Journal of Computers Vol. 35 No. 3, June 2024, pp. 345-361
doi: 10.53106/199115992024063503023

Research on Obstacle Avoidance Technology for Unmanned Aerial Vehicles Based on Panoramic Visual Perception

Xiao-Yan Jiang[1,2*], Mei Han[1], Jun-Kai Zhang[1],
Xiao-Fei Wu[1], Xiao-Yang Zhang[1]

[1] Hebei Institute of Mechanical and Electrical Technology,
Xingtai City 054000, Hebei Province, China
{xiaoyan56782024, meigui9992024, junkai6662024,
xiaowei2323, xiaoyang86863}@126.com

[2] Hebei Province Mechanical and Electrical Equipment Intelligent Sensing and Advanced Control R&D Center,
Xingtai City 054000, Hebei Province, China

Received 1 May 2024; Revised 1 June 2024; Accepted 19 June 2024

Abstract. With the continuous increase of market demand, the application of low altitude rotary wing unmanned aerial vehicles is becoming increasingly widespread, and the requirements for active obstacle avoidance ability and endurance time are also increasing. This article focuses on the active obstacle avoidance technology of drones. The panoramic camera is installed on the drone. For the collected multi camera images, the first step is to perform stitching processing. The improved SURF algorithm is used to process the overlapping defects after image stitching. Then, the image is weighted and fused using the arc function weighted fusion image stitching algorithm to remove the stitching gaps. The fused image is used as input, and the improved YOLOv8 model is used as the target recognition model. After analyzing the basic performance of the model, in order to improve the algorithm's lightweight level, the LAD calculation rule is integrated into the algorithm. Then, the EMA attention mechanism is added to improve the model's recognition ability for specific obstacles, and depth separable convolution is added to enhance the algorithm's ability to extract target features. Finally, an experimental environment was established, and through simulation experiments, the method proposed in this article improved the average recognition accuracy of obstacles by 2.3% while ensuring the endurance of the drone.

Keywords: UAV, panoramic vision, YOLOv8, EMA

1 Introduction

In recent years, with the rapid development of aviation technology, the working space of drones has gradually expanded from mid to high altitude to low altitude environments, and the unstructured low altitude environment has also brought new challenges to the autonomous perception and obstacle avoidance technology of drones At present, obstacle perception of drones in low altitude environments mainly relies on active sensors represented by millimeter wave radar, LiDAR, ultrasound, and passive sensors represented by visual detection such as optoelectronics and infrared Due to its advantages such as light weight, small size, and abundant information, visual sensors are increasingly favored in drone obstacle avoidance technology. Various countries and scholars have made achievements in the visual obstacle avoidance of drones. For example, the School of Engineering at the University of Porto has developed a small fixed wing drone obstacle detection system, which includes ultrasonic sensors, laser rangefinders, and LiDARs [1]. Through the fusion of sensor data and the search ability of genetic algorithms, target recognition and avoidance can be achieved. Fr¨anken, focuses on obstacle recognition and avoidance during drone landing, using telemetry data exchange, GNSS sensor fusion, infrared detection, and other methods to guide drone autonomous landing [2].

For UAV visual obstacle avoidance and in the field of UAV panoramic vision, domestic scholars are also conducting corresponding research. The research goal of this article is UAV panoramic visual obstacle avoidance technology.

* Corresponding Author

2 Related Work

The researchers at the University of Porto and the University of Poland, as typical unmanned aerial vehicle visual obstacle avoidance technologies, share the common feature of integrating visual sensors with other sensors to achieve high-precision perception of the environment and threats However, for the most widely used rotor unmanned aerial vehicles in low altitude environments, the above-mentioned multi-sensor fusion scheme is difficult to achieve a balance between performance, cost, power consumption, and payload. In other words, although multiple sensors can achieve multi-dimensional target perception and improve perception ability, the resulting power increase and algorithm complexity increase. Therefore, it is necessary to consider designing a perception method that relies solely on visual sensors, in order to reduce sensor costs and drone power consumption while ensuring obstacle avoidance capability. Many scholars have also conducted corresponding research on this.

Feng Mei from Guilin University of Electronic Science and Technology uses binocular depth cameras to measure and track the distance of the target in real time, locate the position information, and then convert and output the three-dimensional position data of the drone. The visual data is imported into the YOLOv5s algorithm, and then an optical fast recognition, positioning, and tracking system for the drone is established. The system meets the requirements of high-speed tracking in terms of computational speed [3].

Kaiyuan Tan configured binocular cameras on the basis of quadcopter drones, and then integrated various algorithms such as ORB-SLAM mapping, positioning and obstacle avoidance, and autonomous navigation using ROS system as a carrier to design and develop binocular SLAM drones. After the complete construction of the map, different algorithms were used to plan and screen the optimal path solution of the drone from the starting point to the endpoint, thereby achieving the autonomous navigation function of binocular SLAM drones [4].

Yi Cheng from Tianjin University of Technology uses binocular vision cameras and incorporates an improved YOLOv4 tiny object detection algorithm. She then uses the depth of the solution and the pose information of the drone as the basis for determining the field of view scale of the collision avoidance area, and combines the pixel position of the object detection to achieve visual collision avoidance of the drone [5].

Xunju Ma from North China University of Water Resources and Hydropower analyzed in detail the structure and principles of binocular vision technology, panoramic vision technology, and binocular panoramic vision technology, and proposed the process of obstacle avoidance technology for unmanned aerial vehicles based on binocular panoramic vision and the future development direction of this technology, pointing out the direction for the application of panoramic technology in the field of unmanned aerial vehicles [6].

Yuanyuan Liu from Jilin Agricultural University proposed a fast stitching method for agricultural aerial panoramic images based on optimized SIFT (scale in variant feature transformation) algorithm. Firstly, preprocess the concatenated images, accurately calculate the concatenation conversion model, and then use a multi-resolution fusion algorithm based on the best stitching line for image fusion to obtain panoramic concatenated images. The experimental results show that the information entropy, average gradient, and image contrast of the image stitching method have been significantly improved, and the stitching time has been shortened by more than 90%. The image stitching processing method has good reference significance for the image processing process in this paper [7].

Buyun Wang from Zhengzhou University proposed a target recognition method and path planning rule that integrates point clouds and panoramic images, using a multi-source sensor consisting of LiDAR and panoramic cameras. The innovation lies in the use of spatial geometric information in point cloud data and semantic information in panoramic images to improve the accuracy of target recognition. Firstly, it is necessary to complete the analysis of the panoramic image, obtain the data coordinates and two-dimensional pixel information in the image. Then, after projecting the laser point cloud to generate a panoramic depth map, the point cloud information is integrated into the analyzed panoramic image. For the problem of viewpoint occlusion and occlusion generated by the camera during shooting, by analyzing the continuity and integrity of the target in three-dimensional space, the candidate point cloud is subjected to secondary clustering, and finally the classification of the target is completed. The experimental results show that the recognition accuracy for obstacle targets is 96.64%, 92.68%, and 90.74%, respectively. However, this method mainly involves unmanned aerial vehicles operating at ultra-low altitudes, and the recognition database mainly includes ground targets. Based on this method, it can be considered to increase the database of aerial targets [8].

