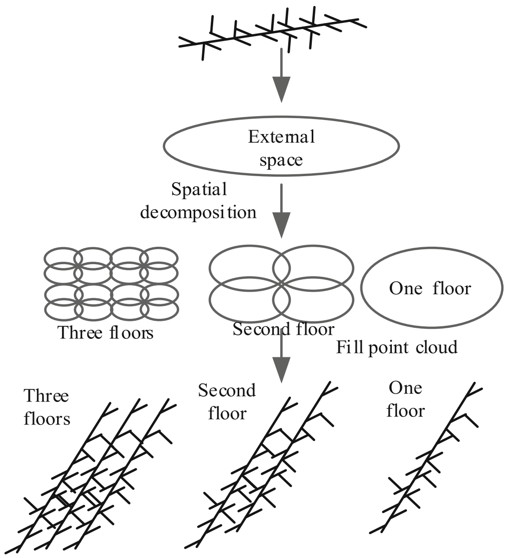

Fig. 2. Schematic diagram of multi-resolution octree construction

The octree is usually used to segment the leading point cloud. The basic idea is: first, get a minimum outer square that contains all the point cloud data and corresponds to the root node of the octree. Then, from different angles, divide the cube into eight smaller cubes, and the cube can continue to split until there is no more cube, or the size of the cube is opposite to the shape of the leaves of the octree. Select the node data to be rendered within the current leaf node range, and automatically adjust the rendering degree of each node through the octree. The image output coordinates for any node can be expressed as:

$$x = J_i \frac{x_{\max} - x_{\min}}{2^{z_i - 1}} \tag{8}$$

$$y = J_i \frac{y_{\max} - y_{\min}}{2^{z_i - 1}} \tag{9}$$

In the above, $x_{\max}$ and $x_{\min}$ represent the maximum and minimum values of the horizontal coordinates respectively; $y_{\max}$ and $y_{\min}$ represent the maximum and minimum values of the vertical coordinates respectively; z_i represents the position of the i-th node in the octree; J_i represents the index position of the i-th node. The detailed data to be rendered is stored in the vertex buffer, and the position and size are set by the vertex shader. Each unit must be determined in a three-dimensional space that can be displayed on the screen. If it cannot, it must be cropped. If the primitive is beyond the visible range, it will be removed. After clipping and deletion, the position coordinates of the vertices are transformed. At this point, the primitive has completed the next step of raster processing, and the coordinates of the actual pixel are obtained through raster element mapping. Since the current pixel has no color information, a shader is required for coloring. Finally, in the frame buffering process, the color information of each layer is written into the frame buffer, thereby completing the image visualization enhancement processing.

3.3 Image Degradation Improvement Processing

The attention residual module makes the attention module embedded in the deep and shallow layers of the network by combining the attention module and the enhancement module, and improves the learning effect of each stage of the network. The algorithm performs global pooling on the input feature map of size $A \times B$ and number C, and obtains the global information between C channels, which can be described as:

$$U = \frac{1}{A \times B} \sum_{x=1}^{A} \sum_{y=1}^{B} m(x, y) \tag{10}$$

In Formula (10), $m(x, y)$ represents the value of the feature map m of the C-th channel at the position coordinate (x, y). Secondly, in order to limit the complexity of the model and ensure the generalization ability of the model, U is excited twice to obtain the weight value of the feature map. Finally, the weights are applied to U, and the recalibrated feature map U is obtained, and the feature map U can be directly fed into the enhancement block. When the attention module is in the shallow layers of the network, it excites informative features in a class-independent manner and reinforces the shared universal features, and when the network is deep, the attention module tends to specialize and respond to different inputs in a class-specific manner. ResNet connects the above-mentioned attention module and enhancement block 4 times alternately, so that the deep and shallow layers of the network are embedded in the attention module, which not only stimulates valuable features, improves the learning effect, but also reduces the phenomenon of image degradation.

3.4 Image Recognition Process

The super pixel method is used to segment the UAV tracking and navigation path image to obtain the super pixel points. The gradient image is binarized using the binarization method, and appropriate thresholds are set for different application scenarios. When the threshold value is large, the pixel is regarded as an edge point; If it is below the threshold value, it is considered as a non edge point; The binary gradient image is segmented by morphological extension operator and erosion operator. The expansion operator is calculated as follows:

$$g' \oplus \gamma = \left\{ i|(\gamma)_i \cap g' \neq \emptyset \right\} \tag{11}$$

In Formula (11), g' represents the image to be expanded; γ represents the structural element. The calculation formula of the corrosion operator is:

$$g'' \oplus \gamma = \left\{ i|(\gamma)_i \subseteq g'' \right\} \tag{12}$$

In Formula (12), g'' represents the image to be eroded. The target image and the segmented image are masked to obtain the area where the UAV tracks the navigation path.

