CNN-Based Salient Target Detection Method of UAV Video Reconnaissance Image

Li Na
Hainan Vocational College of Political Science and Law, Haikou City, Hainan Province, 570100, China

Abstract*—In order to address the challenges of image complexity, capturing subtle information, fluctuating lighting, and dynamic background interference in drone video reconnaissance, this paper proposes a salient object detection method based on convolutional neural network (CNN). This method first preprocesses the drone video reconnaissance images to remove haze and improve image quality. Subsequently, the Faster R-CNN framework was utilized for detection, where in the Region Proposal Network (RPN) stage, the K-means clustering algorithm was used to generate optimized preset anchor boxes for specific datasets to enhance the accuracy of target candidate regions. The Fast R-CNN classification loss function is used to distinguish salient target regions in reconnaissance images, while the regression loss function precisely adjusts the target bounding boxes to ensure accurate detection of salient targets. In response to the potential failure of Faster R-CNN in extreme situations, this paper innovatively introduces a saliency screening strategy based on similarity analysis to finely screen superpixels, preliminarily locate target positions, and further optimize saliency object detection results. In addition, the use of saturation component enhancement and brightness component dual frequency coefficient enhancement techniques in the HSI color space significantly improves the visual effect of salient target images, enhancing image clarity while preserving the natural and soft colors, effectively improving the visual quality of detection results. The experimental results show that this method exhibits significant advantages of high accuracy and low false detection rate in salient object detection of unmanned aerial vehicle (UAV) video reconnaissance images. Especially in complex scenes, it can still stably and accurately identify targets, significantly improving detection performance.

Keywords*—*Regional convolutional neural network; K-means clustering; UAV reconnaissance image; salient target detection; task loss function

I. Introduction

The salient target detection of UAV video reconnaissance image is a method that uses computer vision and depth learning technology to quickly identify and locate key targets from the image taken by UAV. It automatically detects salient targets in the image, such as people, vehicles, buildings, etc., by analyzing the texture, color, shape, and other characteristics of the image [1]. The detection of the salient target of a UAV video reconnaissance image has extensive application value in many fields, such as safety monitoring, public safety, environmental monitoring, disaster rescue, etc. [2]. Through real-time monitoring and early warning, this technology can help people find abnormal situations in time, improve monitoring accuracy and response speed, and provide strong support for preventing and responding to various emergencies [3]. Therefore, the detection of salient targets of UAV video reconnaissance images is of great significance in ensuring public safety and improving emergency response capability.

Jirayupat C, et al. proposed a method of detecting salient targets in UAV video reconnaissance images based on chromatography-mass spectrometry [4]. This method pre-processes the image using the chromatography-mass spectrometry method, extracts the salient target features in the image, and then detects the target using the trained machine learning model. By combining image processing and machine learning, the salient target in the UAV video can be quickly and accurately recognized. This method involves the integration of many technologies, including image processing, chromatography, mass spectrometry, and machine learning. This requires researchers to have extensive professional knowledge and skills, and it is difficult to achieve and optimize. Cherri A K et al. proposed a method of detecting salient targets of UAV video reconnaissance images based on a joint transform correlator [5]. This method reduces the size of the UAV video reconnaissance image to speed up the processing time and increase the storage capacity of the recognition system. By using fringe adjustment joint transform correlation technology, multiple targets based on compression are successfully detected. The use of shift phase encoding and reference phase encoding technology can eliminate the false detection and missed detection caused by multiple expected and unwanted targets, thus realizing the detection of salient targets of UAV video reconnaissance images more accurately. This method reduces the size of the UAV video reconnaissance image, which will lead to the reduction of image resolution and will affect the accuracy of salient target detection and detail recognition. Iván García-Aguilar and others proposed the method of detecting salient targets of UAV video reconnaissance images using CNN and super-resolution [6]. This method improves the resolution of UAV video images through super-resolution technology and then uses a convolutional neural network (CNN) for feature extraction and target detection. Train CNN models to identify and locate salient targets. Combining the advantages of deep learning and super-resolution technology, it can accurately detect salient targets in UAV video reconnaissance images of low resolution. In this method, the super-resolution technology is sensitive to the noise in the image, which will affect the accuracy and reliability of target detection. Ezequiel López-Rubio a b and others proposed a method of detecting salient targets of UAV video reconnaissance images based on deep learning and PTZ camera controller [7]. This method uses deep learning to detect the salient target of a UAV video reconnaissance image. This method includes three modules: target detection, salient target detection, and PTZ camera controller. The deep learning network is used to detect the target in the scene. The salient target detection module automatically detects the salient target using the Dirichlet distributed hybrid model, while the PTZ camera controller allows it to follow and focus on the salient target to achieve the salient target detection of UAV video reconnaissance image. This method uses deep learning to detect the salient target, which faces a challenge in the real-time environment. On low-performance hardware, the processing speed cannot meet the real-time requirements, resulting in the limitation of this method in practical applications.

Faster R-CNN can effectively deal with complex factors in drone video reconnaissance images, such as subtle information, lighting changes, and dynamic backgrounds, ensuring high-precision salient object detection under various conditions. Therefore, this paper proposes a CNN-based method of detecting salient targets of UAV video reconnaissance images. By performing haze removal operations in the preprocessing stage, image quality is improved, and the Faster R-CNN framework combined with K-means clustering is used to optimize anchor boxes, achieving accurate detection of targets in complex scenes. Specifically, in response to the limitations of Faster R-CNN in extreme situations, this paper introduces a superpixel saliency filtering strategy and combines it with image enhancement techniques in the HSI color space. This not only significantly improves the accuracy of object detection and image clarity, but also ensures the authenticity and softness of colors, providing more reliable and efficient technical support for drone video reconnaissance. The specific research approach is as follows:

1) After removing haze from drone video surveillance images, a salient object detection framework for drone video surveillance images is constructed.

2) Preset anchor boxes in RPN and determine the final target bounding box through a multi task loss function.

3) Significant object detection calculation is used for undetectable targets, including calculating target similarity and target connectivity maps to obtain target probabilities.

4) The enhancement of saturation and brightness components based on HSI color space improves the accuracy of salient object detection.

II. Salient Target Detection of UAV Video Reconnaissance Image

A. Dehazing of UAV Video Reconnaissance Image

UAV imaging will be interfered with a variety of factors. Due to the presence of haze in the air or chemical particles and other particles, in the propagation process, the light will be scattered or refracted to varying degrees, resulting in some detailed information not being received by the sensor [8]. The haze image degradation model formula is expressed by Eq. (1):

$$A(x) = Z(x)t(x) + \rho[1 - t(x)] \tag{1}$$

In the equation, $A(x)$is the haze image degradation model, $Z(x)$ is a realistic image, $t(x)$ is medium transmission, ρ is atmospheric scattered light.

Light sources are scattered or reflected by the haze present in the air as they travel through the atmosphere. Then the light reflected through the object is also scattered or reflected by the haze [9]. After that, only the energy of$Z(x)t(x)$can reach the sensor. At the same time, the atmospheric scattered lightρgenerated due to the scattering of various particles in the air will also be absorbed by the sensor.

