RESEARCH ARTICLE

A Dual UAV Cooperative Positioning System With Advanced Target Detection and Localization

**RONG CHANG[1], ANNING PAN[2], KAILONG YU[3], CHENGJIANG ZHOU [ID][3],
AND YANG YANG [ID][3], (Member, IEEE)**

[1] Yuxi Power Supply Bureau of Yunnan Power Grid Company, Yuxi 650032, China
[2] School of Physics and Electronic Information, Yunnan Normal University, Kunming 650500, China
[3] School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China

Corresponding authors: Chengjiang Zhou (chengjiangzhou@foxmail.com) and Yang Yang (yyang_ynu@163.com)

ABSTRACT Utilizing Unmanned Aerial Vehicles (UAVs) plays a pivotal role in the localization of ground construction personnel targets within the specialized operational environments of the energy industry. To ensure the safety of personnel in these areas effectively, this study introduces an innovative dual UAV cooperative positioning system designed to address the challenges of personnel localization in various complex settings. In the first place, the research establishes a dual UAV collaborative remote sensing platform, specifically tailored for executing dual UAV flight control and ground personnel identification and localization tasks. Subsequently, this study proposes a high-precision ground personnel localization solution. This solution integrates the flight data of two UAVs with the pixel coordinates of targets in natural remote sensing imagery, enabling the precise localization of moving targets within complex environments. Moreover, due to the irregularity of ground personnel shapes in imagery, this research introduces an improved rotating target detection network based on the YOLOv7 architecture. This network is specifically optimized for detecting ground targets in remote sensing imagery, significantly enhancing the accuracy of target localization. Extensive experimental validation has demonstrated that this localization system can maintain target localization errors within a 5-meter threshold, effectively enhancing the safety and efficiency of personnel in specific operational areas of the energy industry.

INDEX TERMS Dual-UAV cooperative positioning, rotating target detection, remote sensing platform, YOLOv7.

I. INTRODUCTION

With the development of remote sensing technology, UAVs are used in an increasingly wide range of fields, such as intelligence agriculture [1], [2], [3], [4], [5], power inspection [6], [7], [8], environmental monitoring [9], [10], [11], and intrusion safety monitoring [12]. In these applications, target positioning technology becomes one of the key functions of UAVs, which helps to achieve accurate data collection and monitoring tasks. UAV target positioning is highly flexible and widely applicable, enabling high-precision, real-time target position measurement in complex terrain and harsh environments.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhongyi Guo [ID].

Current UAV-based target positioning methods [13], [14], [15], [16] are broadly divided into two categories, active target positioning and passive target positioning methods. Active target positioning methods usually require UAVs to be equipped with professional sensors such as laser range finders and radars. Through the sensor data, the distance from the target to the UAV is directly obtained to find out the target's position information. Passive target positioning methods usually use only optical sensors, such as multispectral cameras or infrared cameras. These positioning methods can usually be subdivided into methods based on monocular and stereo vision. Monocular vision positioning methods [17] use a monocular camera and can calculate the target's position relative to the UAV by methods such as triangulation or depth estimation using neural networks.

The stereo usually use binocular cameras or simulates binocular cameras by using photos taken by monocular cameras at different locations to form stereo vision [18], and the position coordinates of the target can be found based on parallax information.

However, each of these methods has its own limitations. Active target positioning methods, as UAVs need to carry professional high-precision sensors, not only substantially increase the cost of UAVs, but also increase the power consumption of UAVs and reduce the range of flight. It is not suitable for target positioning work in large-scale areas. Stereo vision localization method based on binocular camera, the localization accuracy and maximum distance receive the limitation of baseline length of binocular camera. A shorter baseline cannot perform accurate depth calculation for distant targets, while a longer baseline will affect the flight status of the UAV and cannot be carried on the UAV. The stereo vision localization method based on monocular camera needs to realize stereo vision through the displacement of the UAV, and this method usually performs poorly for the localization of moving objects due to the large time difference in the acquisition of stereo vision images. The monocular camera-based triangulation localization method requires the use of known reference scales to construct spatial triangles, and the monocular depth estimation method [19] usually requires that the predicted scene has a similar depth distribution trend to the training scene, and that suitable scale information can be found to convert the relative depth to absolute depth. Therefore, both methods are not suitable for target localization in complex scenes.

In order to solve these problems mentioned above, we propose a new solution. This solution uses a cooperative companion flight operation mode of two UAVs to re-identify human targets in the remote sensing images of each of the two UAVs. Based on the respective flight information of the UAVs, a spatial localization model is constructed.

The main contributions of this paper are as follows:

1) A dual UAV collaborative remote sensing platform is built. This platform is used for the deployment of cooperative missions to dual UAVs.
2) A rotating target detection network for remote sensing images is proposed. This network is based on the YOLOv7 network structure and improves the rotating targets. It effectively improves the performance of target localization methods.
3) A high-precision spatial localization model based on the Person Re-identification method is proposed. This method uses the flight information of two UAVs and the target pixel coordinates of remote sensing images as parameters, and is able to precisely locate moving targets in complex scenes.

II. RELATED WORK

We review several currently used monocular UAV-based target positioning methods, as well as several target detection models and person re-identification (ReID) models that are useful for our approach.

A. TARGET POSITIONING METHODS BASED ON UAV

UAV-based target positioning methods typically involve the fields of computer vision, remote sensing, and machine learning. Depending on the flight altitude and camera parameters of the UAV, target localization can be performed by monocular or binocular vision systems. Monocular vision positioning methods usually require ground control points, terrain information, or known reference points. In contrast, binocular vision localization methods use two cameras with relative position relationships to capture images simultaneously and calculate the 3D coordinates of the target by comparing images from different viewpoints.

