

Article

A Stealth–Distance Dynamic Weight Deep Q-Network Algorithm for Three-Dimensional Path Planning of Unmanned Aerial Helicopter

Zeyang Wang, Jun Huang and Mingxu Yi *

School of Aeronautic Science and Technology, Beihang University, Beijing 100191, China

* Correspondence: yimingxu@buaa.edu.cn; Tel.: +86-18201311103

Abstract: Unmanned aerial helicopters (UAHs) have been widely used recently for reconnaissance operations and other risky missions. Meanwhile, the threats to UAHs have been becoming more and more serious, mainly from radar and flights. It is essential for a UAH to select a safe flight path, as well as proper flying attitudes, to evade detection operations, and the stealth abilities of the UAH can be helpful for this. In this paper, a stealth–distance dynamic weight Deep Q-Network (SDDW-DQN) algorithm is proposed for path planning in a UAH. Additionally, the dynamic weight is applied in the reward function, which can reflect the priorities of target distance and stealth in different flight states. For the path-planning simulation, the dynamic model of UAHs and the guidance model of flight are put forward, and the stealth model of UAHs, including the radar cross-section (RCS) and the infrared radiation (IR) intensity of UAHs, is established. The simulation results show that the SDDW-DQN algorithm can be helpful in the evasion by UAHs of radar detection and flight operations, and the dynamic weight can contribute to better path-planning results.

Keywords: SDDW-DQN algorithm; unmanned aerial helicopter; dynamic weight; reward function; path planning



Citation: Wang, Z.; Huang, J.; Yi, M. A Stealth–Distance Dynamic Weight Deep Q-Network Algorithm for Three-Dimensional Path Planning of Unmanned Aerial Helicopter. *Aerospace* **2023**, *10*, 709. <https://doi.org/10.3390/aerospace10080709>

Academic Editor: Hailong Huang

Received: 3 June 2023

Revised: 7 August 2023

Accepted: 10 August 2023

Published: 15 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The development of artificial intelligence and related technologies makes it possible for unmanned aerial vehicles (UAVs) to realize autonomous flying [1–3]. With a smaller size and lighter weight, UAVs have a much lower operation cost and lower maintenance cost. The tiny size of UAVs can also offer a lower detectability and better maneuverability. Furthermore, accidents might happen to UAVs, but UAVs without pilots will not experience any casualties. Thus, UAVs are suitable for dangerous missions, such as reconnaissance, rescue, and operation missions. Compared with fixed-wing UAVs, unmanned aerial helicopters (UAHs) have the ability to hover [4–7]. Due to this, combined with their lower detectability, unmanned helicopters are more suitable for low-altitude stealth missions, such as close air support and specific target raids.

The modern three-dimensional airspace, from the multi-band identification of infrared flights to the high-frequency coverage of radar detection, has put forward great threats to UAHs carrying out missions near the ground. With the development of radar detection technology, it is more and more difficult for UAHs to avoid radar detection, and UAHs might face flight operations after exposure to radar. Without the ability of the high overload maneuver, it is almost impossible for UAHs to dodge incoming flights via an evasive maneuver. For the reasons above, stealth has become crucial to the survivability of UAHs, and is one of the effective methods for UAHs to avoid radar detection, as well as flight operations.

The characteristic signals of UAHs mainly include the radar cross-section (RCS) and infrared radiation (IR), and appear with different values at different azimuth angles, elevation angles, and distances between the UAH and the detector. In the same detection

environment, different positions and attitudes of UAHs lead to different detectabilities, and a UAH has the possibility of maintaining a low detectability by flying in a safe path, with safe attitudes. Therefore, the path planning of a UAH can be important to the survivability of the UAH, and it is essential that a path-planning algorithm is established.

In recent years, researchers have carried out plenty of work on path-planning problems in UAVs. Graph-theory-based methods and stochastic programming methods are two types of existing path-planning techniques. Bhattacharya P et al. [8] have proposed a roadmap approach, and utilized the Voronoi diagram to obtain a path that is a close approximation of the shortest path satisfying the required clearance value set by the user. Wenzheng Chi et al. [9] have presented a generalized-Voronoi-diagram (GVD)-based heuristic path-planning algorithm to generate a heuristic path, guide the sampling process of RRTs, and further improve the motion-planning efficiency of RRTs. To solve the problems in dynamic environments, Chanyoung Song et al. [10] have presented a data structure and algorithm to construct a Voronoi diagram composed of moving circular disks (dynamic Voronoi diagram of moving disks) in the plane.

Graph-theory-based methods and stochastic programming methods can make contributions to the global optimization of path planning, but it is rather difficult for them to give global-optimized path-planning results. Moreover, some of the path-planning algorithms for UAVs based on these methods can only lead UAVs to adapt to the specific static environment set in the simulation. In the autonomous flight of UAVs, intelligent optimization algorithms and machine learning (ML) are adapted to path-planning problems [11–13]. Haoran Zhu et al. [14] have provided an overview of different swarm intelligence (SI) algorithms for UCAV path-planning research, and twelve algorithms that are published in major journals and conference proceedings are surveyed and then applied to UCAV path-planning. However, these SI algorithms mainly solve the problems in two-dimensional path planning, and the x -coordinate of points on the paths must increase monotonically by a certain step size. Some of the optimized paths even travel through the center of one radar-detection area, which is not allowed in path planning. Nating Du et al. [15] have put forward an improved Chimp optimization algorithm (IChOA) based on a somersault foraging strategy with adaptive weight, aiming to solve the three-dimensional path-planning problem. The IChOA mainly focuses on avoiding collision with obstacles in a complex environment, but ignores threats such as radar detection and flight operation. Moreover, the velocity variations in a UAV have not been taken into consideration.

With the development of machine learning (ML) technology, ML has been introduced to path-planning problems by more and more researchers [16–20]. Deep learning (DL) and reinforcement learning (RL) are two typical types of ML. As for path-planning problems with limited state space, in which the number of states is countable, classic ML algorithms can complete path planning. However, the problem of state space explosion will emerge if the state space appears with too many states. Deep reinforcement learning (DRL), the combination of deep learning (DL) and reinforcement learning (RL), has the advantage of avoiding the explosion of state space. Therefore, DRL has been an effective tool in the design of path-planning algorithms. Dongcheng Li et al. [21] have put forward a global–local hybrid UAV path-planning algorithm based on the A* algorithm and Q-learning, with the search strategy, the step size, and the cost function improved, and the OPEN set simplified. The planning time has thereby been shortened, and the execution efficiency of the algorithm has been improved. But the path planning mainly concentrates on avoiding obstacles.