The medium transmission is expressed by Eq. (2):

$$t(x) = h^{-\varepsilon d(x)} \tag{2}$$

The medium transmission is related to the scattering coefficient $\varepsilon$of the atmosphere and also related to the distance $d(x)$from the sensor to the object. When$d(x)$goes to infinity $t(x)$ is close to zero. Then the atmospheric scattered light can be expressed by Eq. (3):

$$\rho = A(x), d(x) \to inf \tag{3}$$

In practical imaging, instead of relying on equations, obtaining the atmospheric scattered light$\rho$is rather a more stable calculation based on Eq. (4). $d(x)$cannot be infinite, but gives a very low transmittance t_0.

$$\rho = \max_{y \in \{x | t(x) \leq t_0\}} A(y) \tag{4}$$

If the atmospheric light in the corresponding area is given, the medium transmission can be calculated based on physical principles, and a clear image can be obtained after the haze is removed. The following procedures are all performed on the dehazed UAV video reconnaissance image.

B. Target Detection of UAV Video Reconnaissance Image Based on Faster R-CNN

After getting a clear image after removing the haze, the corresponding targets can be detected from it. For UAV video reconnaissance, multiple targets need to be monitored at the same time. To detect multiple targets at the same time and ensure detection accuracy, Faster R-CNN is used to build a detection framework, and training is used to adapt to different scenes and target types, to achieve the target detection of UAV video reconnaissance images.

1) Target detection framework: Faster R-CNN (Faster Region-based Convolutional Neural Network) is a new generation of target detection model that optimizes CNN [10]. Faster R-CNN can be seen as a combination of an RPN (Region Proposal Network) and a Fast R-CNN (Region-based Convolutional Network), which share the basic convolutional network [11]. RPN is responsible for generating high-quality proposed target areas in UAV video reconnaissance images, and Fast R-CNN will complete the salient target classification and position regression of the proposed target areas [12]. The improved target detection framework in this paper is shown in Fig. 1. First, the anchor frame is reset by the clustering method to make the reference anchor frame more suitable for the target characteristics of the dataset. Second, a new full connection layer (behind the ROI layer) is added to the Fast R-CNN model to effectively improve the detection performance of the algorithm.

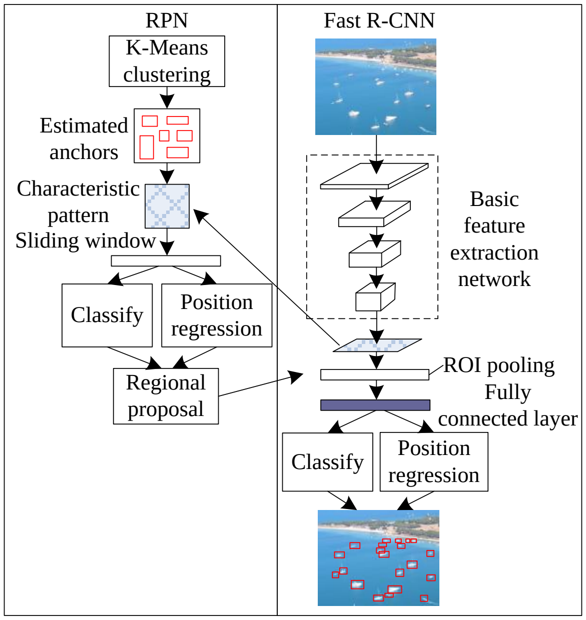

Fig. 1. Improved faster R-CNN target detection framework.

As displayed in Fig. 1, RPN uses a sliding window and anchor mechanism to generate a proposed target area of multi-scale. After the image passes through the basic feature extraction network, a convolution kernel of 3×3 slides on the feature map to get a feature vector of 512 dimensions (corresponding to the VGG model) each time, and then the vector is sent to the full connection layer: (1) Object/non-object binary layer, to predict whether there is a target in the window; (2) Position regression layer, to calculate the position correction of the anchor frame relative to the target boundary frame, and to obtain the position coordinates of the candidate frame. Fast R-CNN and RPN share the feature extraction network. First, ROI pooling is used to obtain the feature representation of each candidate frame, and then ROI features are sent to the full connection layer, and the features are sent to the two parallel task layers: (1) Softmax classification layer, to calculate the probability of candidate frames on the C+1(Class C target + background) class; (2) Position regression layer, to calculate the relative offset between the candidate frame and the target boundary frame, and to further correct the position of the predicted target.

2) Anchor frame setting: In the RPN model, anchor frames with three different aspect ratios are set according to experience. However, for different datasets, the scale of the anchor frame is inconsistent. Choosing an appropriate anchor frame can effectively improve the learning speed of network model training, and can also improve the target detection accuracy of UAV video reconnaissance images [13]. In this paper, K-Means clustering is used to select anchor frames. The algorithm flow is shown in Fig. 2.

- Collect the truth frames of all target samples in the UAV video reconnaissance image set $a = (xmaxmaxmin_{min})$, and set the number of clusters, i.e., the number of anchor frames k, and then randomly selected$k$samples as initial clustering centers;
- Calculate the distances between the remaining target samples and k centers, based on the Euclidean distance. The center with the smallest distance is used as the target sample class;
- Calculate the new clustering center according to the target classification results, and set the clustering center be o. The criterion function is calculated as shown in Eq. (5):

$$I(a,o) = \rho \sum_{i=1}^{k} \left\| a^i - o_c^i \right\|^2 \tag{5}$$

- Repeat steps (2) and (3) until the error of the criterion function is within the allowable range, and end the loop to obtain the final target classification results.

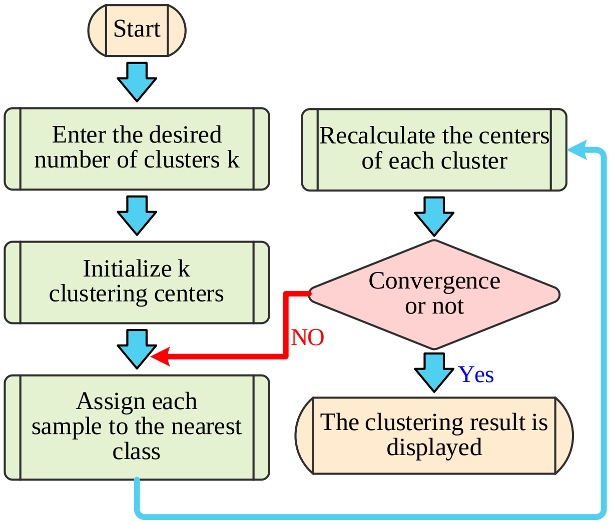
Start
Enter the desired number of clusters k
Initialize k clustering centers
Assign each sample to the nearest class
Recalculate the centers of each cluster
Convergence or not
NO
Yes
The clustering result is displayed

Fig. 2. Clustering flowchart of anchor frame.

3) Multitask loss function: The multitask loss function in the Fast R-CNN model is mainly composed of the classification loss function and regression loss function [14]. The classification loss function is used for the classification of salient targets in the proposed target areas of UAV video reconnaissance images, and the regression loss function is used for the regression of salient target boundary frames in UAV video reconnaissance images [15]. The loss function definition formula for a UAV video reconnaissance image is expressed by Eq. (6):

$$H[(r_i),(s_i)] = \frac{I(a,o)}{N_{cls}} \sum_i H_{cls}(r_i, r_i^*) \tag{6}$$

$$+\alpha \frac{1}{N_{reg}} \sum_i r_i^* H_{reg}(s_i, s_i^*)$$

In the equation, r_i is the prediction probability of the i -th anchor, r_i^* is the predicted probability of the actual boundary frame $\left(GT, Grond \quad Truth\right)$corresponding to the i -th anchor.