Weibiao Chen from Zhejiang Normal University proposed a new object detection method, DSM-YOLO v5, based on a deep separable multi head network structure, to address the issues of high miss rate, low detection success rate, and large model size in traditional drone aerial image object detection algorithms. By adding a small object detection head with a size of 160 × 160 on the YOLO v5 network structure and residual connecting it with the high-level network, the small object detection ability can be improved. At the same time, the Conv module introduces a depthwise separable convolution algorithm, replacing the ordinary convolution with depthwise separable convolution, which can effectively reduce the number of network parameters and reduce the model volume. The experimental results indicate that object detection based on DSM-YOLO v5 network structure is effective mAP@0.5 It is 36.8%, an increase of 3.6 percentage points compared to YOLO v5s, and the overall number of parameters and model volume decrease by 21.1% and 19.7% compared to YOLO v5s. The improvement plan for identifying the model has practical reference significance [9].

In terms of image stitching, Ying Chen from Nanjing University solved the problem of defects when shooting a scene with a large field of view and high resolution, that is, ordinary cameras cannot meet two targets and can only form a scene with a large field of view and low resolution or a high resolution and small field of view. She used four images to stitch together, with some redundancy between each adjacent image, and then concatenated and fused them to achieve the effect of a large field of view and high resolution [10].

In summary, there are relatively complete processing methods for multi-sensor obstacle detection, fast image stitching, and processing in unmanned aerial vehicles. Based on the above research results, this paper further innovates and optimizes them.

With the continuous development of artificial intelligence technology, the enormous role of deep learning in the perception process of drone obstacle avoidance is constantly emerging. By using improved visual recognition algorithms, the recognition ability and speed of obstacles can be improved while reducing the number of sensors. This enables drones to recognize more obstacle features and have stronger environmental adaptability, enabling them to achieve real-time perception of obstacles when equipped with multi camera cameras Based on the above analysis, this article combines deep learning methods with panoramic vision technology to propose an obstacle visual perception method to achieve real-time comprehensive perception of the types, sizes, and relative positions of obstacles by drones.

Therefore, the overall idea of the perception method proposed in this article is as follows, and its overall diagram is shown in Fig. 1.

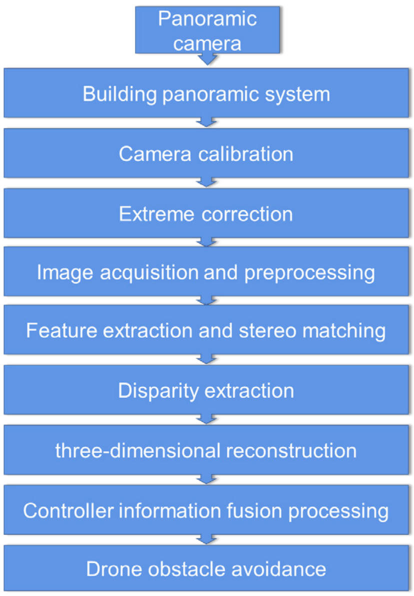

Fig. 1. Overall diagram of panoramic perception

1) The obstacle images captured by the panoramic camera are used for offline training of the deep neural network model, so that the trained model can recognize the type of obstacle from the images captured by the online camera and detect the pixel position and size of the obstacle in the image. Then, it enters the following online perception process;

2) The left camera of the panoramic camera uses a deep learning model to detect and recognize obstacles in the camera image and perform target tracking, obtaining real-time information on the types of obstacles and their

pixel positions and sizes in each frame of the image;

3) By calibrating, correcting, matching, and 3D distance restoring the panoramic camera, the three-dimensional reconstruction of the panoramic camera image is performed to obtain the three-dimensional position information of obstacles relative to the calibrated camera in the drone flight environment;

4) The obstacle information detected by the deep learning model is fused with the three-dimensional position information obtained by the panoramic camera to obtain the type, shape, contour size, and spatial position information of obstacles in the environment relative to the camera, thereby achieving real-time perception of the flying environment of low altitude small unmanned aerial vehicles [11].

In order to fully implement the description of the UAV panoramic perception obstacle avoidance strategy in this article, it is divided into the following chapters. The first and second chapters mainly introduce the research ideas and background of this article, and analyze the research results of some existing scholars. Chapter 3 mainly discusses the method of obtaining panoramic images of drones, using an improved image fusion algorithm to fuse the images, and then analyzing and further processing the graphics through a weighted fusion algorithm. Chapter 4 introduces an improvement plan for the target recognition model, which reduces model size and improves model accuracy by adding attention mechanisms and other means on the basis of YOLOV8. Chapter 5 conducted experimental simulations to verify the scientificity of the algorithm proposed in this paper.

3 Panoramic Image Acquisition

A multi camera panoramic image is a concatenation of images captured by a camera on a drone from multiple angles. After initial registration, there may be obvious stitching gaps or ghosting in the image. After improved image fusion algorithm processing, this problem can be solved to a certain extent. This article conducts experimental analysis on the improved SURF image registration algorithm and the weighted fusion algorithm based on the arc function used.

The panoramic vision system has a field of view angle of up to 360 ° and can obtain multi-directional information. It has broad applications in intelligent detection and robot control, intelligent vehicles, and geographic information acquisition. The panoramic vision system can be mainly divided into three types: multi camera panoramic, catadioptric panoramic, and fisheye lens panoramic. This article uses a multi camera panoramic system, so it is necessary to choose a suitable multi camera panoramic camera in drones. Multi camera panoramic is composed of multiple lenses, and the images collected by each lens are concatenated to obtain a panoramic view. The camera is selected as Ladybug3, a multi camera panoramic camera launched by PointGrey Company in Canada. The panoramic system consists of 6 cameras, including 5 on the side and 1 on the top. This camera system has functions such as color processing, camera correction, and panoramic stitching, and can capture over 80% of the spherical range. The system is widely used in robot navigation. The advantage of a multi camera panoramic system is that it produces better image quality; The disadvantage is that the data volume is large, and real-time performance is insufficient, requiring high hardware requirements such as processors and controllers. Based on the above shortcomings, this chapter mainly discusses the processing of the obtained panoramic images.

3.1 Improving Algorithms for Image Registration

Before detecting obstacles, drones need to create a dataset. When the dataset is large, the complexity of computation will increase significantly. The more feature points extracted, the longer the computation time required, making it difficult to ensure real-time performance. Therefore, it is considered to accelerate the calculation speed by constructing data indexes. The feature points extracted by the SURF [12] algorithm will be presented in a clustering form, and a tree structure will be used to construct a data index to achieve search space level classification and fast matching. This article uses a search engine based $k-d$ -tree [13], and the $k-d$ algorithm is the process of establishing a balanced binary tree, which is actually a recursive process.

The SURF algorithm has good adaptability and stability to changes in image translation, rotation, scaling, etc. It is a robust local feature point detection and description algorithm. Traditional algorithms generally use image downsampling methods to construct different features in the scale space, while the SURF algorithm constructs the scale space by gradually changing the size of the template. Since there is no need for downsampling the image, and multiple layers of images can be processed simultaneously in the scale space, the SURF algorithm has a faster processing speed.