As one of the DRL algorithms, the Deep Q-Network (DQN) algorithm is a method to approximate the Q-learning function through a neural network. DQN methods have been increasingly applied in the field of path planning, and several brilliant algorithms based on it have been put forward [22–25]. Yin Cheng et al. [26] have developed a concise DRL obstacle-avoidance algorithm that designed a comprehensive reward function for behaviors such as obstacle avoidance, target approach, speed correction, and attitude correction in dynamic environments, using the deep Q-network (DQN) architecture, to overcome the usability issue caused by the complicated control law in the traditional analytic approach.

As for UAH path planning, Jiangyi Yao et al. [27] have explored a state-coded Deep Q-Network (SC-DQN) algorithm with symmetric properties, which can effectively avoid randomly moving obstacles, and plan a safe path for a UAH. Radar detection is also taken into consideration in the SC-DQN algorithm. Nevertheless, the formulas of radar detection probability are oversimplified, and the characteristic signals of a UAH (radar cross-section, infrared radiation, etc.) have not been taken into account. The path-planning problem also involves two-dimension paths only. Furthermore, the reward function of the SC-DQN algorithm is only affected by distance. Thus, the proposed DQN algorithm for UAH path planning still has room to improve.

To summarize, few of the existing works on UAH/UAV path planning take the stealth ability of UAHs/UAVs into consideration, and few of the works have introduced the threat of flight operation. According to the analysis above, the research purposes of this paper can be listed as follows:

1. The research aims to develop an ML algorithm of UAH/UAV path planning, with the stealth abilities of UAHs/UAVs considered in the reward function.
2. The research is put forward to take account of not only the threats of radar and IR detections, but also threats of flight operations, in the path-planning simulations of UAHs/UAVs. The related models should also be established.
3. The research is designed to balance evasive maneuvers, as well as the approaching maneuvers of UAHs/UAVs, while making maneuver decisions during path planning.

According to the research purposes above, a new algorithm for UAH path planning, based on the deep Q-network (DQN), is put forward, and a dynamic-weight reward function is designed. The main contributions of the research in this paper are as follows:

1. Considering the detection and operation threats to UAHs, a stealth–distance dynamic weight DQN (SDDW-DQN) algorithm is put forward for UAH path planning under the threat of radar detection and flight operation.
2. The stealth model of a UAH, as well as the guidance model of flight, are established in this paper, and the formulas of detection probability are illustrated in detail.
3. A new reward function of the SDDW-DQN algorithm is designed with dynamic weights, which can be influenced by both the distance to the destination, and the detection probability.

2. Models

2.1. Dynamical Model of UAH

For the simplification of the complexity of the UAH motivation, the roll angle and pitch angle are fixed at 0° . Therefore, the dynamic model of a UAH can be shown as Equation (1).

$$\begin{cases} a_x = \frac{F_x - \text{sgn}(v_x) \cdot C_x v_x^2}{m} \\ a_y = \frac{F_y - \text{sgn}(v_y) \cdot C_y v_y^2}{m} \\ a_z = \frac{F_z - \text{sgn}(v_z) \cdot C_z v_z^2}{m} - g \\ \beta_z = \frac{M_z - \text{sgn}(\omega_z) \cdot C_{\omega_z} \omega_z^2}{I_z} \end{cases} \quad (1)$$

where a_x, a_y, a_z represent the acceleration components on the x, y, z axis. v_x, v_y, v_z represent the velocity components on the x, y, z axis. F_x, F_y, F_z represent the force components on the x, y, z axis. The air resistance is considered to correspond to the square of the velocity, and $C_x, C_y, C_z, C_{\omega_z}$ represent the resistance coefficient. m represents the weight of the UAH, while I_z represents the rotational inertia at the z axis.

Figure 1 shows the schematic diagram of the dynamic model of the UAH, wherein the blue arrows represent thrust forces, the red arrows represent drag forces, and the green

arrow represents gravity. The xyz velocities, as well as the z rotate speed, are expressed by the light-blue thick arrows.

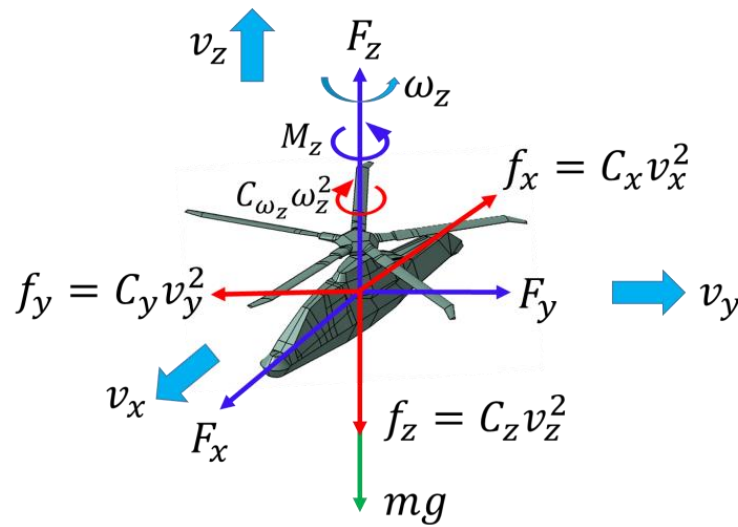


Figure 1. Schematic diagram of the dynamic model of the UAH.

2.2. Dynamic Model and Guidance Model of Flight

With a long operation range and a high operation accuracy, a flight can be one of the most serious threats to unmanned helicopters. Radar guidance and infrared (IR) guidance are two typical guidance modes of flight. Compound guidance mode is applied to advanced flights, with radar guidance and IR guidance combined. Compound guidance mode can improve the operation accuracy of a flight, even under electromagnetic interference.

As active detectors, radars can identify the position, speed, and moving direction of targets according to reflected radar waves. Therefore, proportional navigation (PN) can be applied to radar-guided flights. The guidance law of PN can be described as the normal acceleration (overload) of the flight being proportional to the rotational angular velocity of the line of sight (LOS), and the overloads in the yaw and pitch directions are shown in Equation (2).

$$\begin{cases} n_{mc} = K \cdot \frac{v_m \cos(\gamma_t)}{g} (\dot{\beta} + \tan \varepsilon \tan(\varepsilon + \beta) \dot{\varepsilon}) \\ n_{mh} = \frac{v_m}{g} \frac{K}{\cos(\varepsilon + \beta)} \dot{\varepsilon} \end{cases} \quad (2)$$

where β represents the yaw angle of the LOS, and ε represents the pitch angle of the LOS, as is shown in Figure 2. v_m represents the speed of the flight, g represents the gravitational acceleration, γ_t represents the included angle of the LOS and the flight’s velocity vector, and K represents the proportional guidance coefficient.

