





# Air-to-Ground Active Object Tracking via Reinforcement Learning

Xin Liu, Weiya Ren, Jie Tan , Xiaochuan Zhang, Xiaoguang Ren ,  
and Huadong Dai

Academy of Military Science, Beijing, China  
rxg\_nudt@126.com

**Abstract.** Over the years, active object tracking has emerged as a prominent topic in object tracking. However, most of these methods are unsuitable for tracking ground objects in high-altitude environments. Therefore, the paper proposes an air-to-ground active object tracking method based on reinforcement learning for high-altitude environments, which consists of a state recognition model and a reinforcement learning module. The state recognition model leverages the correlation between observed states and image quality (as measured by object recognition probability) as prior knowledge to guide the training of reinforcement learning. Then, the reinforcement learning module can actively control the PTZ camera to achieve stable tracking and successfully recover tracking after object loss. Additionally, the study introduces a UE-free simulator that increases the efficiency of the training process by over nine times. High-altitude experimental results with the proposed method show significantly enhanced stability and robustness compared to the PID method. Furthermore, the results also indicate that the proposed method can significantly improve the image quality of the observation.

**Keywords:** Active object tracking · Reinforcement learning

## 1 Introduction

Object tracking is a subject of interest for decades, with a focus on robotics and unmanned aerial vehicles (UAV) applications [15]. The purpose of object tracking is to locate the position of a moving target in a sequence of image frames and keep it in the image field of view. Object tracking can be categorized into active object tracking and passive object tracking. The passive method [10] assumes that the camera is stationary and the object is within the camera's field of view. In contrast, the active method requires adjusting the posture of pan-tilt-zoom (PTZ) cameras to maintain continuous visibility within the image, which is more practical and challenging.

Active object tracking has been extensively researched across various fields [4]. However, most studies have focused on ground or low-altitude scenarios, while the research on active object tracking in high-altitude environments is

limited. In addition, previous methods have taken less account of the interference caused by the external environment on the controller and the resulting problem of object disappearance and recovery. However, in high-altitude environments, atmospheric disturbance and complex ground environments can interfere with the tracking process and easily lead to object loss.

To address the problem of active object tracking in a complex high-altitude environment, we propose an air-to-ground (i.e., high-altitude UAV tracks a vehicle on the ground) active object tracking method based on reinforcement learning. Specifically, we build a state recognition model and introduce it to the reinforcement learning training process as prior knowledge, which considers the correlation between the observation states and the quality of the observed images. Then, we design a reinforcement learning module to achieve stable tracking by actively controlling and guiding the PTZ camera, and successfully recovering tracking after object loss.

The results have shown that our proposed active object tracking method is significantly more stable than the PID control method in all proposed scenarios. In particular, in the case of object rectilinear motion, the tracking stability of our model is better than that of the S-curve and random motions. One possible reason is that our method could predict the object’s motion based on its historical information and control the camera accordingly in time, especially for rectilinear motion. Regarding the disturbance, PTZ camera vibration has a certain effect on tracking stability, while the object-specific disturbance (i.e., the object’s speed and direction) has less impact. Such a result indicates that it is important to maintain the stability of the PTZ camera during tracking and that our method has good adaptability regardless of the object’s movement.

Moreover, our method can significantly improve the accuracy of object recognition by automatically adjusting the magnification, i.e., the image quality of the observation is improved. Furthermore, we find that the object has a lower chance of reappearing in the field of view after being lost, even for a short period of time, while our method can quickly retrieve the object and resume tracking. In contrast, the conventional PID control method can only recover tracking in the few cases where the object is still in the field of view after being lost. Therefore, our method is significantly better than the PID control method in terms of robustness during tracking.

In summary, our main contributions include the following:

- An air-to-ground active object tracking reinforcement learning method is proposed to achieve tracking in high-altitude environments, and its tracking stability and robustness outperform the traditional PID method in all the proposed scenarios, i.e., different object motions and disturbance modes.
- A novel reward function involving the state recognition model is proposed and the object recognition probability is utilized as a part of the reward function to effectively guide reinforcement learning to adjust the PTZ camera focus and improve the input image quality.
- A memory-enabled actor-critic neural network is designed specifically for active object tracking, and the training efficiency is significantly improved,

i.e., over nine times compared to the UE simulator, by introducing a UE-free simulator and optimizing the training strategy.