If the recall rate between the i -th anchor boundary frame andGT,$IoU > 0.7$ (at this point $r_i^* r_i^* = 1$), then the anchor is considered as a salient target. If $0.3 < IoU < 0.7$, then the anchor will not participate in the training. If $IoU < 0.3$ (at this point $r_i^* = 0$), then the anchor is considered as the background.

s_i is a vector, indicated as$s_i\{s_x, s_y, s_w, s_v\}$, corresponding to the four parametric coordinates of the boundary frame of the predicted salient target. s_x, $s_y$are corresponding to the central coordinates of the boundary frame of the salient target. s_w, s_v are corresponding to the width and height of the boundary frame of the salient target.$s_i^*$is the coordinate vector of the salient target GTcorresponding to the anchor.

The classified loss is the logarithmic loss of salient target class and non-salient target class, and the calculation formula is expressed by Eq. (7):

$$H_{cls}(r_i, r_i^*) = -lg[r_i^* r_i + (1 - r_i^*)(1 - r_i)] \quad (7)$$

The regression loss is calculated as shown in Eq. (8):

$$H_{reg}(s_i, s_i^*) = B(s_i * -s_i^*) \quad (8)$$

In the equations, B is the defined robust loss function ($smooth\ H_1$) , and the calculation formula is expressed by Eq. (9):

$$smooth\ H_1(x) = \begin{cases} 0.5x^2 & if\ |x| < 1 \\ |x| - 0.5 & otherwise \end{cases} \quad (9)$$

Using a four-dimensional vector$(x, y, w, v)$for the target display window, respectively represents the center point coordinates, width, and height of the target window. Through the regression learning of the salient target boundary frame in the UAV video reconnaissance image, a relational mapping is found, to obtain the regression target window G'which is close to the real target window G by input of the original anchor C through the mapping method, i.e., given $C = (C_x, C_y, C_w, C_v)$, and look for a relational mapping$f$, and make $f(C_x, C_y, C_w, C_v) = (G'_x, G'_y, G'_w, G'_v)$, of which$(G'_x, G'_y, G'_w, G'_v) \approx (G_x, G_y, G_w, G_v)$.

The following two methods of translation transformation and scaling transformation are used to implement the transition from anchor to approximate GT.

The translation transformation is calculated as shown in Eq. (10):

$$\begin{cases} G'_x = C_w d_x(C) + C_x \\ G'_y = C_v d_y(C) + C_y \end{cases} \quad (10)$$

The scaling transformation is calculated as shown in Eq. (11):

$$\begin{cases} G'_w = C_w\, exp[d_w(C)] \\ G'_v = C_v\, exp[d_v(C)] \end{cases} \quad (11)$$

The amount of translation(s_x, s_y), $(s_x^*, s_y^*)$and the scale factor(s_w, s_v), $(s_w^*, s_v^*)$are expressed by Eq. (12), (13), (14), and (15):

$$s_x = (x - x_c)/w_c; s_y = (y - y_c)/v_c \quad (12)$$

$$s_x^* = (x^* - x_c)/w_c; s_y^* = (y^* - y_c)/v_c \quad (13)$$

$$s_w = lg(w/w_c)\, ; s_v = lg(v/v_c) \quad (14)$$

$$s_w^* = lg(w^*/w_c)\, ; s_v^* = lg(v^*/v_c) \quad (15)$$

In the equations, x , y indicate the coordinates x and the coordinates y of the center of the predicted target boundary frame. w, v indicate the width and height of the target boundary frame. The coordinate parameters of the boundary frame of the anchor are respectively expressed as x_c , y_c , w_c , v_c . The coordinate parameters of the boundary frame of GT are respectively expressed asx^*, $* y^*$, w^*, v^*.

C. Detection of Salient Target of UAV Video Reconnaissance Image

The above calculation formula can be understood as, by the regression learning of the target boundary frame, that is, the regression from the anchor boundary frame to the nearby GT boundary frame, a boundary frame of the regression target window G' which is much closer to the actual target window G is obtained. The target is detected according to the boundary frame, but due to some certain occlusion or mistaken recognition of the target as an unrecognized object in the UAV video detection, it is further optimized to achieve salient target detection.

1) Calculation of target similarity: By using Faster R-CNN, through the training of salient target detection in UAV video reconnaissance images, the identification of salient targets within the image can be efficiently identified [16], followed by the extraction of potential targets based on their distinctive features [17]. Following that, begin to create the similarity graph. Faster R-CNN exhibits a notable detection rate; however, it may struggle to detect salient targets in exceptional circumstances. If Faster R-CNN cannot identify the salient target, it will process the entire image as a target.

The likelihood of this window containing a target is denoted by the target similarity score. The pixel-level similarity score [18] is derived from the potential targets to determine the likelihood of a pixel being a component of the target. The pixel-level similarity score is determined as shown in Eq. (16) as follows:

$$PixObj(q) = \sum_{i=1}^{N} e_i\, L_i(x, y) \quad (16)$$

The equation, $e_i$indicates that if the pixel $q$is contained in the target window i detected by Faster R-CNN, $L_i$is a Gaussian filter window of equal dimension to the window, x, $y$is the relative coordinate of pixels$q$ in one of the detection windows, $N$is the number of possible target windows detected by Faster R-CNN.

The sum of the similarity scores of all pixels in the superpixel region is the similarity score of the current superpixel region, defined by Eq. (17).

$$Objectness(Tq_i) = \sum_{i \in M} PixObj(q_j) \quad (17)$$

In the equation, q_j is one pixel in the i -th superpixel regionTq_i. The similarity graph is then further optimized using a threshold-based method, where the threshold is set as 1.5 times the quantity of pixels in the similarity graph divided by the overall number of pixels in the graph.

2) *Calculation of target connectivity graph*: The threshold similarity graph is only a part of the target superpixels that are roughly acquired. This paper adopts the "target connectivity" method, which assigns a value to the predicted target according to the salience value of the superpixel connectivity [19]. Build a graph using superpixels as nodes. The adjacent superpixel nodes have edges, and the weight of these edges is specified as the Euclidean space of the mean Lab color of the two nodes. The target connectivity of the i -th superpixel Tq_i is defined by Eq. (18):

$$F(Tq_i) = \frac{\sum_{m=1}^{N_1} d(Tq_i,Tq_m)\cdot\beta(Tq_m)}{\sum_{m=1}^{N_1} d(Tq_i,Tq_m)\cdot[1-\beta(Tq_m)]} \quad (18)$$

In the equation, $d(Tq_i,Tq_m)$indicates the shortest distance between superpixels Tq_i and Tq_m. If the superpixel Tq_m is predicted as a target in the similarity graph, then assign the value of $\beta(\cdot)$ as 1, and N_1 is the total number of superpixels.

The more superpixel similarities that are predicted as targets, the lower the numerator value and the higher the denominator values, which makes the lower value of F. Set the reciprocal of Fas the target probabilityf_i, which means that the algorithm has identified the possible salient target in the UAV video reconnaissance image. To improve the detection accuracy of the salient target, following image enhancement algorithm is used to improve the detection accuracy of the salient target.

D. *Enhancement of Salient Target of UAV Video Reconnaissance Image Based on HSI Color Space*

First, convert the low illuminance image from RGB space to HSI space [20], and then different enhancement algorithms are used respectively according to component (S) and component(I). The process of this algorithm is: (1) Color space conversion, from RGB space to HSI space; (2) Enhance components$(S)$and(I)respectively by "Piecewise exponential transformation" and "V-transformation + Retinex enhancement + improved fuzzy enhancement"; (3) Return to RGB color space to get the enhanced image. The flow chart is shown in Fig. 3.