Different from radar detection, IR detection is acknowledged as passive detection, and only the orientation of the target can be obtained. Thus, the proper guidance law should be tracking guidance (TG). Equation (3) describes the guidance law of TG.

$$\begin{cases} \beta_m = \beta \\ \varepsilon_m = \varepsilon \end{cases} \quad (3)$$

where β_m and ε_m represent the yaw and pitch angle of the flight’s velocity vector. Thus, the guidance law of TG can be described as: the velocity direction of the flight should be parallel to the LOS.

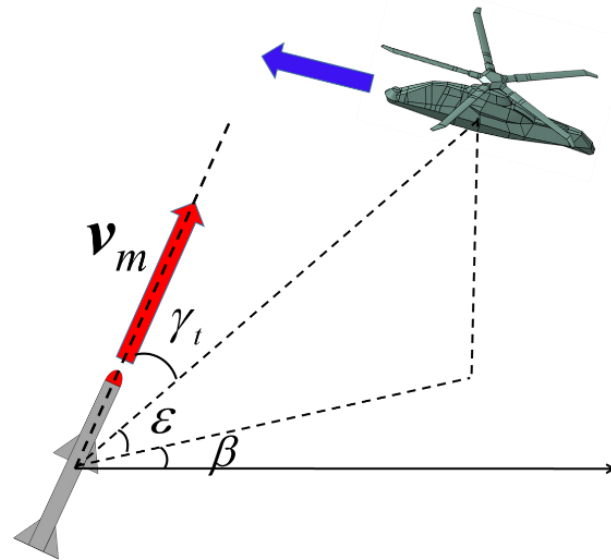


Figure 2. Schematic diagram of proportional navigation (PN).

The compound guidance modes of the flight in this paper include IR guidance and semi-active radar guidance. Figure 3 shows the whole progress of the flight guidance. The guidance progress can be summarized as the flow chart that can be seen in Figure 4, and can be described through the following steps.

- Step 1. Awaiting orders. . .
- Step 2. When the target is continuously detected three times by radar, the moving track of the target can be preliminary identified, and the flight will be launched.
- Step 3. The flight will be controlled by PN when the target is detected by radar. If not, the flight will search for the target itself using infrared detection, and will be controlled by TG when the target is detected via IR detection. If both radar and infrared fail to detect the target, the flight will keep flying in a straight line.
- Step 4. When the distance between the UAH and the flight is shorter than 20 m, the flight will hit the target UAH. Otherwise, the flight will fail to hit UAH, and will execute self-destruction when the UAH has reached the destination area or the flight crashes into the ground ($z \leq 0$).

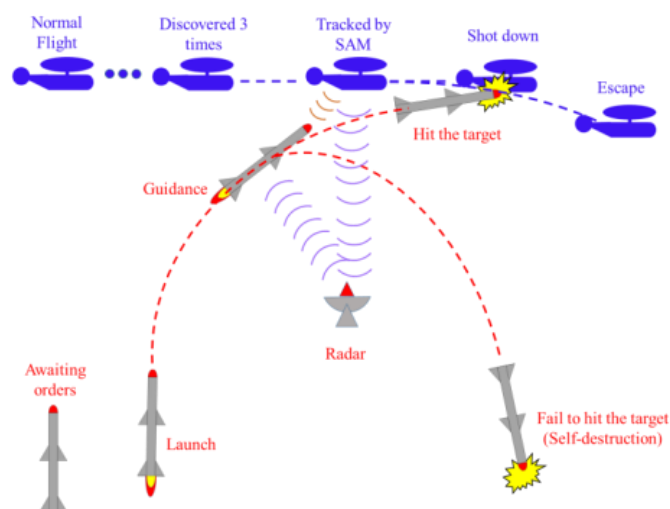


Figure 3. Progress of flight guidance.

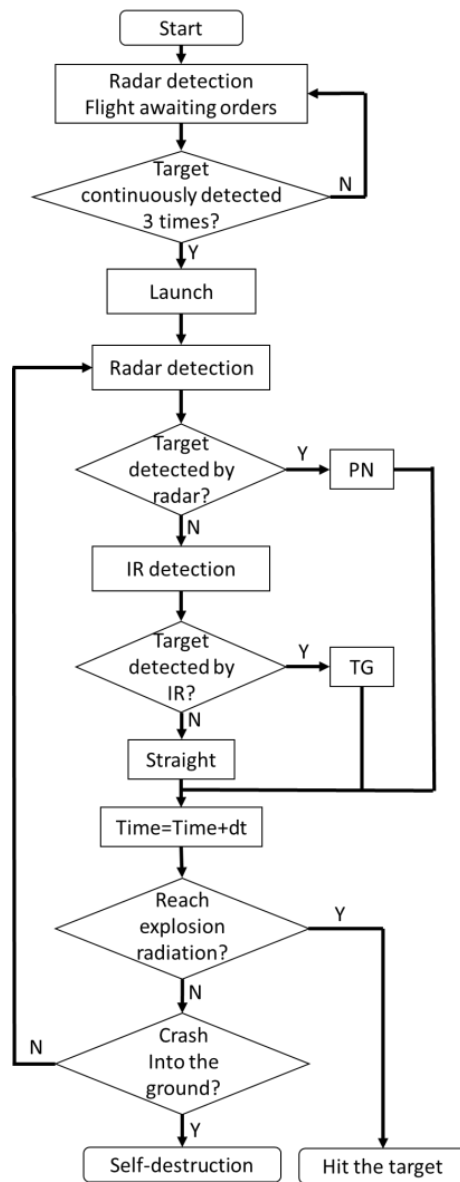


Figure 4. Flow chart of the flight guidance process.

2.3. Stealth and Detection Model

According to the guidance model of flight, the main characteristic signals of the UAH that allow the UAH to be tracked by the flight include the radar cross-section (RCS), as well as infrared radiation (IR). Therefore, the stealth model of a UAH consists of the RCS and IR of the UAH, and the detection model of the UAH consists of the detection probabilities by radars and infrared detectors.