Jianghong et al. [20], explored methods for obtaining target motion information (e.g., position, velocity, and size) during target tracking and emphasized the importance of target detection and identification techniques. Huang et al. [21], proposed a target localization method based on laser ranging and solved the problem of photoelectric target consistency by image tracking, making it applicable to ground and sea surface target localization of UAVs. Qu et al. [22], calculated the target position by solving nonlinear equations, addressed the issue that the target height value could not be accurately determined in the traditional method, and obtained the optimal UAV formation shape through simulation analysis.

While all current methods possess their own advantages in target localization, there are still limitations in localizing moving objects in complex terrain.

B. TARGET DETECTION

Target detection was an important task in computer vision, aiming to identify and localize pedestrians in images or videos. Traditional target detection methods were mainly based on hand-designed features (e.g., HOG and SIFT) and sliding window search techniques. However, these methods faced certain challenges when dealing with complex scenes. In recent years, deep learning techniques made breakthroughs in the field of target detection, bringing new possibilities for pedestrian detection. Deep learning-based target detection methods, such as Faster R-CNN [23], YOLO [24], [25], and SSD [26], achieved good results in various scenarios. These methods were able to automatically learn the feature representation by training a large amount of labeled data, thus achieving highly accurate target detection in complex scenes.

In this field, target detection for UAV remote sensing images became an important research direction. Due to the high mobility and flexibility of UAVs, they could monitor large areas in complex terrains and environments, so they showed greater potential in target detection tasks. However, pedestrian detection from UAV remote sensing images still faced many challenges, such as target size variation, occlusion, and light variation. To address these problems, researchers proposed various improvement strategies [27],

[28], [29], [30], such as introducing multi-scale detection, employing data enhancement techniques, and rotational target detection methods that incorporated angular information. With these improvements, target detection algorithms for UAV remote sensing images achieved better performance in practical applications.

C. PERSON RE-IDENTIFICATION

The task of person re-identification (ReID) was to recognize and match the same pedestrian under different camera views, lighting conditions, and times of day. ReID involved techniques in feature extraction, feature matching, and similarity calculation. In recent years, with the development of deep learning, ReID methods based on convolutional neural networks (CNNs) achieved significant performance improvements. These methods learned feature representations that could distinguish different pedestrians by training a large amount of labeled data. In practical applications, ReID techniques were widely used in scenarios such as video surveillance, intelligent transportation, and security.

III. MATERIALS AND METHODS

The proposed system consists of three main components: (1) dual-UAV collaborative remote sensing platform, (2) target detection method, and (3) person re-identification method.

A. DUAL-UAV COLLABORATIVE REMOTE SENSING PLATFORM

The dual-UAV collaborative remote sensing Platform we construct consists of three main components, including the UAV flight control terminals, cloud services, and deployment and computing terminal.

1) UAV flight control terminal: responsible for receiving and deploying flight missions, while collecting UAV remote sensing images and its camera external parameters (such as latitude, longitude, altitude, pitch angle, etc.) in real time, and uploading these information to the cloud service.
2) Cloud service: mainly responsible for data forwarding, including information of UAV's mission, remote sensing images and camera's external parameters, etc., to provide the required data support for other parts of the platform.
3) Deployment and computing terminal: Visually arrange the flight missions through the map module and receive the data forwarded by the cloud service. Users can visually plan the flight route of the UAV on the map, while monitoring the flight status of the UAV in real time.

In order to ensure the safety of UAV flight, the mission planning needs to be set with reference to Digital Elevation Model (DEM), which can help us understand the terrain features and thus determine the safe and suitable flight altitude to ensure that the UAV obtains better remote sensing data during the mission.

The flight routes of the two UAVs are kept parallel to the mission route direction and located on the left and right sides respectively. The separation distance of the two UAVs is related to the flight altitude of the UAVs. If the lens pitch angle of the two UAVs is about 45 degrees, the horizontal distance between the two UAVs and the mission route is equal to the flight altitude of the UAVs. The lens pitch angle of the UAV varies with the DEM to ensure that the observation area of the two UAVs has the maximum overlap rate and the observation centers overlap up as much as possible. If the UAVs are flown too close together, the parallax will not be significant enough to calculate accurate positioning results.

The remote sensing images are pushed and streamed through the RTMP service and aligned by flight data timestamps. To reduce the computational errors caused by the delay, the two UAVs need to ensure that the timestamps are aligned.

B. TARGET DETECTION

In the UAV remote sensing images, the angle and position differences between the UAV and the target can cause the target to present different angles and features from the ground image. Especially when detecting the human body, due to the distance and angle, it often leads to missed detection or inaccurate bounding box due to too small target image or angular tilt, resulting in less accurate target positioning.

In order to solve this problem, we adapt YOLOv7 [31] for rotating target detection, so that the target can always be surrounded by an accurate bounding box and the detection of small targets is improved. The structure of the model is shown in Figure 2. The model's output is formatted as (cls, x, y, $long\_side$, $short\_side$, θ, $Conf$, $180/2r$), where cls represents the categorical classification of the identified object. The variables x and y denote the horizontal and vertical coordinates of the top-left corner of the detection box. Additionally, $long\_side$ and $short\_side$ specifically indicate the lengths of the longer and shorter sides of the detection box, respectively. θ indicates the angle by which the pixel coordinate system's u-axis is rotated clockwise to align with the longest side. $Conf$ is the probability that the detected object actually exists. $180/2r$ represents the confidence level in the rotation angle of the candidate bounding box for the object. This method converts angle prediction into a classification problem by dividing the range of angles into $2r$ classes, each representing a specific range of angles. It effectively addresses the cyclical nature of angle prediction, resulting in more stable output from the model.