The k-d tree strategy is as follows: each node in the structure is a binary tree with k-dimensional points, and all non leaf nodes can be viewed as dividing the space into two spaces using a hyperplane, which are half the structure of the original space. The subtree on the left side of the node represents the point on the left side of the hyperplane, and the subtree on the right side of the node represents the point on the right side of the hyperplane. In the history of hyperplane selection, the following method is used: each node is related to the dimension perpendicular to the hyperplane in the k-dimension. Therefore, if you choose to partition according to the x-axis, all nodes with x values less than the specified value will appear in the left subtree, and all nodes with x values greater than the specified value will appear in the right subtree. In this way, the hyperplane can be determined using this x-value, with its normal being the unit vector of the x-axis.

Based on the above two algorithm ideas, the matching process is as follows:

$$H\left(X,\sigma\right)=\begin{bmatrix} J_{xx}\left(X,\sigma\right) & J_{xy}\left(X,\sigma\right) \\ J_{xy}\left(X,\sigma\right) & J_{yy}\left(X,\sigma\right) \end{bmatrix}. \tag{1}$$

In the formula, $J_{xx}\left(X,\sigma\right)$ is the convolution of image I and Gaussian second-order partial derivatives at pixel X, and the SURF algorithm obtains the row column expression by using box filters of different sizes:

$$\det\left(H\right)=D_{xx}\times D_{yy}-\left(0.8\times D_{xy}\right)^{3}. \tag{2}$$

A weighting coefficient of 0.8 can balance the error caused by using a box filter approximation. After establishing the scale space, the extreme values of pixels in the scale space can be obtained by approximating the Hessian matrix, and compared with 26 other pixels in adjacent scale spaces to determine the set of feature points through non maximum values. Calculate the $Hear$ wavelet responses of all pixels in the x and y directions within the pixel area with a radius of 6σ around the determined feature points, and accumulate the wavelet response values a_x and a_y in the x and y directions in each $\pi/3$-sector area. Obtain the pixel vector $\left(f_i,h_i\right)$.

$$\begin{cases} f_i=\sum_i a_x+\sum_i a_y \\ h_i=\arctan\dfrac{\sum_i a_x}{\sum_i a_y} \end{cases}. \tag{3}$$

The schematic diagram of the main direction of matching feature points is shown in Fig. 2.

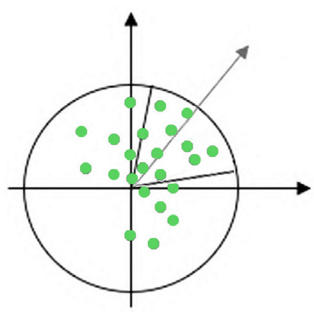

Fig. 2. Feature matching diagram

Then, $k-d$ tree is used to construct a matching index to improve the speed of feature point matching. The $k-d$ algorithm is mainly applied to nearest neighbor search and range search in multi-dimensional space. The data type description of each node in the $k-d$ tree is shown in Table 1.

Table 1. Description of data type for each node in Tree $k-d$

Variable name	Data type	Describe
Data	K-dimensional vector	Data, which refers to k-dimensional data points
Split	Int	Division, direction axis number perpendicular to the division hyperplane (range from 1 to k)
Left	$k-d$ tree	Left tree, a $k-d$ tree composed of all nodes in the left subspace located on the left side of the partition hyperplane

The indexing process of $k-d$ is as follows:

Input: A-dimensional spatial dataset, $T=\left[x_1,x_2,\cdots,x_m\right]$;

$$x_i=\left[x_i^{(1)},x_i^{(2)},\cdots,x_i^{(k)}\right], i=1,2,\cdots,M$$

Output: $k-d$ tree.

The matching index process is shown in Fig. 3.

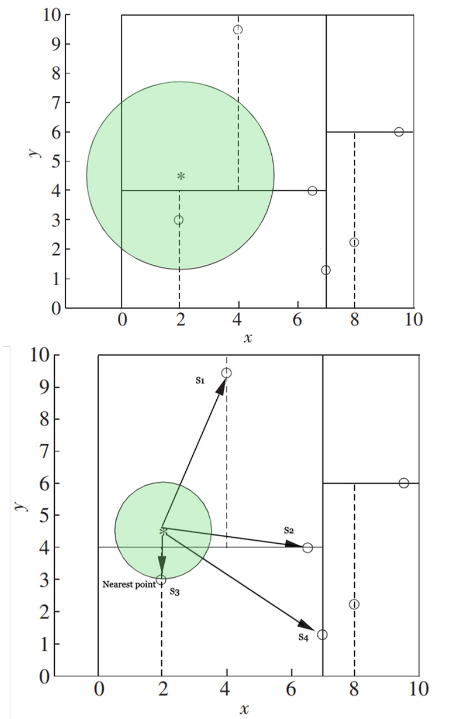

Fig. 3. The matching indexing process of $k-d$ -tree

The original image to be registered is shown in Fig. 4(a), and the image after registration using the SURF image registration algorithm is shown in Fig. 4(b).

(a) Before processing (b) After processing

Fig. 4. Comparison before and after image registration

In summary, based on panoramic vision sensors, drones mainly adopt three modules to achieve fast panoramic image stitching processing after obtaining panoramic images: rapid acquisition of multi camera vision images, improved SURF feature point extraction matching, and panoramic image stitching fusion. Based on the improved SURF image registration algorithm and the weighted fusion algorithm based on circular arc function mentioned in the above chapters, feature extraction is performed on the obtained multi view panoramic visual images to determine the overlapping parts of adjacent images. Matching and stitching fusion are performed based on the information of overlapping feature points, and finally the object processing is completed, preparing for further recognition of target pixels in the image.

3.2 Using Improved Algorithms for Image Fusion

In the process of panoramic image stitching in multi vision, multiple images are involved in stitching and fusion. When the overlapping area content is very complex, there will be stitching gaps or ghosting in the center. This article proposes a weighted fusion algorithm based on circular arc function to obtain nonlinear change weights [14]. The schematic diagram of the image stitching algorithm based on arc function weighted fusion is shown in Fig. 5.

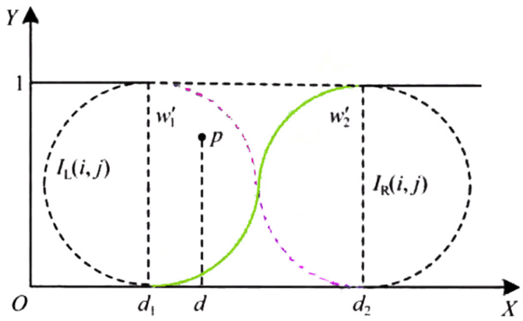

Fig. 5. Schematic diagram of image stitching algorithm

The weight calculation formula is as follows:

$$q_1' = \begin{cases} \dfrac{r+\sqrt{r^2-(d-d_1)^2}}{2r}, d \in [d_1, d_1+r] \\ \dfrac{r+\sqrt{r^2-(d-d_1-2r)^2}}{2r}, d \in [d_1+r, d_2] \end{cases} . \quad \textbf{(4)}$$

$$q_2^{'} = 1 - q_1^{'}. \tag{5}$$

Calculate the pixel values of overlapping area points using weights, and the process is as follows:

$$I(i,j) = q_1^{'}(i,j) I_L(i,j) + q_2^{'} I_R(i,j), (i,j) \in I_L \cap I_R. \tag{6}$$

On the basis of improving the SURF algorithm for image registration, this paper selects images with higher resolution as the images to be concatenated, and verifies the weighted fusion algorithm based on the arc function used. The verification results are shown in Fig. 6.