2.3.1. RCS Model of a UAH

The RCS model of a UAH should be made up of the RCS values at all of the detection orientations, including the azimuth angle from 0° to 360° , and the elevation angle from -90° to 90° . Thus, the establishment of the RCS model of a UAH requires an effective method of RCS prediction. The PO + PTD method can be applied to the calculation of the RCS of a UAH. According to the finite element method (FEM), the geometry model of the target can be divided into triangular elements, as shown in Figure 5, and the PO + PTD method can be applied to the calculation of every triangular facet.

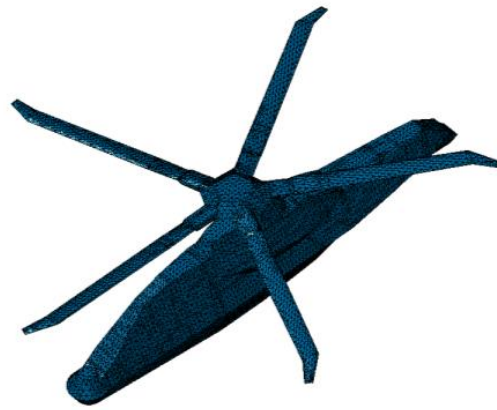


Figure 5. UAH geometry model divided into triangle elements.

The far-field electric field can be obtained by

$$E_S = \frac{e^{-jkr}}{4\pi r} (-j\omega\mu) \int_S J_S e^{jk\hat{s}\cdot r'} dS + jk\hat{s} \frac{1}{\epsilon} \int_S \rho_S e^{jk\hat{s}\cdot r'} dS \tag{4}$$

where E_S represents the scattered electric field, r represents the field point, and r' represents the coordinate vector of the source point. J_S represents the surface current, which can be calculated by

$$J_S = 2\hat{n} \times H_i \tag{5}$$

For the far field, the incoming wave can be considered as a plain wave; thus, the RCS of the target can be written as:

$$\sqrt{\sigma} = \sqrt{4\pi r^2} \frac{1}{|E_i|} \left(-j\omega\mu \frac{e^{-jkr}}{4\pi r} \int_S \hat{e}_s \cdot J_S e^{jk\hat{s}\cdot r'} dS \right) \tag{6}$$

Through substituting Equation (8) into Equation (10), it can be obtained that

$$\sqrt{\sigma} = \sqrt{4\pi r^2} \frac{1}{|E_i|} \left(-j\omega\mu \frac{e^{-jkr}}{4\pi r} \cdot 2 \int_S \hat{e}_s \cdot (\hat{n} \times H_i) e^{jk\hat{s}\cdot r'} dS \right) \tag{7}$$

Taking the relationship between the magnetic field and the electronic field, the calculation of the RCS can be transferred into:

$$\sqrt{\sigma} = \frac{jk}{\sqrt{\pi}} \hat{n} \cdot (\hat{e}_s \times \hat{h}_i) e^{jk\mathbf{w}\cdot\mathbf{r}_0} I' \tag{8}$$

$$\mathbf{w} = \mathbf{s} - \mathbf{i} \tag{9}$$

In Equation (12), the integral item I' can be written as:

$$I' = \begin{cases} \frac{e^{jk\mathbf{w}\cdot\mathbf{r}_m}}{jk|\mathbf{p}|^2} \sum_{m=1}^N \mathbf{p} \cdot L_m \text{sinc}\left(\frac{k\mathbf{w}\cdot L_m}{2}\right), & |\mathbf{p}| \neq 0 \\ A, & |\mathbf{p}| = 0 \end{cases} \tag{10}$$

where N represents the number of sides of every mesh element. $N = 3$ for triangle mesh elements, which is most commonly used in mesh generation.

In order to obtain more accurate RCS results, the edge-diffraction effects of the target should also be taken into consideration, and PTD is applied to the RCS calculation of the

target. More details of PTD can be seen in References [28,29], and the total RCS can be obtained by:

$$\sigma(t) = \left| \sum_{i=1}^{N_F(t)} \left(\sqrt{\sigma_F(t)} \right)_i + \sum_{j=1}^{N_E(t)} \left(\sqrt{\sigma_E(t)} \right)_j \right|^2 \tag{11}$$

where elements with a subscript *E* represents the edge elements, and elements with a subscript *F* represents the face elements.

The RCS distribution of the UAH at each of the orientations can be calculated using the PO+PTD method, and the RCS distributions at the *xy* plane, *yz* plane, and *zx* plane can be seen in Figure 6.

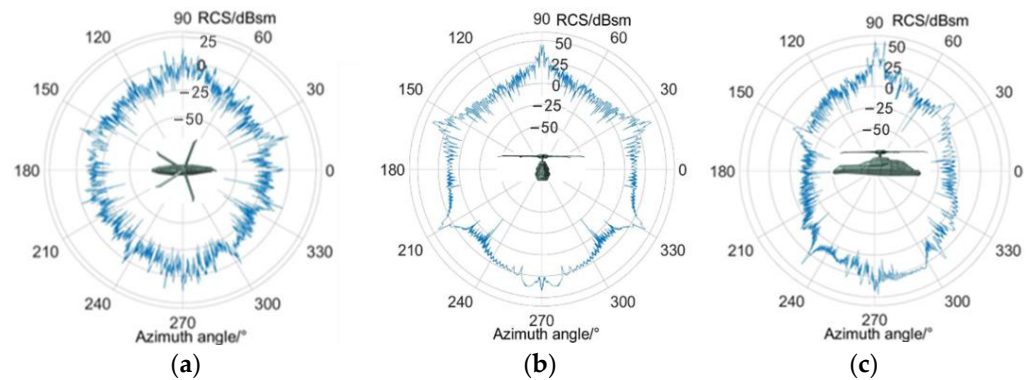


Figure 6. RCS distributions of the UAH at the (a) *xy* plain, (b) *yz* plain, and (c) *zx* plain.

2.3.2. Infrared Radiation (IR) Model of UAH

Similar to the RCS model of a UAH, the infrared radiation (IR) model of a UAH also consists of IR values at all of the detection orientations. An infrared radiation calculation method based on the Monte Carlo and ray-tracking method is presented in Ref. [29], and it can be applied upon the establishment of the IR model.

With the assumption that the reaction in the combustion chamber is complete, without the consideration of the scattering of the medium, the simplified radiation transfer equation (RTE) can be obtained as:

$$\frac{dL(\mathbf{r}, \mathbf{s})}{ds} + \alpha L(\mathbf{r}, \mathbf{s}) = \alpha \frac{\sigma T^4}{\pi} \tag{12}$$

where *r* represents the position vector, *s* represents the direction vector, *L* is the radiation intensity, $\sigma = 5.67032 \times 10^{-8} \text{ W}/(\text{K}^4 \cdot \text{m}^2)$ is the Stefan–Boltzmann constant, and *T* is the temperature.