In our study, angular information is learned using the Circular Smooth Label (CSL) method [32]. CSL is a scheme that implements the idea of angle regression with classification. CSL is described as Equation 1. $g(x)$ is a window function, and the window radius is controlled by r. Each detected target outputs (cls, ϖ, $180/2r$) values, which include cls confidence in the target category, ϖ parameters for the bounding box (x, y, $long\_side$, $short\_side$, θ), $Conf$ for
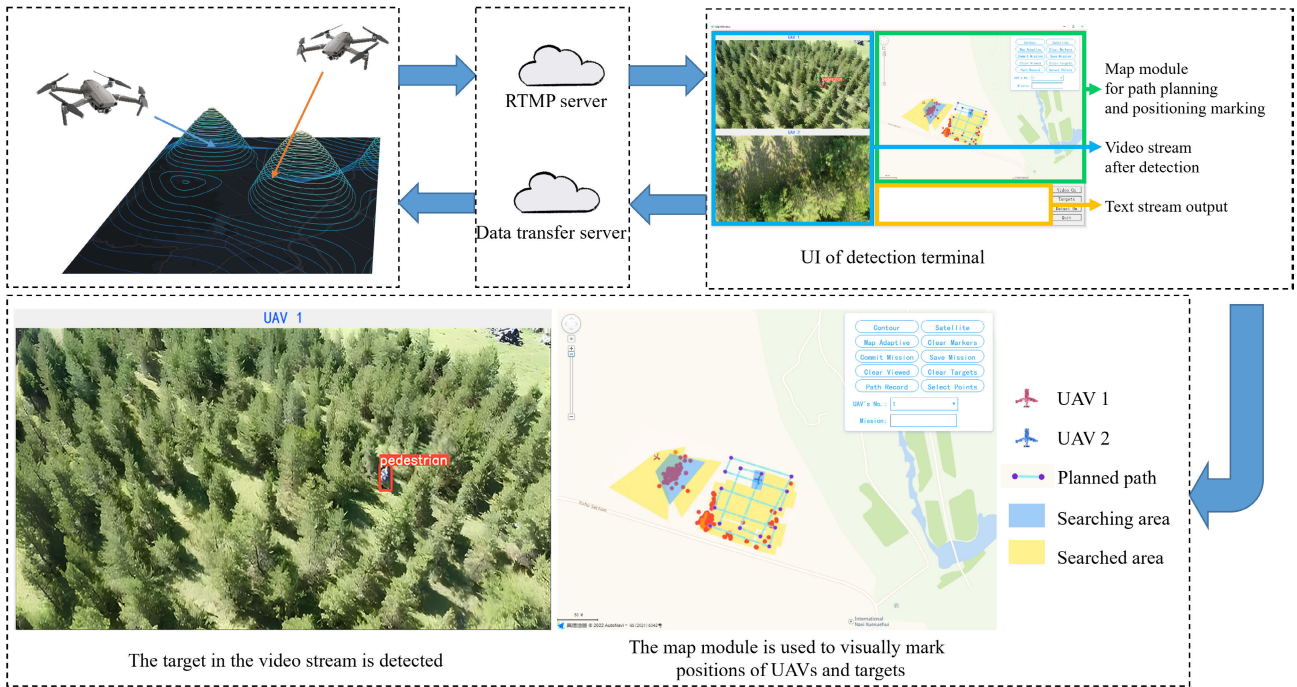
FIGURE 1. Platform structure. Two UAVs independently transmit information to the terminal server, where ground target identification and positioning are performed. The processed information is then sent back to the UAVs for subsequent operations. The real-time video feed from the two UAVs is represented by blue boxes, the green boxes delineate the target areas searched by the UAVs based on mission deployment, and the yellow box displays the UAV's flight log along with the results of the positioning algorithm.
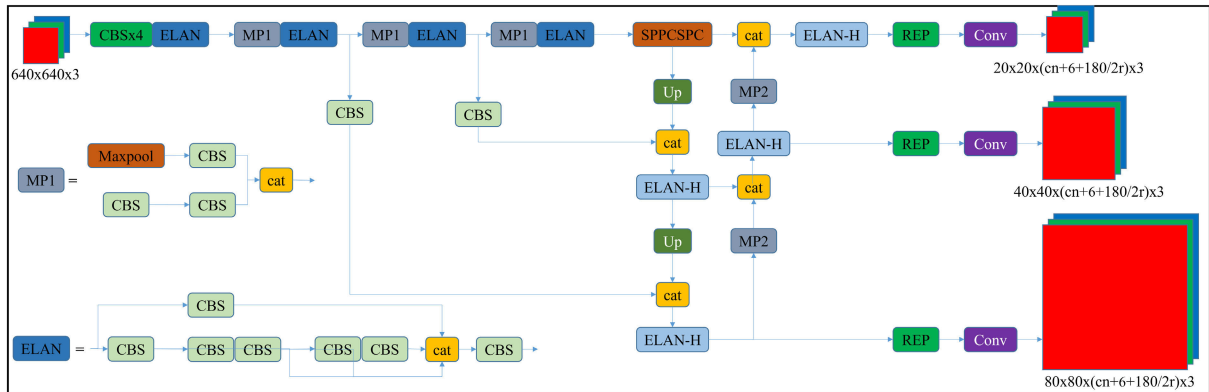

FIGURE 2. Architecture of the adapted YOLOv7 for rotating object detection. The output of the model is modified to (cls, ϖ, $180/2r$) × 3, where ϖ represents the bounding box parameters (x, y, $long\_side$, $short\_side$, θ, $Conf$), where x and y denote the horizontal and vertical coordinates of the top-left corner of the detection box. $long_side$ and $short_side$ indicate the lengths of the longer and shorter sides of the detection box and $180/2r$ is the number of categories for angle classification. cat: Concatenation, Up: Image upsampling. This modification allows the model to predict the angle of the object as a classification problem, effectively eliminating the boundary problem that often occurs in regression-based methods.

the object's existence, and $180/2r$ confidence in the rotation angle.

$$CSL(x) = \begin{cases} g(x), & \theta - r < x < \theta + r \\ 0, & otherwise \end{cases} \quad (1)$$

C. DUAL-UAV STEREOSCOPIC POSITIONING

The main idea of the method is to use Person Re-identification (ReID) method to identify the same target in the remote sensing images of UAV from two different views, and then solve the position coordinates of the target by establishing the spatial coordinate equation relationship to achieve high precision stereo position localization.