## 2 Related Work

The goal of active object tracking is to lock the object by autonomously adjusting the position and attitude of the camera given a visual observation as input [12]. The method has been applied to a range of platforms, including PTZ cameras [13], vehicles [3], and UAVs [12]. For instance, Kyrkou [5] proposed a real-time and lightweight C<sup>3</sup>Net for roadside monitoring. Zhang et al. [16] implemented an end-to-end tracking method for UAVs by introducing GRU into the reinforcement learning network. However, these methods are not suitable for tracking tasks in high-altitude environments due to the relatively close distance between the tracker and the object.

Researchers have studied disturbance factors, including similar objects [12], occlusion [2], and obstacles [6] to increase the robustness of tracking methods. Such as Zhong et al. [18] and Yao et al. [14] improved the robustness of models by introducing occlusion during training. However, they do not consider the disturbance of vibration-induced tracker or target loss during active object tracking.

Combining prior knowledge with reinforcement learning can improve tracking performance, for example, a common approach involves combining PID as a knowledge module [7, 17]. However, the application of the PID is restricted due to problems such as vibration, object loss, and low image quality in high-altitude object tracking scenarios. Thus, we propose an active object tracking method that is well-suited to high-altitude environments and solves such problems.

## 3 Approach

### 3.1 Overview

Active object tracking aims to control the motion of a PTZ camera based on the object’s position in the image. In this paper, we propose an air-to-ground active object tracking reinforcement learning method, as shown in Fig. 1. The method includes two main components: the state recognition model and the reinforcement learning module with the improved Proximal Policy Optimization (PPO) [11] algorithm. A brief description of each of these modules is given below.

In some scenarios, it is necessary to maintain the highest possible image quality during the tracking process to obtain additional information from the observed images, which can further improve tracking performance. For this purpose, we introduce a state recognition model as prior knowledge into reward shaping, which is proposed to establish a relationship between image quality (as measured by object recognition probability) and the observed camera states through supervised learning. The model can guide the motion of PTZ cameras to further improve the quality of observed images and enhance the object tracking performance. Meanwhile, it can also avoid the vast computational burden caused by direct image processing with the reinforcement learning method.

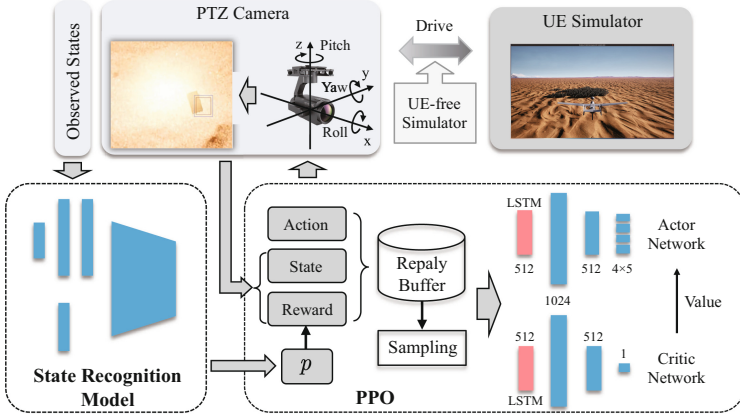


Fig. 1. The overall framework

A PPO-based reinforcement learning module is conducted to control the PTZ camera and, thus, address the high-altitude active object tracking task. To improve the efficiency of reinforcement learning training, we introduce a UE-free simulator. The trained model exhibits excellent tracking performance in the simulator, which is based on the UE engine to create a realistic desert environment with mild undulating terrain, forests, vehicle and UAVs. The movements of UAVs and vehicle follows the laws of physics.