The enhancement algorithms for each stage are described in detail below.

1) *Enhancement of saturation component*: Generally, the enhancement of the saturation component is a simple linear transformation [21]. This paper proposes a piecewise exponential enhancement algorithm, characterized by its nonlinear nature. The saturation levels of distinct areas can be processed individually to enhance the overall visual impact. The salient target image of the UAV video reconnaissance image is divided into three regions based on its saturation level: high, medium, and low. When $x > 0$, $u^x - 1 > x$, and $\lim_{x\to 0}\frac{u^x-1}{x} = 1$, that is, when$x$is extremely small, $u^x - 1$ and$x$are equivalent. Therefore, the low saturation region is stretched by the exponential transformation to enlarge the saturation; for the medium saturation region, only exponential transformation is done to make appropriate adjustments; for the high saturation region, the saturation is reduced appropriately through the reduction of the exponential transformation. The piecewise exponential enhancement algorithm proposed in this paper is expressed by Eq. (19):

$$P'(m,n)=\begin{cases}\eta\left[u^{P(m,n)}-1\right], & P(m,n)\le 0.2\\ u^{P(m,n)}-1, & 0.2<P(m,n)\le 0.7\\ \mu\left[u^{P(m,n)}-1\right], & else\end{cases} \quad (19)$$

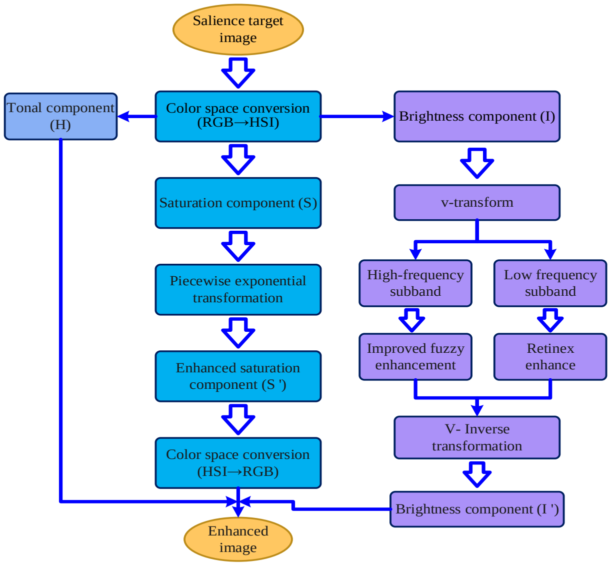
Salience target image
Tonal component (H)
Color space conversion (RGB→HSI)
Brightness component (I)
Saturation component (S)
v-transform
High-frequency subband
Low frequency subband
Piecewise exponential transformation
Improved fuzzy enhancement
Retinex enhance
Enhanced saturation component (S ')
V- Inverse transformation
Color space conversion (HSI→RGB)
Brightness component (I ')
Enhanced image

Fig. 3. Flow chart of image enhancement algorithm.

In the equation, $P$and$P'$are the saturation before and after enhancement, respectively; the parameters η, $\mu$are used to adjust the scaling of the transformation. When$\eta$is 1.2~1.5, and $\mu$is 0.7~0.9, the enhancement effect is the best. In this paper, the value of η is 1.4, and the value of μ is 0.8.

2) *Enhancement of brightness component*: V-transform the brightness component to obtain the V-spectrum matrix. The data of the 1/4 of the upper left corner of the V-spectrum matrix is the low-frequency sub-band $I_L$of the brightness component, with the rest of the data being the high-frequency sub-bandI_H. Different methods of enhancement are used for low and high frequencies.

a) *Low-frequency coefficient enhancement (Retinex):* Because the low-frequency sub-band $I_L$concentrates the overall information of the brightness of the salient target image of the UAV video reconnaissance image, it depicts the general outline of the image, and the Retinex enhancement algorithm can be used.

Assuming that the brightness low-frequency coefficient is$P(x,y)$, the enhanced low-frequency coefficient is calculated by Eq. (20):

$$\begin{aligned} z(x,y) &= ln\,J\,(x,y)\\ &= ln\,P\,(x,y) - ln[D(x,y) * P(x,y)] \end{aligned} \quad (20)$$

$z(x,y)$ stands for the output image after enhancement and $*$ stands for the convolution symbol; $D(x,y)$ is the center-surround function, a Gaussian filter function is usually chosen as shown in Eq. (21):

$$D(x,y) = \sigma \cdot u^{\frac{-(x^2+y^2)}{c^2}} \tag{21}$$

In the equation, c is the Gaussian surround scale; σ is a constant which makes the integral of $D(x,y)$ as 1. The reflected image$J(x,y)$ represents the intrinsic property of an image and carries detailed information about the image.

The key to Retinex theory is to reasonably assume the composition of the salient target image of the UAV video reconnaissance image [22]. If an image is regarded as an image with noise, then the component of the incident light can be regarded as a multiplicative, relatively uniform, and slowly transformed noise [23]. The Retinex algorithm can fairly estimate the noise present at every position within the image and remove it to acquire a noticeable Impact. The obtained image reduces the impact of incident light [24], and retains the reflection attribute of the object essence, that is, the essence of the image.

b) High-frequency coefficient enhancement (Improved fuzzy optimization): In the sub-bandI_H characterized by high frequency, the wavelet coefficient of noise exhibits a diminutive magnitude, which is further reduced compared to the coefficient of the signal. To enhance the details of the salient target image and suppress noise [25], this paper proposes an improved fuzzy enhancement algorithm, which aims to enhance the clarity of details while simultaneously suppressing noise [26]. The specific algorithm process is outlined below:

Step 1: Construct the affiliation function as shown in Eq. (22):

$$E_{mn} = \frac{x_{mn}-x_{min}}{xmin_{max}} \tag{22}$$

Transforming the high-frequency coefficient into fuzzy sets is aiming to normalize the coefficient in the high-frequency subband to the interval$[0,1]$. In the equation, x_{max}andx_{min}stand respectively for the highest and lowest values of the coefficient in the high-frequency subband; x_{mn} is the coefficient.

Step 1: Design the fuzzy affiliation transformation as expressed by Eq. (23):

$$E'_{mn} = \frac{1}{2} + \left(E_{mn} - \frac{1}{2}\right)^{1/3} \tag{23}$$

This function is a nonlinear and monotonically increasing function and $\left(\frac{1}{2},\frac{1}{2}\right)$ is its inflection point. It makes the numbers less than $\frac{1}{2}$ shrinking, and makes the numbers greater than$\frac{1}{2}$amplified. When acting on the high-frequency coefficient, it can achieve the purpose of enhancing the details of the salient target image of the UAV video reconnaissance image while suppressing noise.

Step 1: Convert the fuzzy set to the high-frequency subband as shown in Eq. (24):

$$E''_{mn} = E'_{mn} \cdot (xmin_{max} + x_{min}) \tag{24}$$

The enhanced high-frequency coefficients are obtained.

So far, the enhanced low-frequency and high-frequency coefficients are obtained, and then the enhanced brightness component I' is obtained through V-inverse transformation. Finally, the obtained enhanced saturation$S'$and the enhanced brightness I', and the hue component of the salient target image$(H)$are synthesized and converted to RGB space to output the enhanced salient target image.

III. Experimental Analysis

To verify the effect of the method in this paper on the detection of salient targets of UAV video reconnaissance images, the GPU configuration of experimental hardware equipment is selected as displayed in Table I.