ReID focuses on identifying and matching the same pedestrian under different camera views, lighting conditions and time. After obtaining the pedestrian detection results from the remote sensing images of two UAVs, the pedestrian bounding boxes of the two viewpoints are matched to determine the correspondence of the same target in the two viewpoints.

In our study, we utilized some existing ReID models, such as CTL-Model [33], DLFOS [34]. These models were
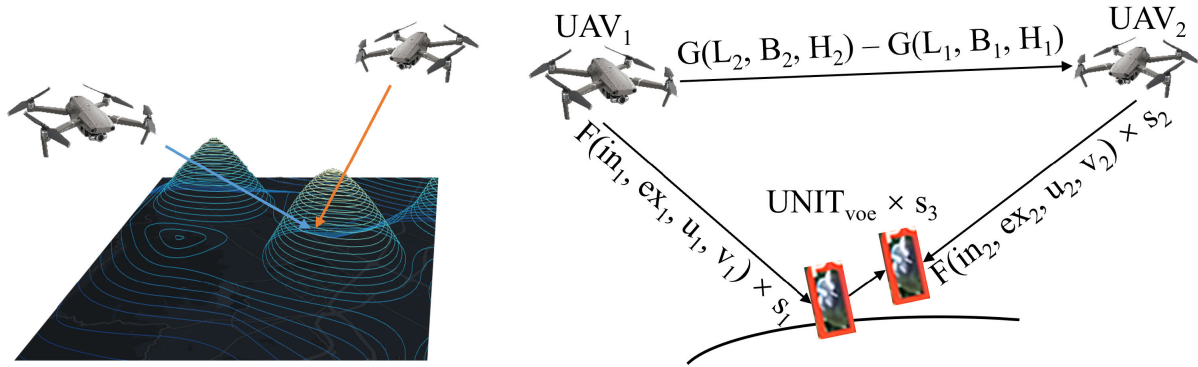
FIGURE 3. Dual UAV stereo positioning schematic. Where, $F(in, ex, u, v)$ represent the mapping relationship between pixel coordinates and direction vectors in the geocentric space-rectangular coordinate system, with *in* being the internal reference of the camera and *ex* being the external reference of the camera. $G(L, B, H)$ denote the mapping relationship between latitude and longitude and the coordinates of the geocentric spatial Cartesian coordinate system, L, B, H denote the latitude, longitude and altitude of the UAV, respectively.
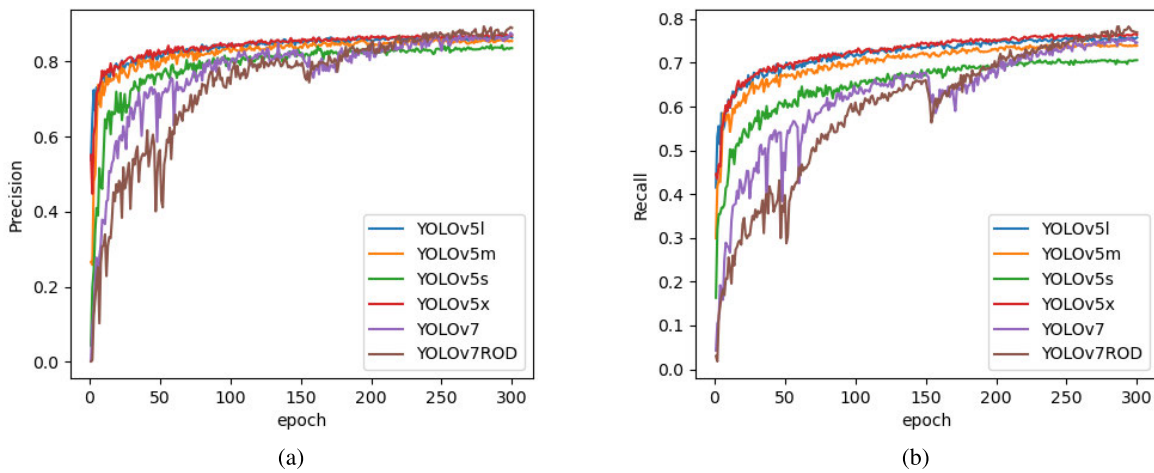

FIGURE 4. The variation trend of precision(a) and recall(b) during training.

chosen because they have proven their effectiveness and robustness in personnel re-identification tasks in previous studies. Although we did not make any modifications to the original ReID model, its application in the context of our study is of significant value. The model plays a key role in identifying the same target in the remote sensing images of each of the two UAVs, which is a critical step in our proposed spatial localization approach.

After determining the correspondence of the targets, the targets are represented in the coordinate systems of the two UAVs, respectively, and the spatial coordinate equation can be constructed by the two representations and the conversion relationship between the two coordinate systems.

The work use $F(in, ex, u, v)$ to represent the mapping relationship between pixel coordinates and direction vectors in the geocentric space-rectangular coordinate system, with *in* being the internal reference of the camera and *ex* being the external reference of the camera. The work use $G(L, B, H)$ to denote the mapping relationship between latitude and longitude and the coordinates of the geocentric spatial Cartesian coordinate system, L, B, H denote the latitude, longitude and altitude of the UAV respectively. The spatial

TABLE 1. Performance of different models on the RTX 2070 Super.