### 3.2 State Recognition Model

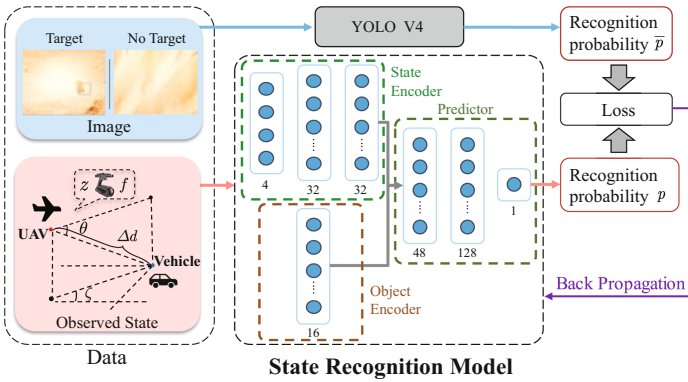


Fig. 2. State recognition model. (together with the observed states)

Figure 2 illustrates the structure of the state recognition model. The model takes the observed states as the input and the object recognition probability measuring the image quality as the output, and establishes the relationship between the

input and output through supervised learning. The observed states consist of five parameters:  $[\Delta d_t, \zeta_t, \theta_t, z_t, f_t]$ , where  $\Delta d_t$  represents the distance between the UAV and the vehicle,  $\zeta_t$  illustrates the azimuth angle of the vehicle relative to the UAV,  $\theta_t$  represents the pitch angle of the PTZ camera,  $z_t$  indicates the camera magnification (the camera is autofocus). In addition,  $f_t$  functions as a status flag to distinguish whether the object is present in the image, assigning a value of 1 if the object exists and 0 otherwise.

As shown in Fig. 2, the network structure of the state recognition model is comprised of three parts: a state encoder, an object encoder and a predictor. The state encoder processes the observed states through three fully connected layers and returns a 32-dimensional vector. The object encoder processes the object status flag in a fully connected layer and outputs a 16-dimensional vector. The predictor joins the two vectors outputted by the state and object encoders, respectively, and feeds them into three fully connected layers to generate the object recognition probability. Moreover, both encoders employ the ReLU activation function, while the predictor applies the softmax activation function. Furthermore, the number of neurons in each network layer is present in Fig. 2.

**Training Process.** We collect 24,000 images with the corresponding observed states from the UE simulator, and train the state recognition model using supervised learning based on pre-trained YOLOv4. Specifically, we take the images as the input and use the object recognition probability  $\bar{p}$  generated by YOLOv4 as the supervision signal, combined with the object recognition probability  $p$  generated by the state recognition model, to form the loss function<sup>1</sup>.

To improve the stability and convergence speed of the learning process, a gradient clipping approach with a threshold of 0.5 was conducted to dynamically adjust the learning rate<sup>2</sup>. The neural network parameters of the state recognition model were optimized using the Adam optimizer during the training process, and the state recognition model showed convergence after 30 iterations.

### 3.3 Reinforcement Learning Module

Active object tracking keeps the object within the field of view by continuously controlling the motion of the PTZ cameras. Such a process can be formulated as a classic reinforcement learning problem, and an improved proximal policy optimization (PPO) reinforcement learning algorithm is employed as an agent. The parameterization of the Markov Decision Process (i.e., state space, action space, and reward shaping) and the network architecture are described below.

The state space  $s_t$  at the moment  $t$  can be defined as given below:

$$s_t = \left[ \frac{x_t}{w}, \frac{y_t}{h}, \frac{z_t}{z_{\max}}, \frac{px_t}{(w \cdot h)}, f_t \right]^T$$

<sup>1</sup>  $Loss = \frac{\sum_{i=1}^N |p - \bar{p}|}{N}$ , where  $N$  represents the batch size of training.

<sup>2</sup> The rule for updating the learning rate:  $lr_{epoch} = \frac{1}{1 + 0.02 \times epoch}$ , where  $epoch$  represents the number of iterations.