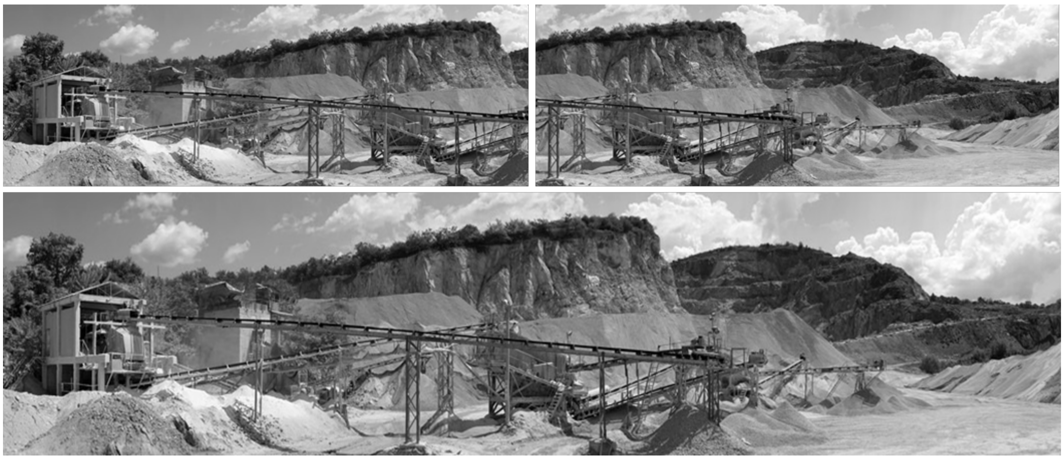

Fig. 6. Image fusion results

In summary, after using panoramic vision sensors, drones can obtain photos from multiple angles, and photos from multiple angles need to be cropped out of duplicate images before being stitched together. To achieve fast panoramic image stitching processing, the main method adopted in this article is the improved SURF algorithm, which is composed of three modules: fast acquisition of multi vision images, improved SURF feature point extraction and matching, and panoramic image stitching fusion. Based on the improved SURF image registration algorithm and the weighted fusion algorithm based on circular arc function mentioned in the above chapters, feature extraction is performed on the obtained multi view panoramic visual images to determine the overlapping parts of adjacent images. Matching and stitching fusion are performed based on the information of overlapping feature points, and finally the object processing is completed, preparing for further recognition of target pixels in the image.

4 Improve YOLOv8 Algorithm

By using the panoramic image processed in the previous chapter as the input image and inputting it into the visual recognition model, obstacles can be identified. This article uses YOLOv8 as the recognition framework. Due to the fact that the energy consumption of drones during mission execution determines the flight time, reducing the energy consumption of unmanned visual algorithm machines is the optimization objective of this article. To reduce energy consumption, algorithm lightweighting should be the main approach. Therefore, in order to improve the lightweighting level of YOLOv8 [15] algorithm, this paper proposes the im-YOLOv8 algorithm. In order to better improve the YOLOV8 algorithm, first analyze the existing structure of YOLOV8. YOLOv8 is an algorithm model released by Ultratics on January 10, 2023. The algorithm is accurately positioned as the next major update of the open-source YOLOv5 and currently supports image classification, object detection, and instance segmentation tasks. YOLOV8 inherits the good genes of YOLOV5 and has also made significant improvements in structure:

First of all, let's take a look at the Backbone module. This module still uses the original structural idea, only replacing the C3 module in YOLOv5 with the C2f module, which has improved the lightweight effect. In addition, YOLOv8 still uses the SPPF module commonly used in YOLOv5 and other architectures, and there are few improvements to Backbone.

Next, let's take a look at the PAN FPN module. YOLOv8 still uses the PAN concept, but has been appropriately upgraded. YOLOv8 has removed the convolutional structure in the PAN FPN upsampling stage from the original structure, and also replaced the C3 module with the C2f module. This structure has also been appropriately lightweight;

Finally, in the Anchor Free module, YOLOv8 abandons the previous Anchor Base structure and adopts the idea of Anchor Free, making feature recognition more flexible. At the same time, in the loss function and sample matching strategy used in the model, $Loss + CIOU$ is used as the classification loss, abandoning the previous IOU matching or unilateral proportional allocation methods, and improving the use of the loss function.

Analyzing the overall structure of YOLOV8, the structure is still quite complex and requires a long running time. In addition, YOLOV8 requires high device parameters. In addition, YOLOV8 is a universally applicable algorithm, and this article needs to place the recognition model in the drone processor. The algorithm's size and computational power consumption should be maximized to not affect the drone's endurance time. At the same time, considering the characteristics of the drone's recognition targets, this article mainly focuses on the following improvements to YOLOV8: scaling the input image to a fixed size, using LAD method for lightweight processing, and then adding a deep convolutional attention mechanism after the feature extraction layer. Through cross dimensional interaction, the feature information of small targets is maximally retained, improving the ability to detect small targets. The improved structure of YOLOv8 is shown in Fig. 7.