In order to further narrow the area, the adjacent nodes are moved, and the control node is at the minimum position of the domain gradient. Calculate all pixel values in the seed point field, update the node position information, cluster labels according to the node position information, and set the initial distance metric. After that, calculate the minimum distance from each node in the neighborhood to the pixel point, and if the total distance is less than the original distance, update the pixel mark to the current node, and the total distance is equal to the initial distance. After the node traversal is completed, the node is located based on the classification mark of the pixel point. In the grayscale image, the grayscale value and coordinate average of the pixels in each classification mark are used as new nodes; the optimization is repeated until the cluster marks no longer change, or the maximum number of repetitions is reached, which improves the connection degree. The superpixels in the image are scanned by a left-to-right, top-down method, and the connected components of the separated single pixel points or regions that are too small are classified.

Fully consider the characteristics of UAV tracking and navigation path, and use the statistical average method to obtain the average interference amplitude. The formula is:

$$\mu = \frac{\sum_{i=1}^{i} d(i)}{\sqrt{Q\left[|d|^2\right]}} \tag{13}$$

In Formula (13), d represents the distance between the target and the background; Q represents the energy parameter. The formula for calculating the average interference phase is:

$$\vartheta = \theta\left(\sum_{i=1}^{i} d(i)\right) \tag{14}$$

In Formula (14), θ represents the phase angle. On the two-dimensional characteristic plane of interference, since the interference amplitude and phase are different dimensional physical quantities, the interference characteristic distance defined by the Gaussian kernel function can be expressed as:

$$d = \exp\left(p(H - H_s) \oplus G^p\right) \tag{15}$$

In Formula (15), H and H_s represent the potential target interference feature and background interference feature respectively; G^p represents the bandwidth matrix; p represents the interference eigenvector norm; $\oplus$ represents the vector outer product. The bandwidth matrix can normalize physical quantities of different dimensions and is suitable for measuring the distance between the target and the background. In the interference space, if the difference between the target and the background is obvious, the farther the vector is from the background feature, the more it represents the path image.

Image recognition usually has many different methods for different applications. According to the characteristics of the target and background of the UAV image, grayscale segmentation is used to achieve image recognition. At the same time, an interface is designed to design different recognition algorithms for different target characteristics, so that the system has good scalability. The selection of the grayscale segmentation threshold adopts the iterative threshold selection method. The iterative formula is:

$$d_{i+1} = \frac{1}{2}\left[\frac{\sum_{h=0}^{d_i} n_h h}{\sum_{h=0}^{d_i} n_h} + \frac{\sum_{h=d+m_i}^{m-1} n_h h}{\sum_{h=d_i+1}^{m-1} n_h}\right] \tag{16}$$

In Formula (16), m represents the number of gray levels; n represents the number of pixels whose gray value is h. Take the initial threshold as 0.5 m, and use this iterative formula until the threshold does not change, that is, when the threshold converges to a certain value, it stops, and this threshold is the optimal segmentation threshold. After threshold segmentation, the image is divided into two parts: background and target, and then the center of gravity of the target is taken as the position of the recognition target. The calculation method of the center of gravity is as follows:

$$x_0 = \frac{\sum_{h=1}^{k} x_h}{k} \tag{17}$$

$$y_0 = \frac{\sum_{h=1}^{k} y_h}{k} \tag{18}$$

In the above formula, k represents the total number of pixels occupied by the target, thereby obtaining the recognition result.

The recognition model is built based on ResNet, the details of the pre-processed image are visualized and rendered, and the global pooling processing is carried out to improve the image degradation phenomenon. The image recognition is completed by the binary method and the statistical average method.

4 Experimental Analysis

4.1 Experimental Setup

The UAV azimuth target positioning, tracking and navigation system is mainly divided into two subsystems: UAV visual navigation system and UAV flight control system. The visual navigation system includes two parts: camera and raspberry pie. The airborne camera captures the target image information; Raspberry Pie is an independent visual processing system. It not only has image processing function, but also can process the captured image information in real time to obtain the information such as target orientation and angle; At the same time, it can also complete the estimation of target motion state, and then provide reference information for UAV flight. The flight control system carries out control operation according to the information given by the visual navigation system, and controls the UAV to accurately complete target tracking. UAV navigation path tracking and navigation system, as shown in Fig. 3.