The spectral radiance of the detector receiving point, according to the Monte Carlo and ray-tracing method, can be calculated by:

$$L_\sigma = L_\sigma^0 \prod_{i=1}^n \tau_{i\sigma} + \sum_{j=1}^{n-1} \left(L_{b\sigma}^j (1 - \tau_{j\sigma}) \prod_{k=j+1}^n \tau_{k\sigma} \right) + L_{b\sigma}^n (1 - \tau_{n\sigma}) \tag{13}$$

where L_σ^0 is the spectral radiance of the wall reverse rays, $L_{b\sigma}^j$ is the spectral radiance, and $\tau_{k\sigma}$ is the *k*-th layer spectral transmittance.

The irradiance at the position of the IR detector can be calculated by:

$$E = \sum_{i=1}^{NB} \sum_{j=1}^N L_{\sigma,(i,j)}^n \cdot \cos \theta_j \cdot \Omega_j \cdot \Delta \sigma_i \cdot 100 \tag{14}$$

The infrared radiation intensity can be written as:

$$I = E \cdot R^2 \tag{15}$$

It can be inferred that the infrared radiation intensity of a UAH can be calculated by the method mentioned above, if the temperature field of the UAH is measured. Considering that the simulation models of the temperature measurement of UAHs are not established in this paper, a decision has been made that the data of the infrared radiation intensity in Ref. [30] are applied to the IR model of the UAH in the path-planning simulation. Figure 7 shows the distribution of the infrared radiation intensity at the *xy* plain, and the IR intensity at the band of 3~5 μm is considered in the path-planning simulation.

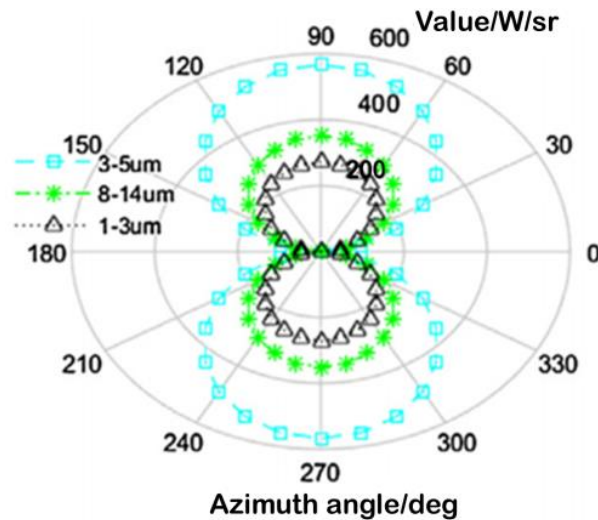


Figure 7. Distribution of the infrared radiation intensity of the UAH at the *xy* plain.

2.3.3. Detection Probabilities

(1) Radar detection probability

For the calculation of the radar detection probability, the signal-to-noise ratio (*SNR*) is indispensable. The signal-to-noise ratio (*SNR*) can be written as Equation (16):

$$\frac{SNR}{SNR_{\min(\hat{p})}} = \frac{\sigma}{\sigma_{cr(\hat{p})}} \tag{16}$$

In Equation (16), σ is the RCS of the UAH, and *SNR* represents the signal-to-noise ratio of the UAH. $SNR_{\min(\hat{p})}$ represents the noise intensity, and (*S/N*) should be the signal-to-noise ratio. $\sigma_{cr(\hat{p})}$ represents the critical radar cross-section of the radar, and $SNR_{\min(\hat{p})}$ represents the minimum *SNR* that can be detected via radar.

The calculation method for the radar detection probability can be obtained from Refs. [27,28], where the Gram–Charlier series, as well as the detection probability of Swerling-type targets, have been researched. The radar detection probability of a UAH can be calculated by:

$$P_{d,radar} = \frac{erfc(V/\sqrt{2})}{2} - \frac{\exp(-V^2/2)}{\sqrt{2\pi}} \left(C_3(V^2 - 1) + C_4V(3 - V^2) - C_6V(V^4 - 10V^2 + 15) \right) \tag{17}$$

where

$$C_3 = -\frac{SNR + 1/3}{\sqrt{n_p}(2SNR + 1)^{1.5}} \tag{18}$$

$$C_3 = \frac{SNR + 1/4}{n_p(2SNR + 1)^2} \quad (19)$$

$$C_6 = C_3^2/2 \quad (20)$$

The element V in Equation (17) can be calculated by:

$$V = \frac{V_T - n_p(1 + SNR)}{\omega} \quad (21)$$

where

$$\omega = \sqrt{n_p(2SNR + 1)} \quad (22)$$

(2) Infrared detection probability

Similar to the calculation of the radar detection probability, the requirement of the SNR of the infrared radiation intensity is also an indispensable part of the infrared detection probability calculation. Ref. [30] has put forward an approach to the calculation of the infrared detection probability.

For infrared detectors, the SNR can be obtained by:

$$SNR_{IR} = \frac{C^2(I_t - I_b)\tau_\alpha}{R^2} \quad (23)$$

where R represents the distance between the UAH and the IR detector. I_t is the infrared radiation intensity of the UAH, while I_b is the infrared radiation intensity of the background. τ_α is the atmospheric attenuation coefficient of infrared radiation, and C is a constant.

After passing the narrowband filter, the noise prominent signal is effectively filtered, and the amplitude probability density of the signal plus noise is expressed as:

$$p(R) = R \cdot \exp\left(-\frac{R^2 + A^2}{2}\right) \cdot I_0(RA) \quad (24)$$

where

$$\begin{cases} R = \rho/\sigma \\ A = a/\sigma \end{cases} \quad (25)$$

In Equation (25), ρ represents the amplitude value of the summation of the infrared noise and the infrared signal of the UAH, and a represents the amplitude of the infrared signal of the UAH.

The detection probability of an infrared detector can be obtained by:

$$\begin{aligned} P_{d,IR} &= \int_{V_0}^{\infty} R \exp\left(-\frac{R^2 + A^2}{2}\right) I_0(RA) dR \\ &= \exp\left(-\frac{V_0^2 + A^2}{2}\right) \cdot \sum_{n=0}^{\infty} \left(\frac{V_0}{A}\right)^n I_n(V_0A) \end{aligned} \quad (26)$$

where V_0 represents the threshold level of the IR detector, and I_n represents the Bessel function.