The UAV used for reconnaissance shooting is shown in Fig. 4. The configuration parameters are shown in Table II.

TABLE I. GPU Configuration

Name	Parameter
Brand	Shadow Chi
Model	GTXTITAN-6GD5
Craftsmanship	28mm+
Stream processor	2688pcs
Core frequency	954 MHze
Video memory capacity	6G GDDR5+
Video memory bit width	384Bite
Video memory frequency	6008MHz
Graphics card power consumption	300W+
Heat-dissipating method	Cooling fan
Size	280mm×127mm×42mm

TABLE II. UAV Configuration Table

Name	Parameter
Model	RQ-8 drone
Maximum safe takeoff weight	7kg
No-load	6.5kg
Load	2kg
Size	The length is 970mm
	Width 920mm, 500mm after folding
	Height 240mm, folded back 180mm
Wheelbase size	1100mm
Propeller size	26inch
Duration of flight	>50min
Maximum speed	36km/h
Wind loading rating	Strong breeze
Operating radius	>10km
Dynamic type	Whoring polymer cell
Working altitude	4500m
Operating temperature	-20 to 60 degrees Celsius

Fig. 4. UAV used for reconnaissance shooting.

The experimental test dataset consists of images captured by a drone as shown in Fig. 4. This dataset consists of 500 frames of video reconnaissance images captured by drones in different scenes, lighting conditions, and dynamic backgrounds. These images contain complex and subtle information, such as hidden targets, varied backgrounds, and challenging factors such as lighting changes. The dataset is divided into two parts: (1) Training set: containing 300 frames of images, used to train a CNN based saliency object detection model. These images cover various possible scenarios and conditions in the dataset to ensure that the model can learn enough features to cope with complex detection tasks. (2) Test set: Contains the remaining 200 frames of images to evaluate the performance of the trained model on unknown data. These images maintain similar diversity to the training set, but are completely independent to ensure the objectivity and accuracy of the test results.

After completing the above preparations, proceed with the experiment according to the following steps:

Step 1: Dataset preparation and annotation

Collect and organize drone video surveillance images, accurately label salient targets in the images through manual means, and provide accurate data foundation for model training.

Step 2: Remove haze from the image

To remove haze from the original image, improve image quality, reduce the impact of haze on subsequent object detection, and enhance detection accuracy.

Step 3: Faster R-CNN model training

Using annotated training set data, train the Faster R-CNN model to learn the features of salient targets and possess the ability to classify and perform bounding box regression.

Step 4: Test Set Evaluation

Evaluate the trained Faster R-CNN model using an independent test set to validate its detection performance on unknown data, including subjective and objective testing metrics such as accuracy, recall, and F1 score.

Step 5: Extreme situation handling and image enhancement

In response to extreme situations where the Faster R-CNN model cannot accurately detect, methods such as superpixel saliency screening are used for supplementary detection, and HSI color space enhancement is applied to the detected saliency target images to improve image clarity and visual effects.

Step 6: Result analysis and optimization

Conduct a comprehensive analysis of the experimental results, evaluate the advantages and disadvantages of the model, and further optimize the model based on the analysis results to improve detection performance and robustness.

In the test set, 1 frame image is selected to perform the target detection experiment using the trained network, and the detection results are plotted as shown in Fig. 5. And compare these detection results with the detection results of the chromatography-mass spectrometry method in study [4] and the joint transform correlator method in study [5], which are shown in Fig. 6 and Fig. 7. The red boxes are the correct targets, the yellow circles are the false detection, and the blue boxes are the missed detection.

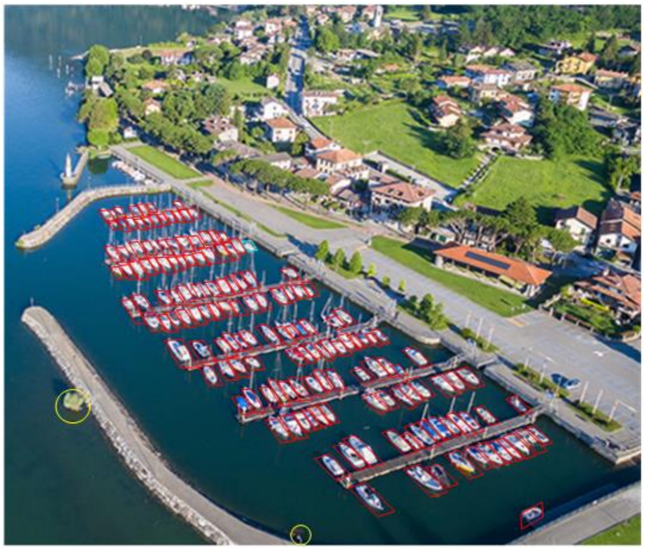

Fig. 5. Detection results of the proposed method.

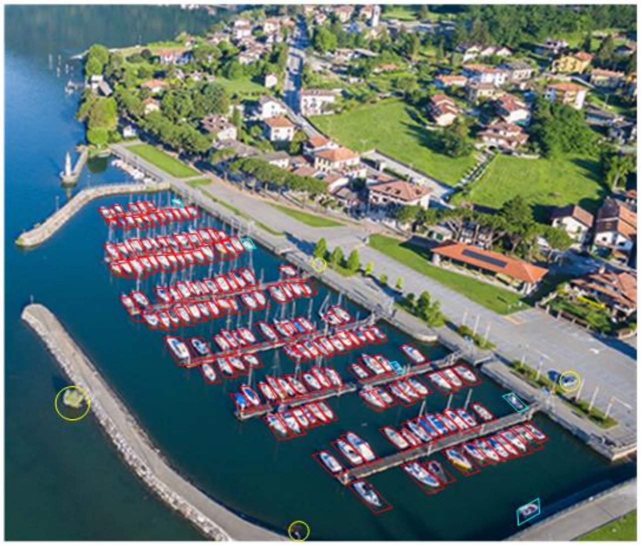

Fig. 6. Test results by chromatography-mass spectrometry.

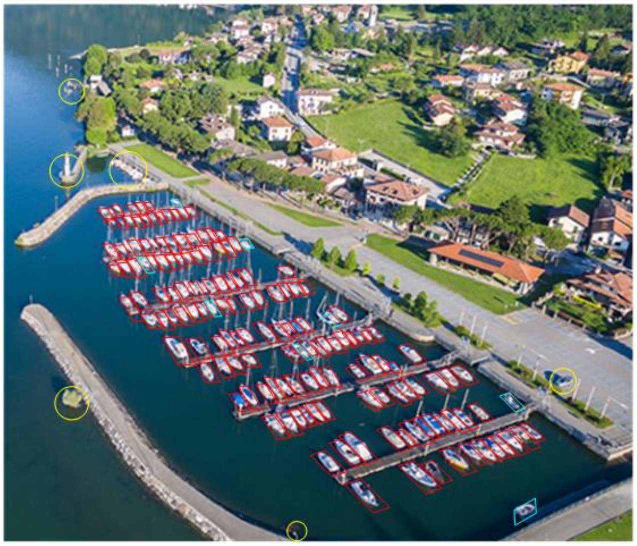

Fig. 7. Detection results of the joint transform correlator method.