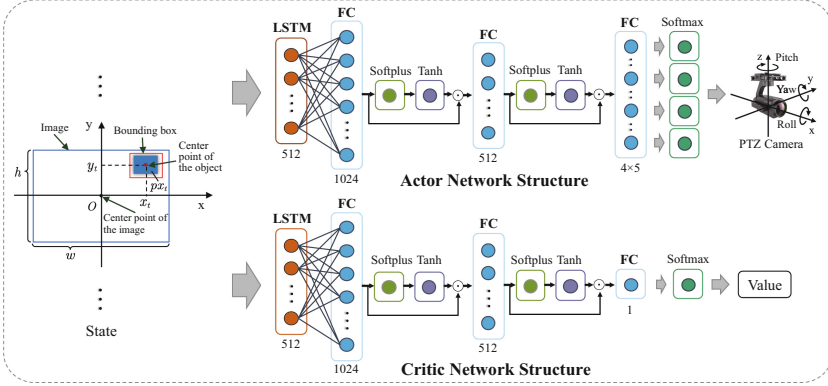


Fig. 3. Reinforcement learning module

where  $x_t$  and  $y_t$  donate the coordinates of the object concerning the center of the field of view,  $w$  and  $h$  represent the width and the height of the image, respectively (as shown in Fig. 3). In addition,  $px_t$  represents the pixel area of the object,  $z_t$  and  $z_{max}$  represent magnification and maximum magnification. In particular, when the object is lost at the time  $t$ ,  $s_t$  is set to  $\left[0, 0, \frac{z_t}{z_{max}}, 0, 0\right]^T$ . In this paper, the values of  $w$ ,  $h$ ,  $z_{max}$  are 1024, 768 and 400x, respectively.

At the time  $t$ , we define the  $action_t$  as  $[pitch_t, yaw_t, roll_t, zoom_t]$ , where  $pitch_t$ ,  $yaw_t$ ,  $roll_t$ , and  $zoom_t$  are integers with values ranging from  $-2$  to  $2$ , and represent the actions of the PTZ camera's pitch angle  $\theta$ , yaw angle  $\psi$ , roll angle  $\phi$ , and camera magnification  $z$ , respectively. When executing the  $action_t$ , the camera state at the time  $t-1$ , i.e.,  $[\theta_{t-1}, \psi_{t-1}, \phi_{t-1}, z_{t-1}]$ , is added with increments  $\alpha \cdot [pitch_t/z_{t-1}, yaw_t/z_{t-1}, roll_t/z_{t-1}, zoom_t \cdot \beta]$  to obtain the required camera state at the time  $t$ , and adjust the PTZ camera. The  $\alpha$  and  $\beta$  are coefficients.

In the proposed active object tracking process, the shaping of the reward function involves: 1) the agent should improve the object recognition probability  $p$  by actively performing actions, and 2) the center of the object should be as close to the center of the image as possible to achieve continuous and stable object tracking. In particular, the agent gets a time penalty when the object is out of the image. Therefore, the reward function is obtained by:

$$r_t = \begin{cases} mp_t - n\sqrt{\left(\frac{x_t}{w}\right)^2 + \left(\frac{y_t}{h}\right)^2}, & f_t = 1 \\ -n, & f_t = 0, \end{cases}$$

where  $m$ ,  $n$  are the coefficients used to limit the total accumulated reward value.

The actor and critic in PPO are represented as neural networks, with the structures shown in Fig. 3. In particular, the output layer of the actor network has four parallel fully-connected networks with five neurons each, corresponding to the four actions passed through a softmax activation function. The critic network has a similar structure to the actor network, except that it has only one

neuron in the output layer that directly outputs the Q value. A Block composed of softplus and tanh activation functions is used after the second and third layers, which makes the networks easier to optimize.

**Training Process.** Reinforcement learning requires iterative optimization through continuous interaction with mass data in the environment. However, the simulation environment built with UE engines usually runs slowly, hindering fast model training. Thus, we propose and construct a UE-free simulator based on environment abstraction, which can provide crucial parameters involved in the simulation environment and significantly improve training efficiency.