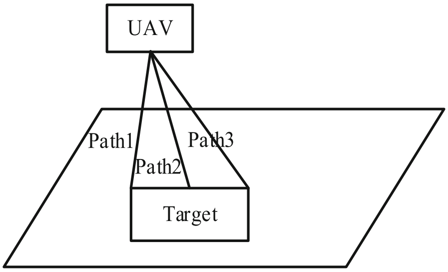

Fig. 3. UAV navigation path tracking and navigation system

As shown in Fig. 3, the UAV is equipped with a visual navigation system to track the ground moving targets in real time. The Raspberry Pi and the camera are used as the visual navigation system, and the Raspberry Pi and the onboard camera are installed directly under the UAV as a visual navigation system to detect the moving targets on the ground in real time during the flight of the UAV. Connect the Raspberry Pi and the UAV flight controller through the serial port, the visual navigation system obtains the deviation information of the target and completes the angle solution, and provides navigation information for the flight controller through the target estimation strategy and control strategy. According to the navigation information provided by the visual navigation system, the flight controller is used to perform control operations to realize the real-time tracking of the target by the UAV. The overall block diagram of the experimental device is shown in Fig. 4.

The designed visual navigation system mainly includes a Raspberry Pi and an external camera. The camera is used to collect real-time images. As an airborne image processing module, the Raspberry Pi is used to perform target detection and image processing on the collected images.

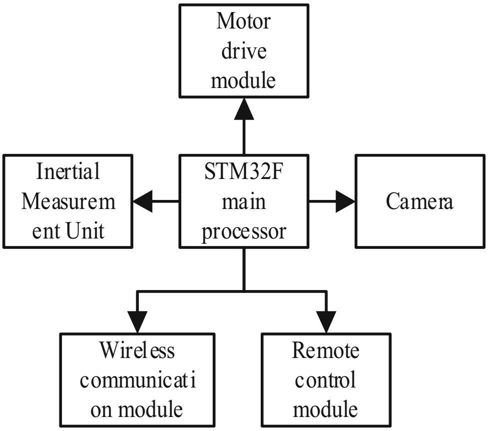

Fig. 4. Overall block diagram of the experimental device

4.2 Experimental Area and Data

The image of the drone tracking the navigation path was taken at 8:00 on April 4, 2020, and the weather was clear but foggy. Using a fixed-wing UAV manufactured by a company, the UAV can quickly and efficiently generate data such as topographic maps, orthophoto maps, and ground digital models. The fuselage is equipped with a Sony DSC-WX220 camera, the lens of which is Sony's G lens.

The image of the UAV tracking the navigation path is shown in Fig. 5.

Fig. 5. UAV tracking and navigation path annotation results

Taking the target annotation results shown in Fig. 5 as the research object, the experimental verification analysis is carried out.

4.3 Experimental Results and Analysis

In the process of identifying image information, it is not only limited to the collected image data, but also the problem of image distortion. Therefore, the convolutional neural network method, the YOLO v3 algorithm, the Cascade R-CNN method and the ResNet recognition technology were used to compare and analyze the image recognition results of the UAV tracking navigation path, as shown in Fig. 6.

It can be seen from Fig. 6 that the images extracted by convolutional neural network method, YOLO v3 algorithm and Cascade R-CNN method are not comprehensive, resulting in partial loss of information. However, the images extracted by ResNet recognition technology are comprehensive, consistent with the selected areas shown in Fig. 5, and can obtain all image information.

In order to prove that the researched technology has a good anti-interference effect, the signal-to-noise ratio of the interference images of these four methods is compared and analyzed again, and the comparison results are shown in Fig. 7.

It can be seen from Fig. 7 that the convolutional neural network method, YOLO v3 algorithm and Cascade R-CNN method have low signal-to-noise ratio, while the ResNet recognition technology has introduced the ResNet technology to improve the signal-to-noise ratio of the interfering image, of which the highest signal-to-noise ratio is 30.5dB, enriching the important information of the image and obtaining the image enhancement effect.

Based on this, these four methods are used to compare and analyze the image recognition accuracy, as shown in Table 1.

It can be seen from Table 1 that the image recognition accuracy using the convolutional neural network method, the YOLO v3 algorithm, and the Cascade R-CNN method is the highest at 0.52, 0.56, 0.69, and 0.96. The highest image recognition accuracy using the ResNet method is 0.96, indicating that the image recognition accuracy using this method is high.