With the detection probabilities of both radar and infrared detectors, the total detection probability can be expressed as:

$$P_d = 1 - (1 - P_{d,radar})(1 - P_{d,IR}) \quad (27)$$

Considering that the detection probabilities of both radar and infrared detectors are the functions of the detection distance and detection orientation, the total detection probability can also be determined by the detection distance and detection in real time.

3. Design of SDDW-DQN Algorithm for Path Planning

In this section, the design of the SDDW-DQN algorithm for the path planning of a UAH is illustrated in detail, including the reward function, as well as the structure of the algorithm. First of all, a reward function with dynamic weights is designed, aiming to solve the path-planning problems of UAHs facing different threat environments. Then, the structure of the path-planning algorithm is established, and the new state elements and action elements are designed, with the consideration of stealth effects. Finally, the SDDW-DQN is developed.

3.1. Reward Function Design of SDDW-DQN

The reward function plays a crucial role in the reinforcement learning algorithm and, therefore, the design of a proper reward function can be important. Many of the existing works on path planning with reinforcement learning algorithms use reward functions with static weights or coefficients, and there are difficulties when facing dynamic threat environments. In order to solve the problems mentioned above, a new reward function with different weights under different situations is designed in this paper, and its expression can be illustrated by Equation (28).

$$r = - \left(w_1 \frac{d}{d_0} + w_2 P_d \right) \quad (28)$$

where d_0 represents the initial distance between the UAH and the target, d represents the distance between the UAH and the target in real time, and P_d represents the total detection probability of the flight by radar and IR detectors. w_1 represents the weight of the distance element (d/d_0), while w_2 represents the weight of the stealth element (P_d). The values of the two weights should satisfy:

$$w_1 + w_2 = 1 \quad (29)$$

Both w_1 and w_2 are dynamic weights, which vary based on changes in the threat environment. For the path-planning simulation, there are three situations of the threat environment, and the values of w_1 and w_2 are designed according to different situations.

Situation 1: The UAH is detected by radar, and the flight is waiting for orders.

Under Situation 1, the UAH should try to avoid radar detection, as well as approach the target and, thus, a greater detection probability and a farther distance might both bring punishment. According to the required condition of launching the flight, the value of the weights should vary from the times that the UAH has been detected. Therefore, the weights are designed as:

$$\begin{cases} w_1 = 1 - P_d^{\tan(\frac{2-n}{4}\pi)} \\ w_2 = P_d^{\tan(\frac{2-n}{4}\pi)} \end{cases}, n \leq 2 \quad (30)$$

where n represents the times that the UAH has been detected.

From Equation (30), it can be seen that the weight of the stealth element increases when the UAH is detected by radar once more and, meanwhile, the weight of the distance element decreases.

When $n = 0$, the UAH has not been detected, and the distance element should be given priority. Therefore, the values of the weights are $w_1 = 1$, $w_2 = 0$, meaning that the UAH is trying to approach the target, without worrying about detection.

When $n = 1$, the existence of the UAH is discovered via radar, and the threat emerges. Thus, the values of the weights are $w_1 = 1 - P_d$, $w_2 = P_d$, and the UAH starts to take detection into consideration. In addition, the weight of the stealth element (w_2) can be higher if the detection probability is higher, and the UAH has a higher tendency to avoid detection, to prevent itself from being continuously detected twice.

If $n = 2$, the UAH is in danger for the reason that the flight will launch and operation the UAH if it is continuously detected by radar once more. Therefore, the UAH must try its

best to maneuver away from radar at any cost, and the values of the weights are $w_1 = 0$, $w_2 = 1$, according to Equation (30). The stealth element has the highest priority.

Situation 2: the flight is launched, and the UAH is under tflight operation.

Situation 2 is dangerous, and the UAH must manage to evade the flight operation, otherwise the UAH will be shot down, and the mission will be a complete failure. Differently to fixed-wing aircrafts, helicopters have much lower speeds and much lower maneuver overloads, and it is hard, even impossible, for helicopters to avoid flight operations via evasive maneuvers. The UAH must take advantage of its stealth abilities and make it harder for radar and the flight to find the target, by suppressing the total detection probability. The weight values of the reward function can be designed as:

$$\begin{cases} w_1 = 1, w_2 = 0 (P_d < 0.2) \\ w_1 = \frac{5-10P_d}{3}, w_2 = \frac{10P_d-2}{3} (0.2 \leq P_d < 0.5) \\ w_1 = 0, w_2 = 1 (P_d \geq 0.5) \end{cases} \quad (31)$$

The design of the weight values under Situation 2, as expressed in Equation (31), can be explained as: the detection probability increases fast when it reaches the value of 0.5. On the other hand, it is rather hard for the detectors to find the UAH when the detection probability is under a threshold, which is set to be 0.2 in this paper. According to Equation (31), the stealth element occupies all the priority when the total detection probability reaches 0.5, and the UAH must suppress the fast increase in the total detection probability at any cost. On the contrary, however, the UAH should mainly consider approaching the target when the total detection probability is low enough. This is because the UAH has already been at the position far away from the radar and the flight, and the total detection probability is too low to detect and track the UAH. Then, the UAH might travel too far away from the target if it still considers suppressing the total detection probability, by maneuvering away even at a low total detection probability, and this can lead to a waste of time.

Generally speaking, the design of the weight values under Situation 2 can give consideration to both suppressing the total detection probability, and approaching the target at the proper moments, with proper dynamic weights.

Situation 3: the flight has failed to hit the UAH.

The fuel of the flight is limited, and its thrust cannot last for too long, and air friction might also be consuming the flight's mechanical energy, especially under high speeds and high overloads. Thus, the flight might fail to follow the track of the UAH if it has lost the target too many times, for the reason of having flown for too long. Under Situation 3, the UAH can enjoy a safe flight after the self-destruction of the flight, and can concentrate on approaching the target. Thus, the weight values of the reward function under Situation 3 should be:

$$\begin{cases} w_1 = 1 \\ w_2 = 0 \end{cases} \quad (32)$$

The path-planning simulation will end with one new episode if one of the following three events happens, and different values will be given to the reward function.

1. The UAH crashes into the ground. $r = -100$.
2. The UAH is shot down by the flight (when the distance between the UAH and the flight is 20 m or shorter). $r = -50$.
3. The UAH reaches the target within its working range (when the distance between the UAH and the target is 100 m or shorter). $r = 0$.