Model	Precision	Recall	mAP@.5	S(ms)
YOLOv5s	0.848	0.673	0.726	**5.5**
YOLOv5m	0.881	0.693	0.752	7.1
YOLOv5l	**0.904**	0.698	0.763	10.0
YOLOv5x	0.889	0.726	0.778	16.1
YOLOv7	0.873	0.741	0.822	8.5
YOLOv7ROD	0.881	**0.773**	**0.838**	8.6

coordinate equation can be expressed as:

$$F(in_1, ex_1, u_1, v_1) \times s_1 - F(in_2, ex_2, u_2, v_2) \times s_2$$
$$+ UNIT_{voe} \times s_3 = G(L_2, B_2, H_2) - G(L_1, B_1, H_1),$$
$$(2)$$

The parameters related to both UAVs are distinguished by the subscripts 1, 2. s_1, s_2 represent the scaling of the distance of the target relative to the UAV along the vector direction with respect to the vector length, and s_3 represents the error distance, where a smaller error distance means a more accurate result is obtained. Use $[a_i, b_i, c_i]$ to represent the

FIGURE 5. Partial pictures of dataset.

TABLE 2. Part of the UAV's parameters.

	Mavic 2 Pro	Phantom 4
FOV	77°	94°
Hover precision	vertical: ± 0.5m	vertical: ± 0.5m
	horizontal: ± 1.5m	horizontal: ± 1.5m
Maximum take-off altitude	6km	6km
Maximum flight time	31min(25km/h)	28min
Effective distance with maximum signal	FCC:10km	FCC:5km
Quality of real-time picture transmission	720P@30fps	720P@30fps
Delay of Picture transmission	120 - 130 ms	220ms

FIGURE 6. Target detection is performed for targets in remote sensing images taken at different locations at the same moment, and the correspondence between targets is established using the ReID method. The two images in the same column are images of the same set of different locations taken at the same time, and we have marked the same targets with the same color for identification. More obvious is the second set of images, where the red anchor box is the corresponding target after re-identification, and the pink is the target not found in the corresponding figure.

direction vector of the UAV pointing to the target in the geocentric space right-angle coordinate system, and $[a, b, c]$ to represent the vector of UAV 1 pointing to 2. Since the UAV observation positions are different, using the three basis vectors separately to represent the error direction ensures that at least one of the three equations has a unique solution. The equation can be expressed as follows:

$$
\begin{bmatrix} a_1 & a_2 & 1 \\ b_1 & b_2 & 0 \\ c_1 & c_2 & 0 \end{bmatrix} \times \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad (3)
$$

$$
\begin{bmatrix} a_1 & a_2 & 0 \\ b_1 & b_2 & 1 \\ c_1 & c_2 & 0 \end{bmatrix} \times \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad (4)
$$

$$
\begin{bmatrix} a_1 & a_2 & 0 \\ b_1 & b_2 & 0 \\ c_1 & c_2 & 1 \end{bmatrix} \times \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad (5)
$$

Taking the set of results with the smallest error distance, s_3, the final localization result of the target can be expressed as:

$$
GPS_{target} = \frac{1}{2}\{F(in_1, ex_1, u_1, v_1) + G(L_1, B_1, H_1) \\ + F(in_2, ex_2, u_2, v_2) + G(L_2, B_2, H_2)\}, \quad (6)
$$

IV. EXPERIMENTS

According to the classification of flight mode, UAVs can be divided into fixed-wing UAVs, helicopters, multi-rotor UAVs and so on. Compared with UAVs of other flight methods, multi-rotor UAVs can take off and land vertically without a runway and hover in the air, and it has the best stability and ease of use. Considering these advantages, we choose to use a multi-rotor UAV as a remote sensing device for my experiments.

Two models of UAVs are used in the experiment: the DJI Mavic 2 Pro and the DJI Phantom 4. The use of different UAV models aim to verify that accurate positioning tasks can be accomplished using any UAV model that can provide the necessary parameters for the calculations. The UAV models are shown in Table 2.

In this section, we demonstrate the effectiveness of this system in a practical application. The experiment is divided into four parts: (1) target detection, and (2) target Positioning.