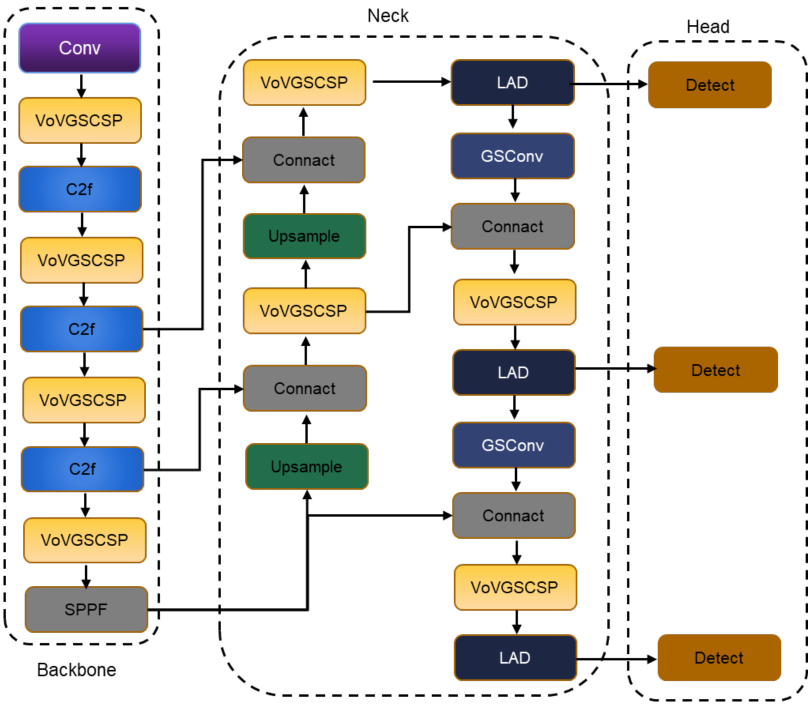

Fig. 7. Schematic diagram of improving YOLOv8 algorithm structure

4.1 Lightweight Adaptive Downsampling Weight LAD Method

The LAD method dynamically adjusts the weights of each position during the downsampling process through an adaptive weight downsampling adjustment mechanism. This design allows the algorithm to focus more accurately on key feature information, effectively reducing the loss of key information features. Compared to traditional convolutional architectures, the introduction of adaptive weight value adjustment mechanism significantly reduces the number of algorithm parameters, improves modeling ability and universality, and achieves the goal of lightweight processing of the algorithm. At the same time, the algorithm enhances the sensitivity of multi-scale and multi-dimensional targets, improving the accuracy and robustness of object detection. When applied to complex and diverse scenarios, the application of this method can better adapt to scene changes and effectively deal with various uncertainty factors, expanding the scope of application.

The workflow of the LAD method is divided into two steps: 1) Perform average pooling on the input feature map, with a convolution kernel size of 3 × 3 and a step size of 1, to obtain the downsampled feature map. Then, perform a 1x1 convolution on the feature map, which receives the pooled feature map. Use the realrange function to rearrange the dimensions of the attention map obtained after the convolution operation, and perform a softmax operation on the rearranged feature map to normalize the weights at each position to probability values. 2) Referring to the Focus structure in the YOLO v5 algorithm, the design idea of the Focus structure is to divide the input feature map into different blocks, rearrange and combine the blocks to increase the receptive field and extract richer features. Although the Focus operation can provide some performance improvement, it faces problems such as increasing the computational load and memory consumption of the algorithm, and may introduce unnecessary computational complexity. This article designs a "cheap" group convolution operation based on the Focus structure to extract richer features. Firstly, the original feature map is divided into multiple groups based on the number of channels, and each group is convolved separately. The purpose is to downsample the input feature map, reduce its size, and increase the number of channels. Then, the rearrange operation is used to rearrange the dimensions of the feature map, which is then multiplied element by element with the attention tensor to achieve adaptive weight fusion, thereby better integrating attention information and feature information. The downsampled feature map is multiplied by the attention weight and summed on the last dimension to obtain the final downsampling result.

The LAD method utilizes adaptive weight downsampling operation to enhance the expression ability of features, reducing the number of parameters and floating-point computation of the algorithm. The method dynamically adjusts the weight of each position during the downsampling process through an adaptive weight downsampling adjustment mechanism. This design allows the algorithm to focus more accurately on key feature information, effectively reducing the loss of key information features [16]. The workflow of the LAD method is shown in Fig. 8.

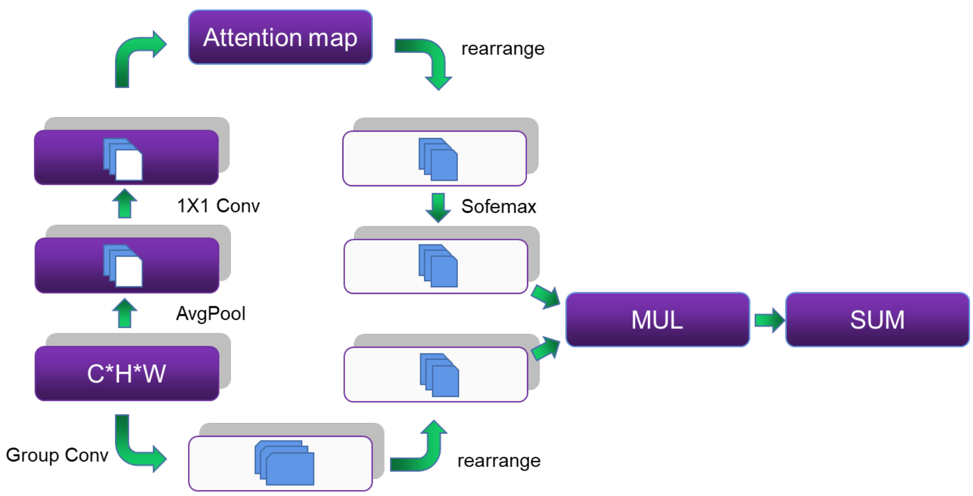

Fig. 8. LAD algorithm structure

The downsampled feature map is multiplied by the attention weight and summed on the last dimension to obtain the final downsampling result. Normalize the weights at each position to probability values as follows:

$$LAD(F) = \sum \left\{ g^{3\times3} \times \sigma\left[g^{1\times1}\left(APL(F) \right) \right] \right\}. \tag{7}$$

g is a convolution operation, 1×1 and 3×3 are convolution kernel sizes, σ is an $soft\max$ operation, APL is an adaptive average pooling operation, and LAD operation can reduce the number of parameters. The rule for reducing the number of parameters is as follows:

$$K \times K \times C_1 \times C_2 \times n^3 \times ATT. \quad \textbf{(8)}$$

In the formula, n represents the number of groups, ATT represents the attention weight map, and after the parameter reduction rule, the parameter amount is the original $1/n$.

Convolutional layers play a decisive role in the process of model feature extraction. Traditional convolutions are mostly static convolutions, while static convolutions only have one static convolution kernel to process the input image. The ability to extract features is insufficient, which can easily lead to information loss or redundancy. Dynamic convolution can perform linear weighting on multiple convolutional kernels, dynamically adjusting their shape and size based on the characteristics of input data to adapt to different input data and improve the performance of convolutional neural networks. For a given input, the weights of each convolutional kernel share the same attention scalar, which limits their ability to capture rich contextual clues and fails to fully utilize the dynamic convolutional properties.

4.2 Integrating Efficient Multi-scale Convolution

In YOLOv8, a large amount of standard convolutions are used to extract image features in the network, which consumes a lot of time. Therefore, lightweight models such as MobileNet and ShuffleNets are used to improve the speed of the detector by using depthwise separable convolution (DSC) operations. Drone target detection requires high detection speed and accuracy of the model, so the Neck network is integrated into the YOLOv8n network to improve precision measurement. On the basis of GSConv, the lightweight structure of GS bottleneck and VoVGSCSP [17] is introduced, which ensures the accuracy of the model and can more efficiently complete aerial target detection tasks. The structure is shown in Fig. 9.