We propose an improved training procedure for the PPO algorithm (pseudocode below) to address the challenge of recovering the tracking process when the object has been lost for a long time. The core idea is to accumulate the rewards  $r_t$  for different moments satisfying certain criteria into a variable  $r_{sum}$  after each interaction between the agent and the environment, i.e., UE-free simulator. Then, the current episode is terminated when the value of  $r_{sum}$  falls below a predefined threshold, and a new round of training is started.

---

**Algorithm 1.** Improved PPO Algorithm.

---

**Require:** Initialize parameters  $\theta$  of actor network and  $\phi$  of critic network, respectively.

```

for episode = 1,2,... do
  Initialize set to store trajectory  $D \leftarrow \emptyset$ ;
  Initialize flag to determine whether to end the episode  $r_{sum} \leftarrow 0$ ;
  Randomly initialize state  $s_0$ ;
  for step = 1,2,... do
    Get the state  $s_t$  from the UE-free simulator;
    Run old policy  $\pi_{\theta'}$  to select action  $action_t$ ;
    Execute action  $action_t$ , receive  $r_t$  and obtained status  $s_{t+1}$ ;
    if  $r_t > 0$  then
       $r_{sum} \leftarrow \min([0.02, r_t]) + r_{sum}$ ;
    else
       $r_{sum} \leftarrow \max([-0.01, r_t]) + r_{sum}$ ;
    end if
    if  $r_{sum} < -2.0$  then
      break;
    end if
    Collect trajectory  $D \leftarrow D \cup (s_t, action_t, r_t)$ ;
    if the data in  $D$  reaches mini-batch and step % 50 == 0 then
      for Iteration = 1,2 do
        Sample a random mini-batch trajectory from buffer  $D$ ;
        Compute advantage estimates  $\hat{A}_t$  with GAE;
        Update  $\theta$  and  $\phi$  with clipped surrogate objective and value functions loss;
      end for
    end if
  end for
end for

```

---

In the training process, we initialize the actor and critic parameters, interact with the UE-free simulator to obtain the training data and store it in a replay buffer. Then, we sample a mini-batch of 256 from the replay buffer as our dataset and use Adam (the learning rate of actor and critic is  $1e-4$  and  $2e-4$ ) to optimize the network. The generalized advantage estimation (GAE) parameter  $\lambda$  and clipping  $\epsilon$  are set at 0.95 and 0.2. The maximum global episodes and maximum steps  $N$  are 8K and 400. For the reward function, the values of the  $\gamma$ ,  $m$  and  $n$  are 0.99, 0.1 and 0.1, and the action coefficient  $\alpha$ ,  $\beta$  are 50 and 5, respectively. We use TensorFlow as a deep learning framework to train the actor and critic networks with a PC containing an AMD Ryzen 7-5800H (3.20 GHz  $\times$  16) processor, 16 GB of RAM, and an NVIDIA RTX 3050 with 4 GB of VRAM.

## 4 Experimental Setup and Results Analysis

### 4.1 Baseline and Evaluation Criteria

As a baseline, we choose the commonly used and effective PID controller to control the pitch and yaw actions of the PTZ camera. The PID formula under discrete control is:

$$u_k = K_p \cdot e_k + K_i \sum_{j=0}^k e_j + K_d (e_k - e_{k-1}),$$

where at time  $k$ ,  $u_k$  represents the increment of pitch and yaw actions, and  $e_k$  represents the Euclidean distance from the image center to the object centroid. Moreover,  $K_p$ ,  $K_i$  and  $K_d$  are the tuning hyperparameters, which can be set to  $-0.005$ ,  $0.003$  and  $0.003$ , respectively, after multiple experiments. Furthermore, the magnification in the PID controller is set to 50 times.