From Fig. 5, it can be seen clearly that the method in this study successfully detects all targets, except one missing detection and two false detections. This result shows that the method in this study has high accuracy and reliability in the detection of salient targets of UAV video reconnaissance images. Through the analysis of the results, it can be seen that the method in this paper provides an effective way for the accurate detection of salient targets, with high feasibility and significant effectiveness. In contrast, there are five missed and four false detections in the detection results of the chromatography-mass spectrometry method shown in Fig. 6. This data is significantly higher than the data of the method in this paper, indicating that the effect of the method of chromatography-mass spectrometry in target detection is not ideal. The detection results of the joint transform correlator method in Fig. 7 are even more disappointing, with 8 missed and 6 false detections. This result shows the shortcomings of the method of joint transform correlator in multi-target detection. In conclusion, by comparing the target detection effect of the three methods, it can be seen that the method in this paper has high accuracy and low false detection rate in multi-target detection, and can play an important role in UAV video reconnaissance.

To verify the performance of the salient target detection method in this paper, 1 frame of image is selected in the test set for the experiment. Also compare the result with the experimental result of the chromatography-mass spectrometry method and the joint transform correlator method, as shown in Fig. 8 to Fig. 11.

Fig. 8 shows a scene with multiple targets and a complex background, increasing the difficulty of target detection. By comparing the detection results in Fig. 9, Fig. 10, Fig. 11, and Fig. 8, the advantages of this method in dealing with such complex scenes are evident. In the detection result of this method, salient targets are detected, and there is no false detection or missed detection. On the contrary, both the chromatography-mass spectrometry method and the joint transform correlator method have problems in processing Fig. 8. There are three missed detections in the detection of salient targets by the chromatography-mass spectrometry method, and the detected targets are not clear, which has the problem of shadow occlusion. The detection effect of the joint transform correlator method is worse. There are not only five missed detections but also the green belt in the middle of the road has been detected. In addition, the detected target is accompanied by obvious shadows, which makes the target unclear. By conducting a comparative analysis, it is concluded that the method in this study can detect salient targets more accurately and clearly, which proves the effectiveness of this method in complex scenes.

The frame images in the test set are selected for the image enhancement experiment. Compared with the joint transform correlator method and the chromatography-mass spectrometry method, the original plots and the experimental results of the three methods are shown in Fig. 12 to Fig. 15.

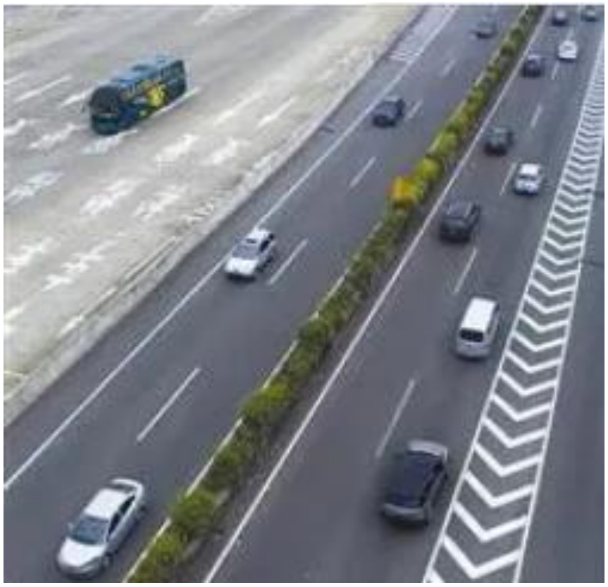

Fig. 8. Original drawing.

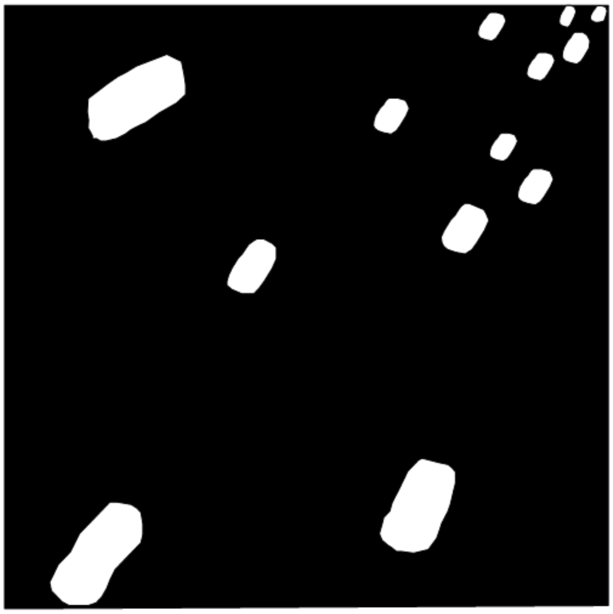

Fig. 9. Result of salient target detection of the proposed method.

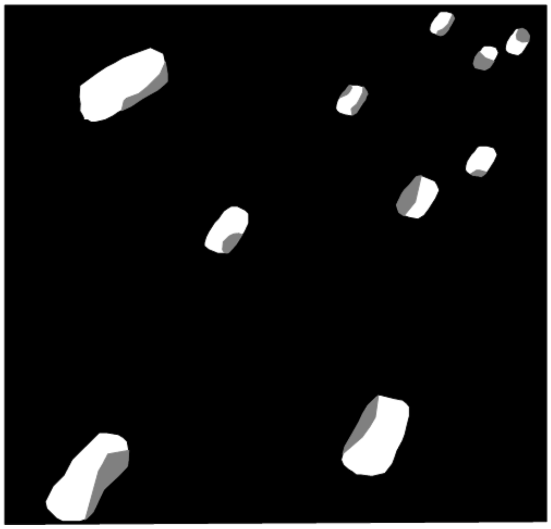

Fig. 10. Result of salient target detection by chromatography-mass spectrometry.

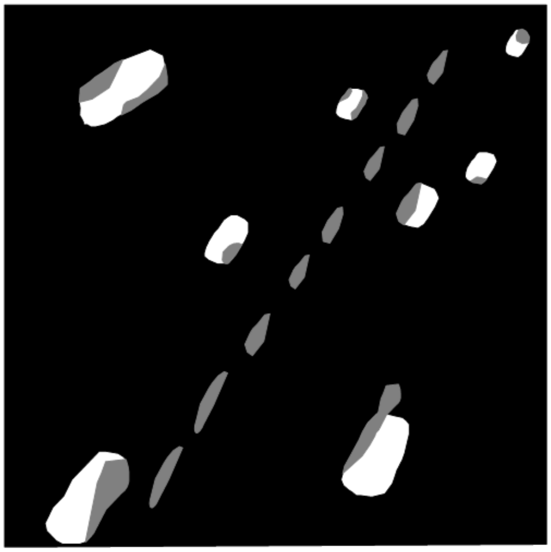

Fig. 11. Result of the salient target detection by joint transform correlator.

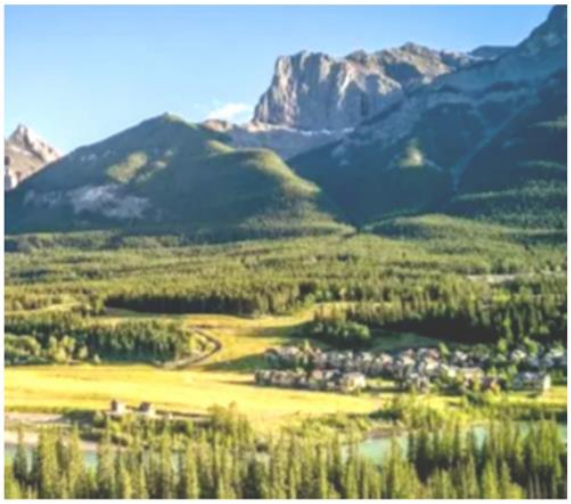

Fig. 12. Original drawing.