3.2. Structure and Elements of SDDW-DQN

The structure of the SDDW-DQN algorithm for UAH path planning is shown in Figure 8, with the applications of the designed reward functions, guidance laws of flight, and stealth model of the UAH.

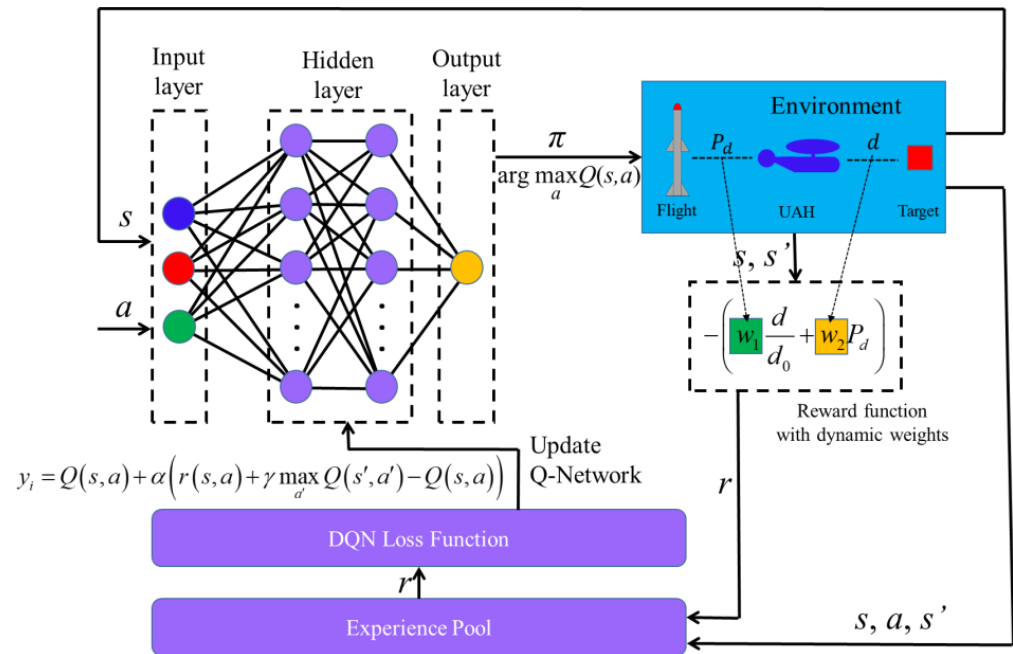


Figure 8. Structure of the SDDW-DQN algorithm for UAH path planning.

Figure 8 shows that the input layer of SDDW-DQN consists of two parts (state and action), but three points (blue, red, and green). This is because the state includes the state of the UAH (blue point), as well as state of the flight (red point). The green point represents the action of the UAH. The state elements of the UAH can be seen in Table 1, while the state elements of the flight can be seen in Table 2, and the action elements of the UAH are listed in Table 3.

Table 1. State elements of the UAH.

Symbol of State Elements	Meanings
x_h	X-coordinate of UAH
y_h	Y-coordinate of UAH
z_h	Z-coordinate of UAH
ϕ	Yaw angle of UAH
v_{xh}	X-velocity of UAH
v_{yh}	Y-velocity of UAH
v_{zh}	Z-velocity of UAH
$\dot{\phi}$	Yaw rate of UAH

Table 2. State elements of the flight.

Symbol of State Elements	Meanings
x_m	X-coordinate of flight
y_m	Y-coordinate of flight
z_m	Z-coordinate of flight
v_m	Speed of flight
β	Yaw angle of flight
ϵ	Pitch angle of flight
n_{sm}	State mark of flight
n_{con}	Numbers of continuous discoveries
n_{cur}	Current discovery state
t_f	Time of flight

Table 3. Action elements of the UAH.

Symbol of State Elements	Meanings
F_x	X-axis component of thrust
F_y	Y-axis component of thrust
F_z	Z-axis component of thrust
M_z	Z-axis moment

First of all, the basic parameters of the environment, UAH, and flight, and the initial state elements of the UAH, as well as the flight, are imported to the input layer. Next, the action under the current state will be selected according to the epsilon-greedy policy. The state of the UAH will be updated after the operation of the action and, meanwhile, the state of the flight will be updated based on the discovery of detection, as well as guidance laws. The reward will also be calculated. Then, the four elements (state, action, new state, reward) will be stored into the experience pool.

The learning starts when the loop of the path planning breaks, for the reason of either success or failure. The reward of all the points on the path is replayed according to the experience pool. The input of the DQN learner will be the current state s and the action a .

The current state s is defined as:

$$s = [s_h, s_m]^T \quad (33)$$

where

$$s_h = [x_h, y_h, z_h, \phi, v_{xh}, v_{yh}, v_{zh}, \dot{\phi}]^T \quad (34)$$

$$s_m = [x_m, y_m, z_m, v_m, \beta, \varepsilon, n_{sm}, n_{con}, n_{cur}, t_f]^T \quad (35)$$

The definitions of the state elements in Equations (33)–(35) can be seen in Tables 1–3. $Q(s, a)$, the output of DQN, is considered as the evaluate value. The target value is defined as:

$$y = Q(s, a) + \alpha \left(r(s, a) + \gamma \max_{a'} Q(s', a') - Q(s, a) \right) \quad (36)$$

Therefore $Q(s, a)$, $Q(s', a')$ and the reward value r are used to obtain the mean-square error of the DQN loss function, and the DQN network will be updated based on the stochastic gradient descent policy.

4. Simulation

4.1. Environments and Parameters

(1) Positions

The simulation environment is set as a plain field, where the expressions of the positions of elements in the environment are rectangular coordinates ($[x, y, z]$). The initial position of the UAH is set to be [0 m, 0 m, 500 m], while the position of the radar is initialized as [200 m, 100 m, 0 m]. The flight is considered to be coincident with the radar, and its initiated position is also [200 m, 100 m, 0 m]. The position of the destination is set to be [400 m, 300 m, 0 m].

(2) Attitudes and velocities

The initial velocity, as well as the initial rotation speed of the UAH, is fixed at 0, simulating the state of hovering at $t = 0$. The initial yaw angle of the UAH is set to be 0° . As for the flight, the pitch angle is initialized to be 90° , indicating that the flight stays vertically upward before being launched, in case it crashes into the ground the first second after it is launched.

(3) Parameters

The parameters of the simulation mainly include the performance parameters of the UAH and the flight. Tables 4 and 5 have listed the main parameters of the simulation. The density of the air is set as a constant value of 1.27 kg/m^3 , for the reason that the variety in the air density can be ignored at the relatively low heights at which the UAH travels.