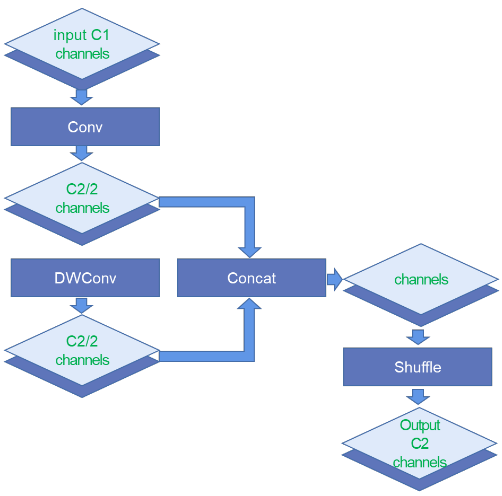

Fig. 9. Schematic diagram of deep convolution operation

GSConv concatenates different hierarchical features obtained from standard convolution and depthwise separable convolution, and then through shuffle operation, completely mixes the information from standard convolution into the output of depthwise separable convolution. The GS bottleneck structure is an enhancement module based on GSConv, which further enhances the network's ability to process features. The VoV-GSCSP structure is a cross stage partial network module designed using a one-time aggregation method to improve feature utilization efficiency and network performance. Its structure is shown in Fig. 10. After the above operations, a network structure incorporating lightweight processing methods has been integrated to achieve algorithm lightweighting, while also improving feature expression ability. Under the same conditions, the computational load of drone processing visual information has been reduced, power consumption has been reduced, and endurance time has been increased.

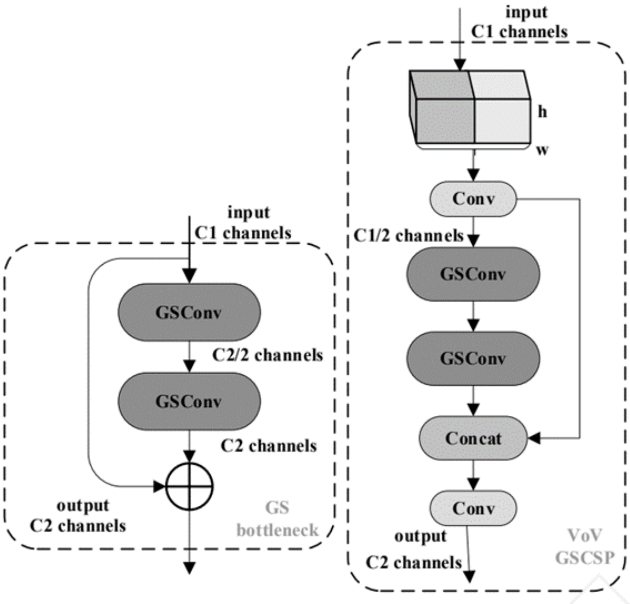

Fig. 10. GS bottleneck and VoVGSCSP structure framework

4.3 The Integration of Attention Mechanisms

The attention mechanism can capture important local information, enabling the model to focus on detecting relevant features of the target, and plays an important role in various computer vision tasks. EMA [18] is an efficient multi-scale attention mechanism that reshapes some channels into batch dimensions, avoiding channel dimensionality reduction, thus preserving the information of each channel and reducing computational costs. EMA not only adjusts the channel weights of parallel sub networks by encoding global information, but also fuses the output features of two parallel sub networks through cross dimensional interaction. The structure of the EMA module is shown in Fig. 11.

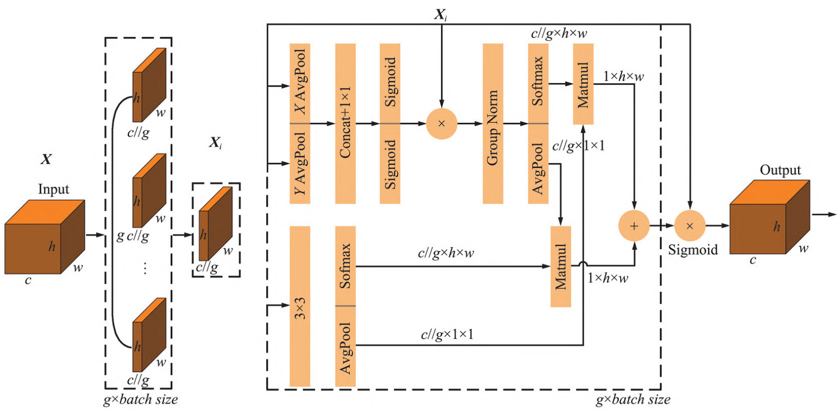

Fig. 11. EMA structure

EMA extracts the weight descriptors of the grouped feature map through two parallel paths on branch 1×1 and one on branch 3×3 . In branch 1×1 , two global average pooling operations are used to encode the channel along two spatial directions, and the two encoding features 1×1 are connected to prevent dimensionality reduction on the branch; Decompose the output after 3×3 convolution into 2 vectors and use 2 *Sigmoid* nonlinear functions to fit the binary distribution on the linear convolution; Finally, cross channel interaction is achieved by aggregating channel attention through multiplication. Use one 3×3 convolution in branch 3×3 to capture multi-scale feature representations. Using global average pooling to encode the global spatial information in the outputs of branches 1×1 and 3×3 , the outputs will be transformed into corresponding dimensional shapes.

Finally, a non-linear matrix dot product operation was added to multiply the results of the parallel processing mentioned above, resulting in a spatial attention map that can collect spatial information at different scales. The final output of EMA is the same size as the input x , making it easy to add directly to the YOLOv8 network.

After the above improvements, the entire module has overcome the problem of weak localization and recognition capabilities for small targets. Due to the interference of normal sized targets and the influence of network deletion of unclear information, small targets are easily affected during the feature extraction process, resulting in a continuous decrease in small target information and unsatisfactory small target detection performance. Therefore, in the network, a more effective bidirectional feature pyramid approach can be used, which can use a simple and efficient weighted feature fusion mechanism to replace the originally simple additive feature fusion and retain more gradient features. The use of weighted feature fusion method can retain more original feature information while reducing model parameters, ultimately obtaining a satisfactory feature recognition model.

5 Experimental Results and Analysis

Based on the content of the third and fourth sections, this section mainly builds an experimental platform to verify the recognition ability of the bee colony algorithm and the processing ability of panoramic images. The scarcity of existing datasets is an obstacle to the experiment in this article. Therefore, this article first builds a simulated flight scene using Unity3D software, and uses the images in the scene as the dataset to ensure the reliability and stability of the data.

5.1 Building an Experimental Environment

The experiment used the Win10 operating system and trained and tested the network based on CPU and GPU. The deep learning framework was Pytorch 1.13.1 and CUDA 11.1. The specific experimental platform and configuration are shown in Table 2. During network training, it is necessary to set some key parameters. The experimental parameters are as follows: batch size is set to 32, input image size is $640*640$, and the training set, validation set, and test set are divided in a ratio of 8:1:1:1. The number of training iterations is 300.