(a) Convolutional neural network method

(b) YOLO vz algorithm

(c) Cascade R-CNN method

(d) ResNet recognition technology

Fig. 6. Comparative analysis of image extraction results by different methods

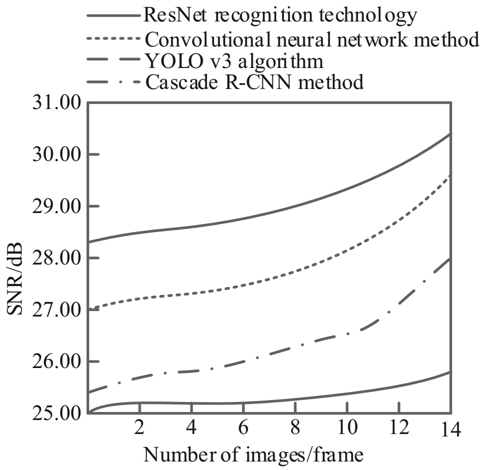

Fig. 7. Comparative analysis of signal-to-noise ratio of different methods

Table 1. Analysis of recognition accuracy and loss degree of recognition process by different methods

Iterations/time	Convolutional neural network methods	YOLO v3 algorithm	Cascade R-CNN method	ResNet
200	0.39	0.49	0.65	0.92
400	0.32	0.48	0.67	0.93
600	0.48	0.45	0.69	0.93
800	0.45	0.49	0.67	0.95
1000	0.45	0.53	0.69	0.96
1200	0.52	0.56	0.69	0.93

5 Conclusion

ResNet-based UAV tracking and navigation path image recognition technology uses the actual image to understand the position and attitude information. The technique avoids overfitting, converges quickly, and the residual learning network makes deep network training easy. The highest image recognition accuracy is 0.96. The ResNet algorithm has good recognition effect and robustness, and has a certain promotion effect in other machine vision application fields. Although the ResNet algorithm achieves the expected image recognition effect, there are still shortcomings. The number of images used in this experiment is small and the relative noise is small. There may be some differences in the actual recognition process. In the future research, it is necessary to continuously improve the recognition effect of the proposed method to cope with the recognition of massive path images and the interference of greater noise.

References

1. Li, G.: Design of UAV navigation system based on image recognition. J. Agricult. Mech. Res. **43**(01), 114–118 (2021)
2. Xue, S., Zhang, Z., Lv, Q., et al.: Image recognition method of anti UAV system based on convolutional neural network. Infrared Laser Eng. **49**(07), 250–257 (2020)
3. Zhao, X.: The study and application of a UAV image recognition technology system. China Rural Water Hydropower **05**, 195–200 (2022)
4. Liu, Y., Yang, F., Hu, P.: Parallel FPN algorithm based on cascade R-CNN for object detection from UAV aerial images. Laser Optoelectron. Prog. **57**(20), 302–309 (2020)
5. Lou, X., Huang, Y., Fang, L., et al.: Measuring loblolly pine crowns with drone imagery through deep learning. J. Forest. Res. **33**, 1–12 (2021)
6. Nemer, I., Sheltami, T., Ahmad, I., et al.: RF-based UAV detection and identification using hierarchical learning approach. Sensors **21**(6), 1947–1955 (2021)
7. Wan, J., Ma, L.: Optimal generation method of GPS navigation UAV transportation path. Comput. Simulat. **38**(11), 37–41 (2021)
8. Wang, Z., Peng, H.: Voronoi diagram-based optimized traversal path of unmanned aerial vehicle for acquisition data of nodes in wireless sensor networks. Mod. Mach. Tool Automat. Manuf. Techniq. **2021**(09), 67–70+74 (2021)
9. Shang, K., Zheng, X., Wang, M., et al.: Image semantic segmentation-based navigation method for UAV auto-landing. J. Chin. Inertial Technol. **28**(05), 586–594 (2020)
10. Jia, H., Wang, L., Fan, D.: The application of UAV LiDAR and tilt photography in the early identification of geo-hazards. Chin. J. Geolog. Hazard Control **32**(2), 60–65 (2021)
11. Triet, N.A., Phuong, N.D., O'Regan, D., et al.: Approximate solution of the backward problem for Kirchhoff's model of parabolic type with discrete random noise. Comput. Math. Appl. **80**(3), 453–470 (2020)
12. Yuan, H., Zhang, S., Chen, G., et al.: Underwater image fish recognition technology based on transfer learning and image enhancement. J. Coast. Res. **105**(SI):124–128 (2020)