A. TARGET DETECTION

We train YOLOv7 using a remote sensing dataset we constructed. The dataset has two sources, one part is from the public dataset VisDrone2019 and the other part is from the remote sensing images we collected with the UAV. Parts of
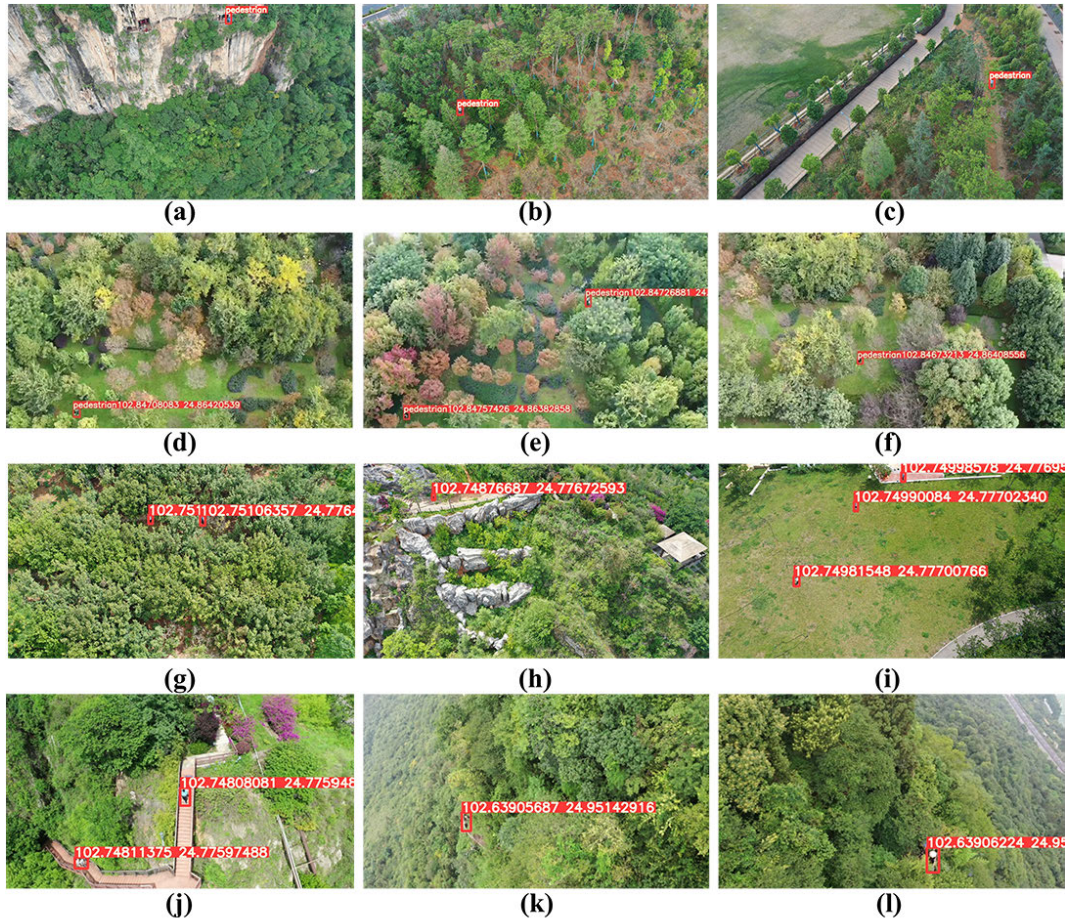
FIGURE 7. Target positioning result display. The detected targets are selected by the target corresponding color box and marked with the positioning result.

TABLE 3. Target positioning result.

Target	True longitude	True latitude	Positioning lng	Positioning lat	Error(m)
a1	102.640 111 06°	24.951 548 48°	102.640 137 10°	24.951 525 47°	5.79
b1	102.859 217 08°	24.868 319 75°	102.859 232 215°	24.868 314 08°	2.53
c1	102.859 889 855°	24.868 687 76°	102.859 905 52°	24.868 679 18°	2.30
d1	102.847 077 73°	24.864 212 38°	102.847 080 83°	24.864 205 39°	2.50
e1	102.847 552 01°	24.863 834 21°	102.847 574 26°	24.863 828 58°	3.21
e2	102.847 256 63°	24.863 895 54°	102.847 268 81°	24.863 892 46°	1.755
f1	102.846 724 73°	24.864 110 98°	102.846 732 13°	24.864 085 56°	3.51
g1	102.751 009 80°	24.776 486 35°	102.751 001 03°	24.776 471 37°	1.96
g2	102.751 072 92°	24.776 421 59°	102.751 063 57°	24.776 405 62°	2.09
h1	102.748 796 78°	24.776 721 28°	102.748 766 87°	24.776 725 93°	5.04
i1	102.749 808 07°	24.777 012 05°	102.749 815 47°	24.777 007 65°	1.265
i2	102.749 893 16°	24.777 027 97°	102.749 900 84°	24.777 023 40°	1.31
i3	102.749 972 73°	24.776 965 85°	102.749 985 78°	24.776 958 09°	2.23
j1	102.748 050 46°	24.775 945 20°	102.748 080 81°	24.775 948 15°	3.88
j2	102.748 078 38°	24.775 971 45°	102.748 113 75°	24.775 974 88°	4.53
k1	102.639 120 85°	24.952 119 31°	102.639 129 03°	24.952 077 62°	5.70
l1	102.639 089 45°	24.951 474 45°	102.639 056 87°	24.951 429 16°	7.56

the training and test sets are shown in Figure 5. To find a better model approach, we also trained several model variants of YOLOv5. The test results are shown in Table 5. The column S

in the table represents the time taken by the model to make predictions for a single image. The trend of precision and recall during training is shown in Figure 4.

B. TARGET POSITIONING

In order to compare the effectiveness of the localization platform in different scenarios, we selected different types of scenarios for localization experiments. These scenes include gentle areas (Yunnan Normal University, 24°51′51″ N 102°50′49″ E), forests (Gudian Wetland Park, 24°46′34″ N 102°44′57″ E), mountains and cliffs (Xishan Forest Park, 24°57′6″ N 102°38′24″ E). Each detected target and the predicted position are labeled in Figure 7. The error results of the target localization are shown in Table 3.

The positioning error distance is obtained by calculating the spatial distance between the actual and predicted coordinates. The results shown in the table indicate that in complex scenarios, 76.5% of the position accuracy errors remain within 5*m*. Positioning accuracy in extremely steep environments is generally lower than in other environments. However, the error can still be kept within 8*m*. In conclusion, this approach can meet the positioning requirements in complex situations.

V. CONCLUSION

This paper presents a new method for targeting using dual UAVs. The method improves the accuracy of target localization in complex scenes. Our proposed method requires establishing correspondence between targets in dual UAV remote sensing images. The correspondence relationships are predicted by a trained person re-identification model. Using the internal and external parameters of the dual UAV, we establish the equation of spatial three-dimensional coordinate relations, which in turn enables us to find the exact position of the target. This method accomplishes accurate localization of moving targets in complex terrain. In our future work, we will further explore ground target localization methods in complex multi-view remote sensing scenes by expanding datasets and incorporating semantic guidance. We aim to apply these methods to subsequent tracking and semantic analysis tasks. Additionally, we strive to enhance the computational speed and detection accuracy of ground target localization methods.