Table 2. Experimental platform and configuration

Configure environment	Version model
Operating system	Linux ubuntu 20.04
CPU	Intel Xeon Platinum 8255C
GPU	NVIDIA GeForce RTX 3080 Ti
RAM	40GB
Programming language	Python3.8.0
Deep learning framework	PyTorch1.13.1
Computational framework	CUDA11.3

5.2 Establishment of Datasets

The dataset must include various typical drone flight operations, including different weather environments, lighting environments, and different situations of obstacles For object detection problems in deep learning, the dataset generally consists of actual images and their annotations of the detected object However, for the specific issue of low altitude obstacle perception by drones, limited by the acquisition cost, the number of obstacle real-life images that can be captured by onboard cameras is relatively limited, and the shooting angles and distances are not rich enough to comprehensively reflect the characteristic information of obstacles Therefore, if only a limited number of real-life images of obstacles are used as the dataset, the trained model may not meet the high accuracy detection requirements In order to expand the dataset of obstacles, this article uses Unity3D software to build realistic flight simulation scenes as a supplement to obstacle reality images The dataset production method used in this article is not only simple, convenient, and cost-effective, but also produces a good generalization effect of the trained model. Some images of the dataset are shown in Fig. 12.

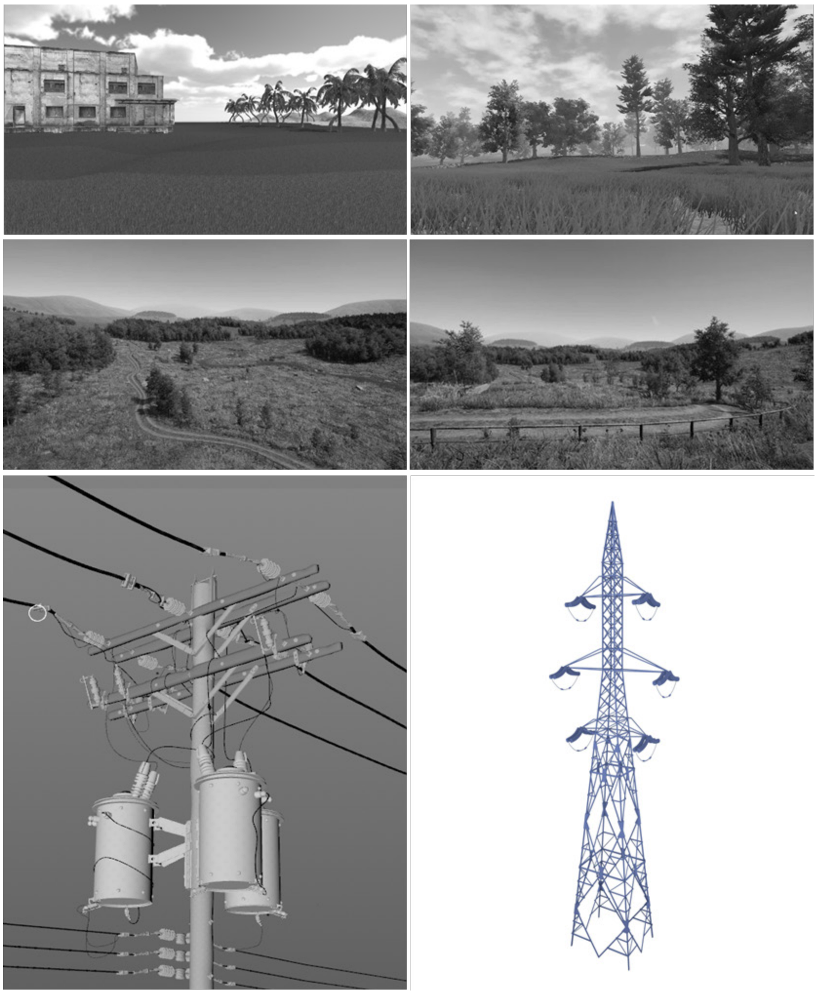

Fig. 12. Display of images in the training set section

5.3 Ablation Experiment and Result Analysis

This article selects evaluation indicators including precision (P), recall (R), average precision (AP), mean average precision (mAP), computational complexity (FLOPs), number of parameters (Params). In order to better verify the effectiveness of the improved model, the detection efficiency of the model before and after lightweight is verified under the same experimental parameters. The comparison object is YOLOv8. The comparison results are shown in Fig. 13.

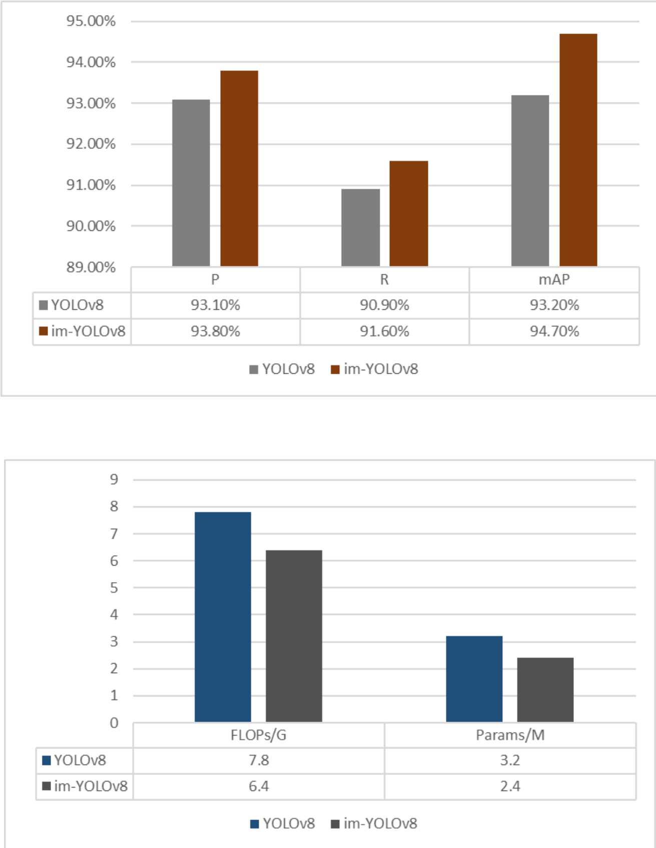

Fig. 13. Comparison of effects before and after improvement

5.4 Panoramic Detection Experiment

In order to verify the panoramic visual obstacle avoidance technology discussed in this article, a drone was first used to take panoramic photos, and the multi camera visual images obtained from multiple angles in the horizontal direction as shown in Fig. 6 were used for testing. The images were then panoramic concatenated and used for testing. The spliced panoramic image of multi vision is shown in Fig. 14, and the obstacle detection image is shown in Fig. 14.

Fig. 14. Identification results

From Fig. 6, it can be seen that the stitching image has eliminated stitching gaps or ghosting, resulting in better image quality for subsequent target detection. From Fig. 14, it can be seen that the spliced images can be recognized. There is no misalignment in house recognition, but identifying the steel truss as a high-voltage line support can accurately identify obstacles even though the name is incorrect.

6 Conclusion

This article focuses on the obstacle avoidance technology of drone panoramic vision. Firstly, the images captured by the panoramic camera are concatenated, and the SURF algorithm is used to analyze and overlap defects in the images. Then, the image is weighted and fused using the arc function weighted fusion image stitching algorithm. The fused image is used as input, and the improved YOLOv8 algorithm is used for target recognition. In order to improve the algorithm performance, the LAD lightweight idea is integrated into the algorithm to reduce the volume of the algorithm. Then, the EMA attention mechanism is added to improve the recognition ability of the algorithm. At the same time, depth separable convolution is added to improve the algorithm's ability to extract target features. Finally, through simulation experiments, the average accuracy of the algorithm was improved by 2.3%. At the same time, it has also demonstrated good recognition and image processing performance in real panoramic image recognition and image processing, meeting experimental requirements. Based on this, this article will continue to deepen the research on panoramic image recognition methods. At the same time, panoramic recognition can also be applied in fields such as unmanned water quality detection ships, and has good market prospects.