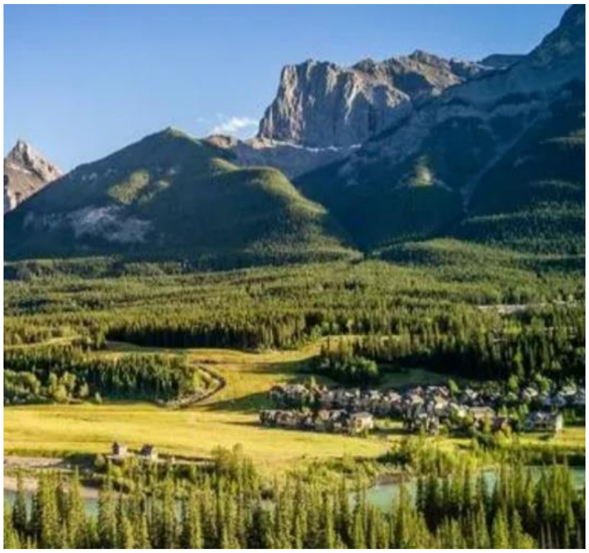

Fig. 13. Enhancement effect of the proposed method.

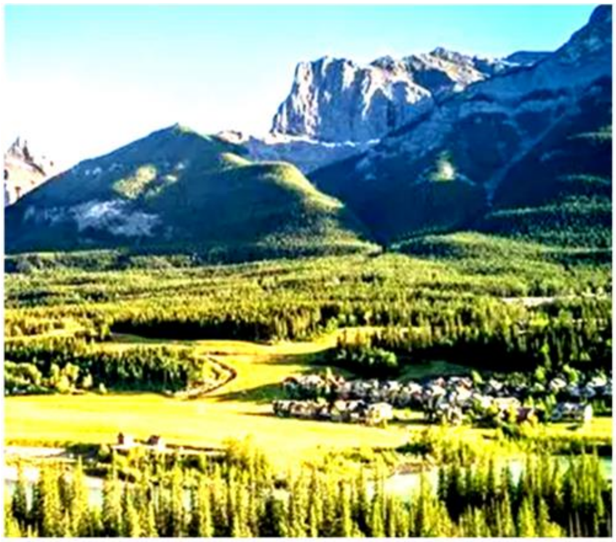

Fig. 14. Enhancement effect of joint transform correlator.

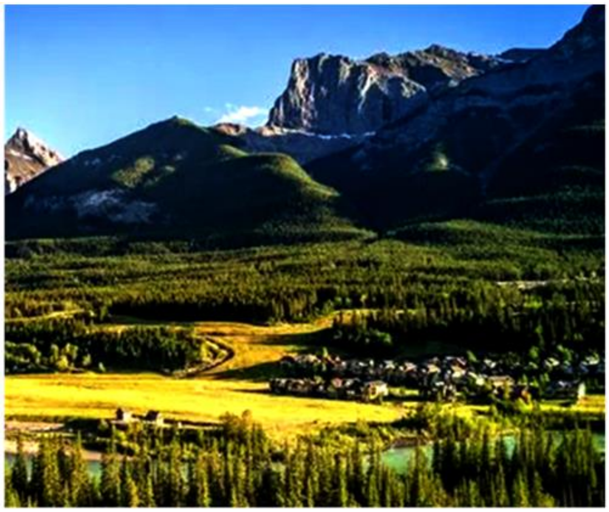

Fig. 15. Enhancement effect of chromatography-mass spectrometry method.

By comparing Fig. 13 and Fig. 12, the advantages of this method in image enhancement are apparent. The original image in Fig. 12 is slightly fuzzy and the details are not clear enough, but after the enhancement processing of this method, the image in Fig. 13 becomes very clear and the details are presented vividly. More importantly, the color of the enhanced image is soft, which is closer to the visual effect of the real scene, providing a good basis for subsequent tasks such as target detection. In contrast, although the joint transform correlator method in Fig. 14 has some effect of enhancement, the color is too bright, even some dazzling, giving a sense of unnatural. This kind of too-bright color will cover up some important details, causing trouble for subsequent tasks. The chromatography-mass spectrometry method in Fig. 15 has insufficient effect of enhancement. The overall image is dark and some details are not clear enough. Such an enhancement effect will make subsequent tasks such as target detection more difficult. In conclusion, the method in this paper performs well in the enhancement of UAV video reconnaissance images. It can not only significantly improve the image clarity, but also maintain the authenticity and softness of colors. The enhancement effect has a positive role in promoting the detection of salient targets.

To further validate the object detection performance of the proposed method, the detection performance of the three methods was compared using accuracy, recall, F1 score, and average detection time as objective indicators. The results are shown in Table III.

TABLE III. UAV CONFIGURATION TABLE

Method	Precision	Recall	F1 Score	Average detection time (seconds)
Chromatography-mass Spectrometry Method	0.75	0.68	0.71	5.2
Joint Transform Correlator	0.80	0.72	0.76	4.8
Proposed Method	0.90	0.85	0.87	0.3

According to Table III, the chromatography-mass spectrometry method shows relatively low accuracy and recall, with values of 0.75 and 0.68, respectively. This indicates that the method has certain errors in distinguishing significant and non-significant targets, and may miss some targets. The joint transformation correlator method has improved in accuracy and recall, reaching 0.80 and 0.72 respectively, demonstrating better object detection capability. The method proposed in this paper performs well in both accuracy and recall, reaching 0.90 and 0.85 respectively, significantly higher than the other two methods, indicating that this method can more accurately identify and detect significant targets. The F1 score of proposed method reached 0.87, which is much higher than that of the chromatography-mass spectrometry method (0.71) and the combined transform correlation method (0.76), further verifying the superiority of the proposed method. In terms of detection speed, the method proposed in this paper demonstrates significant advantages, with an average detection time of only 0.3 seconds, far lower than the chromatography-mass spectrometry method (5.2 seconds) and the combined transform correlation method (4.8 seconds). This indicates that the method proposed in this paper has higher real-time performance in processing drone video reconnaissance images. In summary, the CNN based method for detecting salient objects in unmanned aerial vehicle (UAV) video reconnaissance images proposed in this paper outperforms the compared methods in terms of accuracy, recall, F1 score, and detection speed. The experimental results show that the method proposed in this paper can accurately and efficiently detect salient targets in drone video reconnaissance images, which is of great significance for improving the efficiency and accuracy of reconnaissance work.

IV. CONCLUSION

With the popularization of UAV technology, the application of UAVs in the field of video reconnaissance is more and more extensive. However, the images taken by UAVs are often affected by factors such as illumination variation, camera angles, etc., which brings certain difficulties to target detection. To solve the problem, this study proposes a CNN-based salient target detection method for UAV video reconnaissance images. According to Faster R-CNN, the salient target detection of UAV video reconnaissance images is realized. For the salient target that cannot be detected in extreme cases, the detection is completed using the salient target calculation method of this paper. To make the detection effect better, the salient target enhancement algorithm is set to complete the salient target detection of UAV video reconnaissance image. Through experimental verification, the method of this paper has high accuracy and low false detection rate, can detect salient targets more accurately and clearly, and performs well in the enhancement of UAV video reconnaissance images. However, the effectiveness of this method largely depends on the quality and diversity of the training dataset. If the training dataset cannot fully cover various complex scenarios that may be encountered in practical applications, the generalization ability of the model may be limited. In addition, image differences under different drone platforms and shooting conditions may also affect the detection performance of the model. Therefore, in the future, addressing this issue will be the focus.