REFERENCES

[1] A. D. Boursianis, M. S. Papadopoulou, P. Diamantoulakis, A. Liopa-Tsakalidi, P. Barouchas, G. Salahas, G. Karagiannidis, S. Wan, and S. K. Goudos, "Internet of Things (IoT) and agricultural unmanned aerial vehicles (UAVs) in smart farming: A comprehensive review," *Internet Things*, vol. 18, May 2022, Art. no. 100187. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2542660520300238

[2] P. K. R. Maddikunta, S. Hakak, M. Alazab, S. Bhattacharya, T. R. Gadekallu, W. Z. Khan, and Q.-V. Pham, "Unmanned aerial vehicles in smart agriculture: Applications, requirements, and challenges," *IEEE Sensors J.*, vol. 21, no. 16, pp. 17608–17619, Aug. 2021.

[3] D. Popescu, F. Stoican, G. Stamatescu, L. Ichim, and C. Dragana, "Advanced UAV-WSN system for intelligent monitoring in precision agriculture," *Sensors*, vol. 20, no. 3, p. 817, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/3/817

[4] H. Hildmann and E. Kovacs, "Review: Using unmanned aerial vehicles (UAVs) as mobile sensing platforms (MSPs) for disaster response, civil security and public safety," *Drones*, vol. 3, no. 3, p. 59, Jul. 2019. [Online]. Available: https://www.mdpi.com/2504-446X/3/3/59

[5] S. H. Alsamhi, F. Afghah, R. Sahal, A. Hawbani, M. A. A. Al-Qaness, B. Lee, and M. Guizani, "Green Internet of Things using UAVs in B5G networks: A review of applications and strategies," *Ad Hoc Netw.*, vol. 117, Jun. 2021, Art. no. 102505. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1570870521000639

[6] H. Shakhatreh, A. H. Sawalmeh, A. Al-Fuqaha, Z. Dou, E. Almaita, I. Khalil, N. S. Othman, A. Khreishah, and M. Guizani, "Unmanned aerial vehicles (UAVs): A survey on civil applications and key research challenges," *IEEE Access*, vol. 7, pp. 48572–48634, 2019.

[7] M. N. Boukoberine, Z. Zhou, and M. Benbouzid, "A critical review on unmanned aerial vehicles power supply and energy management: Solutions, strategies, and prospects," *Appl. Energy*, vol. 255, Dec. 2019, Art. no. 113823. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261919315107

[8] D. Sadykova, D. Pernebayeva, M. Bagheri, and A. James, "IN-YOLO: Real-time detection of outdoor high voltage insulators using UAV imaging," *IEEE Trans. Power Del.*, vol. 35, no. 3, pp. 1599–1601, Jun. 2020.

[9] N. Mohamed, J. Al-Jaroodi, I. Jawhar, A. Idries, and F. Mohammed, "Unmanned aerial vehicles applications in future smart cities," *Technological Forecasting Social Change*, vol. 153, Apr. 2020, Art. no. 119293. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0040162517314968

[10] J.-C. Padró, F.-J. Muñoz, J. Planas, and X. Pons, "Comparison of four UAV georeferencing methods for environmental monitoring purposes focusing on the combined use with airborne and satellite remote sensing platforms," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 75, pp. 130–140, Mar. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0303243418306421

[11] K. Topouzelis, A. Papakonstantinou, and S. P. Garaba, "Detection of floating plastics from satellite and unmanned aerial systems (plastic litter project 2018)," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 79, pp. 175–183, Jul. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0303243419301163

[12] H. Sachdeva, S. Gupta, A. Misra, K. Chauhan, and M. Dave, "Improving privacy and security in unmanned aerial vehicles network using blockchain," 2022, *arXiv:2201.06100*.

[13] M. Vrba, Y. Stasinchuk, T. Báča, V. Spurný, M. Petrlík, D. Hert, D. Žaitlík, and M. Saska, "Autonomous capture of agile flying objects using UAVs: The MBZIRC 2020 challenge," *Robot. Auto. Syst.*, vol. 149, Mar. 2022, Art. no. 103970. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921889021002396

[14] I. Catalano, J. P. Queralta, and T. Westerlund, "Evaluating the performance of multi-scan integration for UAV LiDAR-based tracking," in *Proc. Int. Conf. FinDrones*, 2023, pp. 85–95.

[15] S. Zhang, J. Li, C. Yang, Y. Yang, and X. Hu, "Vision-based UAV positioning method assisted by relative attitude classification," in *Proc. 5th Int. Conf. Mathematics Artif. Intell.* New York, NY, USA: Association for Computing Machinery, 2020, pp. 154–160, doi: 10.1145/3395260.3395263.

[16] M. A. Kassab, A. Maher, F. Elkazzaz, and Z. Baochang, "UAV target tracking by detection via deep neural networks," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2019, pp. 139–144.

[17] Y. Zhou, Z. Yu, and Z. Ma, "UAV based indoor localization and objection detection," *Frontiers Neurorobotics*, vol. 16, Jul. 2022, Art. no. 914353. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnbot.2022.914353

[18] J.-C. Trujillo, R. Munguia, S. Urzua, E. Guerra, and A. Grau, "Monocular visual SLAM based on a cooperative UAV-target system," *Sensors*, vol. 20, no. 12, p. 3531, Jun. 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/12/3531

[19] K. Yu, H. Li, L. Xing, T. Wen, D. Fu, Y. Yang, C. Zhou, R. Chang, S. Zhao, L. Xing, and H. Bai, "Scene-aware refinement network for unsupervised monocular depth estimation in ultra-low altitude oblique photography of UAV," *ISPRS J. Photogramm. Remote Sens.*, vol. 205, pp. 284–300, Nov. 2023.