Regarding the further research direction of this article and the prospects for obstacle avoidance and navigation technology of unmanned aerial vehicles, a brief summary will also be made at the end of this article. Miniaturization is a development direction for future civilian unmanned aerial vehicles. Low cost, high precision, and high reliability integrated navigation systems for small unmanned aerial vehicles are rigid requirements for obstacle avoidance and path planning. With the continuous upgrading of domestic BeiDou navigation systems and the combination of satellite positioning systems to improve the obstacle avoidance and path planning capabilities of unmanned aerial vehicles, it has a good technical foundation. Therefore, the further research direction is to carry out navigation and positioning technology research based on the combination of medium and low precision inertial navigation systems and satellite deep integration, as well as the combination with visual tracking technology. The further exploration directions are summarized as follows:

1) In practical situations, the motion of drones has a certain degree of uncertainty. The noise in drone image recognition is often not Gaussian white noise, but colored noise with certain spatiotemporal correlation or abnormal characteristics. Studying combined recognition algorithms under colored noise interference conditions is of great significance for improving the accuracy and robustness of drone recognition and obstacle avoidance;

2) The combination of visual recognition and satellite navigation technology has advantages that traditional combination methods cannot match under conditions of high dynamics, weak signals, strong interference, etc. Its complexity and implementation difficulty are also greater. It is necessary to further research on inertial assisted loop stability and fault-tolerant redundancy technology in combination with engineering application requirements, to solve the bottleneck of deep combination technology application in specific scenarios;

3) With the increasingly complex usage environment of future small unmanned aerial vehicles, intelligent unmanned aerial vehicles will undoubtedly occupy a place in future unmanned aerial vehicle technology. Integrated navigation technology under satellite free conditions has become a key development direction of future navigation technology. Integrated navigation technology based on visual tracking technology does not rely on external auxiliary information and can autonomously navigate for a long time. It is a key means to solve the problem of long-term high-precision navigation under satellite free conditions. In the future, research will focus on the calibration and suppression of observation errors in combination navigation methods based on visual tracking technology, further improving the performance of combination navigation methods based on visual tracking technology, and completely solving the problem of long-term high-precision navigation in satellite free environments.

7 Acknowledgement

Research on panoramic vision perception system in low-altitude rotorcraft UAVs (QN2023041), The source of the project is: 2023 Hebei Province Higher Education Scientific Research Youth Fund Project.

References

[1] M. Portugal, A.-C. Marta, Optimal Multi-Sensor Obstacle Detection System for Small Fixed-Wing UAVs, Modelling 5(1)(2024) 16-36.
[2] D. Fr¨anken, A. Hupper, Unified tracking and fusion for airborne collision avoidance using log-polar coordinates. in: Proc. The15th IEEE International Conference on Information Fusion, 2012.
[3] F. Mei, X.-Y. Gao, S.-C. Deng, W.-M. Li, X. Zou, UAV rapid identification, positioning and tracking system based on binocular ranging and YOLOv5s, Modern Electronics Technique 46(10)(2023) 181-186.
[4] K.-Y. Tan, J.-L. Zhu, J. Deng, L.-Q. Wang, J.-Y. Yang, SLAM Quadrotor UAV Based on Binocular Vision, Mechanical & Electrical Engineering Technology 51(9)(2022) 83-87+141.
[5] Y. Cheng, T.-L. Zheng, Binocular Visual Obstacle Avoidance of UAV Based on Deep Learning, Electronics Optics & Control 28(10)(2021) 31-35.
[6] X.-J. Ma, R.-Z. Zhang, Z.-L. Ma, D.-K. Yan, Research on UAV Obstacle Avoidance Technology Based on Binocular Panoramic Vision, Mechanical & Electrical Engineering Technology 48(3)(2019) 16-19.
[7] Y.-Y. Liu, M. He, Y.-Y. Wang, Y. Sun, X.-B. Gao, Fast stitching for the farmland aerial panoramic images based on optimized SIFT algorithm, Transactions of the Chinese Society of Agricultural Engineering 39(1)(2023) 117-125.
[8] B.-Y. Wang, H.-W. Li, S. Zhao, Road side multi-object recognition by integrating point cloud and panoramic image, Bulletin of Surveying and Mapping (10)(2023) 40-46.
[9] W.-B. Chen, X.-J. Jia, X.-B. Zhu, E.-F. Ran, H. Xie, Target Detection for UAV Image Based on DSM-YOLOv5, Computer Engineering and Applications 59(18)(2023) 226-233.
[10] Y. Chen, Y. Ding, L. Wang, Research on Multiview Panoramic Imaging System Based on Linear Array, Electric Engineering (24)(2019) 153-156+172.
[11] Z. Kou, J.-F. Wu, H.-L. Wang, W.-Y. Ruan, Obstacle visual sensing based on deep learning for low-altitude small unmanned aerial vehicles, Scientia Sinica (Informationis) 50(5)(2020) 692-703.
[12] X.-G. Cao, J.-L. Yang, W. Wan, X.-Y. Wang, Feature Extraction and Registration of Snow Covered Image Based on SURF Algorithm, Geomatics & Spatial Information Technology 46(5)(2023) 1-4+8+12.
[13] X.-F. Zhang, X. Ma, C.-Y. Mu, Research on image matching based on quadtree fusion of SIFT and K-D tree, Electronic Measurement Technology 44(22)(2021) 121-127.
[14] Y.-L. Huo, Q.-S. Zhang, J.-S. Zhang, H.-D. Fan, An image dehazing algorithm using weighted fusion with the Sigmoid function, Computer Engineering & Science 45(9)(2023) 1621-1628.
[15] H.-X. Cheng, Q.-Y. Qiao, X.-L. Luo, S.-J. Yu, Object Detection Algorithm for UAV Aerial Image Based on Improved YOLOv8, Radio Engineering 54(4)(2024) 871-881.
[16] X.-G. Xue, J.-W. Gao, B.-H. Zhang, L.-Q. Huang, Research on topic extraction algorithm based on Map Reduce parallel LAD model, Journal of Fuzhou University (Natural Science Edition) 44(5)(2016) 644-648.
[17] X.-Y. Niu, P.-J. Mao, Y.-T. Duan, X.-H. Lou, Research on Lightweight ImprovedAlgorithm for Indoor Target Detection Based on YOLOv5s, Computer Engineering and Applications 60(3)(2024) 109-118.
[18] H.-Y. Zhao, H. Zhu, X.-G. Hou, Traditional Custume Image Semantic Segmentation Based on Improved EMA Unit Based on Improved EMA Unit, Journal of Beijing University of Posts and Telecommunications 45(1)(2022) 69-74.