The performance evaluation includes the following three criteria:

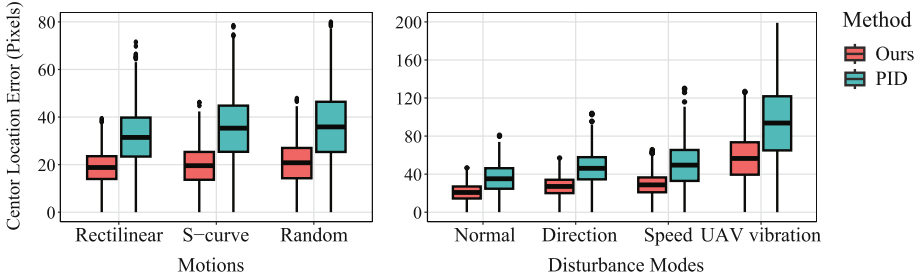
- **Stability.** The stability of the tracking process is measured in terms of center location error, which represents the Euclidean distance (in pixels) between the object centroid and the image center in a step. Continues smaller values of center location errors indicate better stability.
- **Robustness.** Ro is used to evaluating the robustness of the active tracker, which is the percentage of frames in which the tracker loses the object during the tracking process. Smaller Ro means better robustness.
- **Image quality.** Object recognition probability is adopted to measure image quality. Higher probability indicates better image quality is obtained during the object tracking process.

### 4.2 Experiments and Results

To conduct the experiments, we randomly initialized the starting position and orientation of the vehicle and the UAV, with the vehicle moving at 12 m/s and the UAV flying at 300 m altitude. Initially, the camera was set at a magnification of 50x and precisely aimed at the vehicle.



**Stability.** We compared the object tracking stability for three different vehicle motions, i.e., rectilinear, S-curve and random. In addition, we introduced three disturbance modes, i.e., the vehicle speed and direction, and the PTZ camera vibration, to further verify the tracking stability. Regarding the vehicle, the speed changes randomly in the range of 0 to 20m/s and the direction turns arbitrarily. For the PTZ camera, we applied a slight vibration by setting random changes in the pitch and roll angles, which caused the object to vibrate within the camera viewfinder frame. Moreover, we also compared the results obtained in each scenario with the tracking performance of the PID control method.



**Fig. 4.** Comparison of the tracking stability obtained in different scenarios

Figure 4 shows a series of box plots depicting the distribution of center location errors of our proposed and PID control methods in 30 episodes (i.e., 12000 steps) for each of the above-mentioned scenarios. According to Fig. 4, the center location errors of our method are almost two times smaller than those of the PID control method in all scenarios. To further evaluate the significance of the difference for each scenario, we calculated the Wilcoxon Signed Rank test [9] and Cliffs Delta Effect Size [1] on the center location errors of our method and PID control method. The results of the statistical tests reveal that the difference between the center location errors of our method and the PID control method in each scenario is significant (i.e.,  $p$ -value  $< 0.05$ ), and with a large effect size<sup>3</sup>. Thus, **the proposed method is significantly more stable than the PID control method in object tracking for all proposed scenarios.**

To evaluate the significance of different scenarios on the stability of our proposed tracking method, we performed a Scott-Knott effect size difference (ESD) test to group the different scenarios into statistically distinct ranks based on their center location errors. Tables 1 and 2 illustrate the ranks of tracking stability for three object motions and four disturbance modes, respectively. From Table 1, we found that the three object motions are distributed in two distinct groups, and in particular, the center location errors of S-curve and random motions (group #1) are relatively higher than rectilinear motion. Thus, **in the case of object**

<sup>3</sup> The magnitude is assessed using the thresholds provided by Romano et al. [8],  $|\text{delta}| > 0.474$  is large.