[20] W. Jianhong, R. A. Ramirez-Mendoza, and T. Xiaojun, "Target tracking algorithms for multi-UAVs formation cooperative detection," *Syst. Sci. Control Eng.*, vol. 9, no. 1, pp. 417–429, Jan. 2021, doi: 10.1080/21642583.2021.1916789.

[21] Z. Huang, Z. Huang, X. Tang, and S. Zou, "Study on multipoint-ranging-target-positioning technologies for UAV," in *Proc. 3rd Int. Conf. Electron. Inf. Technol. Comput. Eng. (EITCE)*, Oct. 2019, pp. 1405–1411.

[22] Y. Qu, B. Yang, Y. Song, and S. Xu, "Multi-UAVs cooperative ground moving target tracking and positioning method," in *Proc. 39th Chin. Control Conf. (CCC)*, Jul. 2020, pp. 3471–3476.

[23] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[24] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[25] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Oct. 2017, pp. 6517–6525.

[26] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Computer Vision—ECCV*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, pp. 21–37.

[27] Y. Xu, G. Yu, X. Wu, Y. Wang, and Y. Ma, "An enhanced viola-jones vehicle detection method from unmanned aerial vehicles imagery," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 7, pp. 1845–1856, Jul. 2017.

[28] Y. Hu, X. Wu, G. Zheng, and X. Liu, "Object detection of UAV for anti-UAV based on improved YOLO v3," in *Proc. Chin. Control Conf. (CCC)*, Jul. 2019, pp. 8386–8390.

[29] X. Guo, X. Li, Q. Pan, P. Yue, and J. Wang, "An object detection algorithm for UAV reconnaissance image based on deep convolution network," in *Proc. Int. Conf. Sens. Imag.*, E. T. Quinto, N. Ida, M. Jiang, and A. K. Louis, Eds. Cham, Switzerland: Springer, 2019, pp. 53–64.

[30] R. Niu, Y. Qu, and Z. Wang, "UAV detection based on improved YOLOv4 object detection model," in *Proc. 2nd Int. Conf. Big Data Artif. Intell. Softw. Eng. (ICBASE)*, Sep. 2021, pp. 25–29.

[31] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. EEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2022, pp. 7464–7475.

[32] X. Yang and J. Yan, "Arbitrary-oriented object detection with circular smooth label," in *Computer Vision—ECCV*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham, Switzerland: Springer, 2020, pp. 677–694.

[33] M. Wieczorek, B. Rychalska, and J. Dabrowski, "On the unreasonable effectiveness of centroids in image retrieval," in *Proc. Int. Conf. Neural Inf. Process.*, 2021, pp. 212–223.

[34] M. J. Khan and M. Rahman, "Person re-identification by discriminative local features of overlapping stripes," *Symmetry*, vol. 12, no. 4, p. 647, Apr. 2020. [Online]. Available: https://www.mdpi.com/2073-8994/12/4/647

ANNING PAN received the bachelor's degree in computer science and technology from Yunnan Normal University, in 2015, and the master's degree in computer application technology, in 2018. She is currently pursuing the Ph.D. degree in optical engineering with Yunnan Normal University. Since 2018, she has been a Teaching Assistant with the Big Data School, Baoshan College. Her published academic achievements include more than ten academic articles, including more than five SCI and EI retrieval articles. Her research interests include computer vision, pattern recognition, and remote sensing image registration. Her research results can be found in the academic homepage at https://www.researchgate.net/profile/Anning-Pan/research.

KAILONG YU received the bachelor's degree in computer science and technology from Minjiang University, in 2020. He is currently pursuing the master's degree in computer technology engineering with Yunnan Normal University. His current research interests include deep learning, remote sensing image processing, and monocular depth estimation.

CHENGJIANG ZHOU received the bachelor's degree in automation from North China University of Science and Technology, in 2014, and the master's degree in control theory and control engineering and the Ph.D. degree in metallurgical control engineering from Kunming University of Science and Technology, China, in 2017 and 2020, respectively. He is currently a Teacher with the Information School, Yunnan Normal University, China. He has published more than 30 academic articles, including more than 20 SCI articles. His research interests include signal processing and fault diagnosis research, mechanical equipment fault detection, mechanical equipment fault diagnosis, and mechanical fault diagnosis based on image processing.

RONG CHANG received the master's degree in computer application technology from Kunming University of Science and Technology. He is currently a Senior Engineer. In the past three years, he has published one monograph and more than ten Chinese core and EI-indexed scientific and technological articles, three of which have won the Second Prize of the Southern Power Grid Technology Forum, one of which has won the First Prize of the National Electric Power Industry Association's Paper, and have been granted more than 25 patents. As the Project Leader, he has undertaken multiple key power grid projects with a total budget of 12.0827 million yuan. His research interests include safety production monitoring, power information systems, platform software infrastructure operation and maintenance, data management, digital construction, and digital application innovation.

YANG YANG (Member, IEEE) received the master's degree in computer science from Waseda University, Tokyo, Japan, in 2007, and the Ph.D. degree in computer science from the National University of Singapore, Singapore, in 2013. He is currently a Professor with the School of Information Science and Technology, Yunnan Normal University, Kunming, China. His published academic achievements include more than 90 academic articles, including more than 85 SCI and EI retrieval articles. His research interests include computer vision, remote sensing, geography information systems, and medical imaging. His research results can be found in the academic homepage at https://www.researchgate.net/profile/Yang-Yang-219.

● ● ●