COMPETING OF INTERESTS

The authors declare no competing of interests.

AUTHORSHIP CONTRIBUTION STATEMENT

Li Na: Writing-Original draft preparation, Conceptualization, Supervision, Project administration.

DATA AVAILABILITY

On Request

DECLARATIONS

Not applicable

CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

AUTHORS' STATEMENT

The manuscript has been read and approved by all the authors, the requirements for authorship, as stated earlier in this document, have been met, and each author believes that the manuscript represents honest work.

Funding

Not applicable

Ethical approval

All authors have been personally and actively involved in substantial work leading to the paper, and will take public responsibility for its content.

Competing of Interests

The authors declare no competing of interests.

References

[1] Y. Hu, X. Wang, W. Li, X. Hei, and G. Xie, “A New Ship Target Detecting Method Based on Saliency in SAR Image,” in 2021 7th Annual International Conference on Network and Information Systems for Computers (ICNISC), IEEE, 2021, pp. 241–245.

[2] V. R. S. Mani, A. Saravanaselvan, and N. Arumugam, “Performance comparison of CNN, QNN and BNN deep neural networks for real-time object detection using ZYNQ FPGA node,” Microelectronics J, vol. 119, p. 105319, 2022.

[3] K. Lee et al., “STEM image analysis based on deep learning: identification of vacancy defects and polymorphs of MoS2,” Nano Lett, vol. 22, no. 12, pp. 4677–4685, 2022.

[4] C. Jirayupat et al., “Image Processing and Machine Learning for Automated Identification of Chemo-/Biomarkers in Chromatography–Mass Spectrometry,” Anal Chem, vol. 93, no. 44, pp. 14708–14715, 2021.

[5] A. K. Cherri and A. S. Nazar, “Class-associative multiple target recognition for highly compressed color images in a joint transform correlator,” Optical Engineering, vol. 61, no. 12, p. 123102, 2022.

[6] I. García-Aguilar, R. M. Luque-Baena, and E. López-Rubio, “Improved detection of small objects in road network sequences using CNN and super resolution,” Expert Syst, vol. 39, no. 2, p. e12930, 2022.

[7] E. López-Rubio, M. A. Molina-Cabello, F. M. Castro, R. M. Luque-Baena, M. J. Marín-Jiménez, and N. Guil, “Anomalous object detection by active search with PTZ cameras,” Expert Syst Appl, vol. 181, p. 115150, 2021.

[8] M. Gazzea, M. Pacevicius, D. O. Dammann, A. Sapronova, T. M. Lunde, and R. Arghandeh, “Automated power lines vegetation monitoring using high-resolution satellite imagery,” IEEE Transactions on Power Delivery, vol. 37, no. 1, pp. 308–316, 2021.

[9] A. W. S. Putra, H. Kato, and T. Maruyama, “Infrared LED marker for target recognition in indoor and outdoor applications of optical wireless power transmission system,” Jpn J Appl Phys, vol. 59, no. SO, p. SOOD06, 2020.

[10] R. Akter, V.-S. Doan, T. Huynh-The, and D.-S. Kim, “RFDOA-Net: An efficient ConvNet for RF-based DOA estimation in UAV surveillance systems,” IEEE Trans Veh Technol, vol. 70, no. 11, pp. 12209–12214, 2021.

[11] D. Mishra, S. K. Singh, R. K. Singh, and D. Kedia, “Multi-scale network (MsSG-CNN) for joint image and saliency map learning-based compression,” Neurocomputing, vol. 460, pp. 95–105, 2021.

[12] K. Ogohara and R. Gichu, “Automated segmentation of textured dust storms on mars remote sensing images using an encoder-decoder type convolutional neural network,” Comput Geosci, vol. 160, p. 105043, 2022.

[13] Z.-H. Lin, A. Y. Chen, and S.-H. Hsieh, “Temporal image analytics for abnormal construction activity identification,” Autom Constr, vol. 124, p. 103572, 2021.

[14] S. Molavi Vardanjani, A. Fathi, and K. Moradkhani, “Grsnet: gated residual supervision network for pixel-wise building segmentation in remote sensing imagery,” Int J Remote Sens, vol. 43, no. 13, pp. 4872–4887, 2022.

[15] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, “Scalable recurrent neural network for hyperspectral image classification,” J Supercomput, vol. 76, no. 11, pp. 8866–8882, 2020.

[16] A. Ferdowsi, M. A. Abd-Elmagid, W. Saad, and H. S. Dhillon, “Neural combinatorial deep reinforcement learning for age-optimal joint trajectory and scheduling design in UAV-assisted networks,” IEEE Journal on Selected Areas in Communications, vol. 39, no. 5, pp. 1250–1265, 2021.

[17] P. D. Ledger, B. A. Wilson, A. A. S. Amad, and W. R. B. Lionheart, “Identification of Metallic Objects using Spectral MPT Signatures: Object Characterisation and Invariants,” arXiv preprint arXiv:2012.10376, 2020.

[18] S. El Mohtar, B. Ait-El-Fquih, O. Knio, I. Lakkis, and I. Hoteit, “Bayesian identification of oil spill source parameters from image contours,” Mar Pollut Bull, vol. 169, p. 112514, 2021.

[19] Z. Wang, J. Wang, K. Yang, L. Wang, F. Su, and X. Chen, “Semantic segmentation of high-resolution remote sensing images based on a class feature attention mechanism fused with Deeplabv3+,” Comput Geosci, vol. 158, p. 104969, 2022.

[20] A. Patra, A. Saha, and K. Bhattacharya, “High-resolution image multiplexing using amplitude grating for remote sensing applications,” Optical Engineering, vol. 60, no. 7, p. 73104, 2021.

[21] S. V. S. Diddi and L.-W. Ko, “Course-grained multi-scale EMD based fuzzy entropy for multi-target classification during simultaneous SSVEP-RSVP hybrid BCI paradigm,” International Journal of Fuzzy Systems, vol. 24, no. 5, pp. 2157–2173, 2022.

[22] R. Theagarajan et al., “Integrating deep learning-based data driven and model-based approaches for inverse synthetic aperture radar target recognition,” Optical Engineering, vol. 59, no. 5, p. 51407, 2020.

[23] X. Chen, R. Proietti, C.-Y. Liu, and S. J. Ben Yoo, “A multi-task-learning-based transfer deep reinforcement learning design for autonomic optical networks,” IEEE Journal on Selected Areas in Communications, vol. 39, no. 9, pp. 2878–2889, 2021.

[24] M. Goudarzi, M. Palaniswami, and R. Buyya, “A distributed deep reinforcement learning technique for application placement in edge and fog computing environments,” IEEE Trans Mob Comput, vol. 22, no. 5, pp. 2491–2505, 2021.

[25] S. Gupta, P. K. Rai, A. Kumar, P. K. Yalavarthy, and L. R. Cenkeramaddi, “Target classification by mmWave FMCW radars using machine learning on range-angle images,” IEEE Sens J, vol. 21, no. 18, pp. 19993–20001, 2021.

[26] A. Rizik, E. Tavanti, H. Chible, D. D. Caviglia, and A. Randazzo, “Cost-efficient FMCW radar for multi-target classification in security gate monitoring,” IEEE Sens J, vol. 21, no. 18, pp. 20447–20461, 2021.