**Table 1.** Ranks of object motions according to the Scott-Knott ESD tests

Group	Object Motions
1	S-curve, random
2	rectilinear

**Table 2.** Ranks of disturbance modes according to the Scott-Knott ESD tests

Group	Disturbance Modes
1	PTZ camera vibration
2	object speed and direction
3	normal

**rectilinear motion, the tracking stability of our model is better than that of the S-curve and random motions.**

Similarly, Table 2 shows that the four disturbance modes are distributed in three distinct groups, and the center location errors obtained in the PTZ camera vibration scenario (group #1) are considerably higher than the other three. In addition, the center location errors are also significantly higher when adding disturbance to object speed and direction than in the normal situation. The results indicate that **the disturbance of the PTZ camera has the greatest impact on the object tracking stability, followed by the object-specific disturbance.**

**Image Quality.** The object recognition probabilities of 210 episodes obtained from the above seven scenarios using our method and the PID control method at the initial PTZ camera magnification (50x), respectively, all have an average of about 0.123. In other words, both methods have an accuracy of only about 0.123 for object recognition at the initial moment. During the tracking process, our proposed method improves the object recognition probability close to 1 in a short time (about 40 steps) and continues until the end of the tracking task by controlling the PTZ camera magnification (zoom in to approximately 400x). However, the PID control method has almost no improvement in the object recognition probability during the tracking process and ends with an average accuracy of 0.124. Therefore, **our method can significantly improve the accuracy of object recognition by automatically adjusting the magnification, i.e., the image quality of the observation is improved.**

**Robustness.** We evaluate the robustness of the proposed method by comparing the tracking performance after losing the object due to various reasons (e.g., interference or occlusion) during the UAV flight with the PID control method. We set the object vehicle to move randomly within an episode (400 steps) and to be lost (i.e., no longer receiving the tracking signals) at step 100. After the loss, the movements of the object vehicle and the UAV remain constant, and an attempt is made to re-observe and re-track the object at step 120. Since the object vehicle and the UAV continue to remain in motion, there are two situations when the object is re-observed at step 120 through the PTZ camera, 1) the object is still in the field of view (*inFoV*) and 2) the object is no longer observed, i.e., out of the field of view (*outFoV*).

Table 3 shows the number of occurrences (numbers in parentheses) for *inFoV* and *outFoV* after 50 experiments using our method and the PID control method,

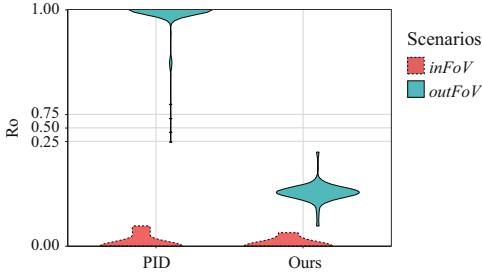


Fig. 5. Distribution of Ro

Table 3. Ro values in different scenarios

Scenarios	PID	Ours
<i>inFoV</i>	0.20% (9)	0.12% (6)
<i>outFoV</i>	95.69% (41)	2.84% (44)
<b>Total</b>	<b>78.50%</b>	<b>2.51%</b>

respectively, and the corresponding average value of Ro, i.e., the percentage of frames in which the tracker lost the object after step 120. In addition, the “Total” row represents the average Ro values in a total of 50 experiments. From Table 3, *outFoV* has a much higher probability of occurrence than *inFoV*, which means **the object has a lower chance of reappearing in the field of view after being lost for a short period of time** (e.g., only 20 steps).

Moreover, Table 3 also reveals the lower Ro values for our method, especially in the *outFoV* situation. This can be further visualized in Fig. 5, where the violin plots comparing the change in Ro values for *inFoV* and *outFoV* situations can be visualized for each method. The more elongated the shape of the violin, the larger the variance in the corresponding group; and the wider the violin plot, the higher the density. By observing Fig. 5, we note that the Ro values of our method in the *inFoV* situation are slightly lower than those of the PID control method. In stark contrast, the Ro values of our method in the *outFoV* situation are surprisingly smaller than those of the PID control method. These results highlight that **our method is significantly better than the PID control method in terms of robustness during tracking, especially when re-tracking after object loss.**

## 5 Conclusion

This paper proposes an air-to-ground active object tracking method based on reinforcement learning for a high-altitude tracking environment. The method consists of two parts: a state recognition model and an improved PPO algorithm. By incorporating the state recognition model’s output into the reward function, the proposed reinforcement learning algorithm adjusts the PTZ camera to improve image quality during tracking. Moreover, a UE-free simulator is introduced to accelerate the training process. The experimental results indicate that our proposed method offers a higher level of robustness and stability in the tracking process as compared to PID control method.

However, due to the limitations of experimental conditions, future work will focus on deploying our method in physical environments and solving more disturbance factors in high-altitude environments to further enhance its applicability.

## References

1. Cliff, N.: Ordinal Methods for Behavioral Data Analysis. Psychology Press, London (2014)
2. Cui, Y., Hou, B., Wu, Q., Ren, B., Wang, S., Jiao, L.: Remote sensing object tracking with deep reinforcement learning under occlusion. *IEEE Trans. Geoscience Remote Sens.* **60**, 1–13 (2021)
3. Devo, A., Dionigi, A., Costante, G.: Enhancing continuous control of mobile robots for end-to-end visual active tracking. *Robot. Auton. Syst.* **142**, 103799 (2021)
4. Jeong, H., Hassani, H., Morari, M., Lee, D.D., Pappas, G.J.: Deep reinforcement learning for active target tracking. In: 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 1825–1831. IEEE (2021)
5. Kyrkou, C.: C<sup>3</sup>Net: end-to-end deep learning for efficient real-time visual active camera control. *J. Real-Time Image Proc.* **18**, 1421–1433 (2021)
6. Luo, Y., et al.: Calibration-free monocular vision-based robot manipulations with occlusion awareness. *IEEE Access* **9**, 85265–85276 (2021)
7. Ma, X., Wang, Y., Yang, S., Niu, W., Ma, W.: Trajectory tracking of an underwater glider in current based on deep reinforcement learning. In: OCEANS 2021, San Diego-Porto, pp. 1–7. IEEE (2021)
8. Romano, J., Kromrey, J.D., Coraggio, J., Skowronek, J.: Appropriate statistics for ordinal level data: should we really be using t-test and Cohen’s d for evaluating group differences on the NSSE and other surveys. In: Annual Meeting of the Florida Association of Institutional Research, vol. 177, p. 34 (2006)
9. Rosner, B., Glynn, R.J., Lee, M.L.T.: The Wilcoxon signed rank test for paired comparisons of clustered data. *Biometrics* **62**(1), 185–192 (2006)
10. Ross, D.L., Lim, J.L., et al.: Incremental learning for robust visual tracking. *Int. J. Comput. Vis.* **77**(1r3), 125r141 (2008)
11. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms. arXiv preprint [arXiv:1707.06347](https://arxiv.org/abs/1707.06347) (2017)
12. Xi, M., Zhou, Y., Chen, Z., Zhou, W., Li, H.: Anti-distractor active object tracking in 3D environments. *IEEE Trans. Circuits Syst. Video Technol.* **32**(6), 3697–3707 (2021)
13. Yang, J., Tang, Z., Pei, Z., Song, X.: A novel motion-intelligence-based control algorithm for object tracking by controlling pan-tilt automatically. *Math. Probl. Eng.* (2019)
14. Yao, B.: GARAT: generative adversarial learning for robust and accurate tracking. *Neural Netw.* **148**, 206–218 (2022)
15. Yun, S., Choi, J., Yoo, Y., Yun, K., Choi, J.Y.: Action-driven visual object tracking with deep reinforcement learning. *IEEE Trans. Neural Netw. Learn. Syst.* **29**(6), 2239–2252 (2018)
16. Zhang, H., He, P., Zhang, M., Chen, D., Neretin, E., Li, B.: UAV target tracking method based on deep reinforcement learning. In: 2022 International Conference on Cyber-Physical Social Intelligence (ICCSI), pp. 274–277. IEEE (2022)
17. Zhao, W., Meng, Z., Wang, K., Zhang, J., Lu, S.: Hierarchical active tracking control for UAVs via deep reinforcement learning. *Appl. Sci.* **11**(22), 10595 (2021)
18. Zhong, F., Sun, P., Luo, W., Yan, T., Wang, Y.: Towards distraction-robust active visual tracking. In: International Conference on Machine Learning, pp. 12782–12792. PMLR (2021)