Table 4. Performance parameters of the UAH.

Parameters	Symbols	Values
Weight of UAH	m	5000 kg
Z-rotational inertia	I_z	35,000 $\text{kg}\cdot\text{m}^2$
X-resistance coefficient	C_x	2.3
Y-resistance coefficient	C_y	2.3
Z-resistance coefficient	C_z	2.3
Z-rotation resistance coefficient	C_{ω_z}	2.3
X-thrust force	F_x	−15,000~15,000 N
Y-thrust force	F_y	−15,000~15,000 N
Z-thrust force	F_z	0~64,000 N (max)
Z-torque	M_z	−10,500~10,500 N·m
Working range	$r_{working}$	100 m

Table 5. Performance parameters of the flight.

Parameters	Symbols	Values
Weight of flight	m_{flight}	100 kg
Max lift coefficient	m_{flight}	0.3
Lift–drag ratio	$(L/D)_{flight}$	10
Zero-lift drag coefficient	$C_{D0,flight}$	0.05
Proportional navigation coefficient	K	3
Rocket thrust	F_t	12,000 N
Max thrust duration	$t_{ft,max}$	8 s
Operation radiation	r_{op}	20 m

4.2. Assumptions

For the simplification of the path-planning problem in this research, the assumptions of the simulation in this paper can be seen as follows:

- (1) It is assumed that the UAH is set to be the blue side, while the target, radar, and flight are set to be the red side. The symbols of points on the UAH paths are circles, and the symbols of points on the flight tracks are triangles.
- (2) For the simplification of the complexity of the UAH motivation, the roll angle and pitch angle are fixed at 0° . Only the yaw maneuvers of the UAH are taken into consideration.
- (3) The flight launcher has the same position as the radar, and both of them are static. During the simulation, the flight launcher only fires one flight to operation the UAH. If the flight has failed to hit UAH, and has operated self-destruction, the flight launcher appears as no threat to the UAH.

5. Results and Discussions

For the verification of the SDDW-DQN algorithm for UAH path planning developed in this paper, both the SDDW-DQN algorithm and ordinary DQN algorithms with different constant weights in reward functions are applied in the UAH path-planning simulation. Then, the comparisons of the path-planning simulation results using different DQN algorithms will be analyzed and discussed.

5.1. Converge Results of Different Cases

For the exploration of the effects of different weights in the reward function, several simulation cases using DQN algorithms with different weights have been considered in the path-planning simulation.

Case1: Path planning using DQN algorithms with static weights of $w_1 = 1, w_2 = 0$.

Case2: Path planning using DQN algorithms with static weights of $w_1 = 0.75, w_2 = 0.25$.

Case3: Path planning using DQN algorithms with static weights of $w_1 = 0.5, w_2 = 0.5$.

Case4: Path planning using DQN algorithms with static weights of $w_1 = 0.25, w_2 = 0.75$.

Case5: Path planning using DQN algorithms with static weights of $w_1 = 0.01, w_2 = 0.99$.

Case6: Path planning using DQN algorithms with dynamic weights.

Figure 9 shows the reward curves of the six cases. For a few stages of episodes at the beginning of the training, the reward values of all the cases appear at -100 , which illustrate the meaning of the initial random exploration period of UAH path planning. During the random exploration period, the UAH inevitably crashes into the ground, whatever the weights of the reward function. The reward curves from Case1 to Case4 start to converge at about the 1000~1500 training stages, but only converge to close to the value of -50 .

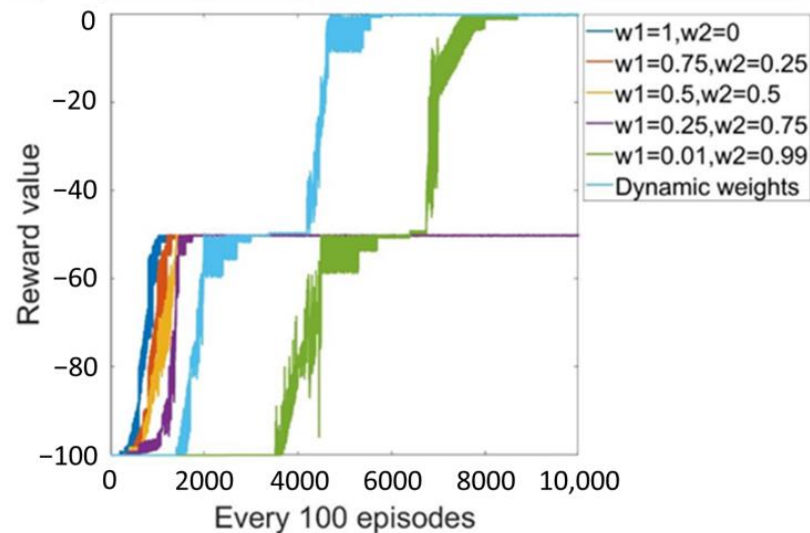


Figure 9. Reward curves of cases with different weights.

As for Case5 and Case6, both of the reward curves exist with two convergences. The first convergence of the reward curve of Case5 appears after approximately 6000 training stages, and the second comes out after about 8000 training stages, much later than Case1 to Case4. The two convergences of Case6 emerge earlier than those of Case5, after about 3000 training stages, and about 6000 training stages, but still much later than Case1 to Case4. However, the reward values of Case5 and Case6 can eventually converge to close to the value of 0 at their second convergences, but the reward values of Case1 and Case4 cannot.

Combined with the reward that UAH obtains under different consequences (crash into the ground, shot down by the flight, or operation the target successfully), it can be concluded from Figure 9 that the stealth element (detection probability) plays an important role in the reward function of the DQN path-planning algorithm. In Case1, Case2, Case3, and Case4, the UAH might be easily detected by radar, and shot down by the flight, if the weight of the factor of stealth is not high enough. On the contrary, the reward curve might appear to have a slow convergence if the weight of the factor of distance is rather low, for the reason that the paths of the UAH at the early exploration periods might cover large areas containing positions that are too far away from the target. This can lead to a waste of time.

5.2. Result Discussions of Different Cases

5.2.1. Result Discussions of Case1

Figure 10 shows the path of the UAH, as well as the track of flight in Case1. It can be observed from Figure 10 that the UAH travels almost straight towards the target, without evasive maneuvers, and it is easily shot down by the flight. The flight launches at the time spot of $t = 0.3$ s, and hits the UAH at the time spot of $t = 3.8$ s.

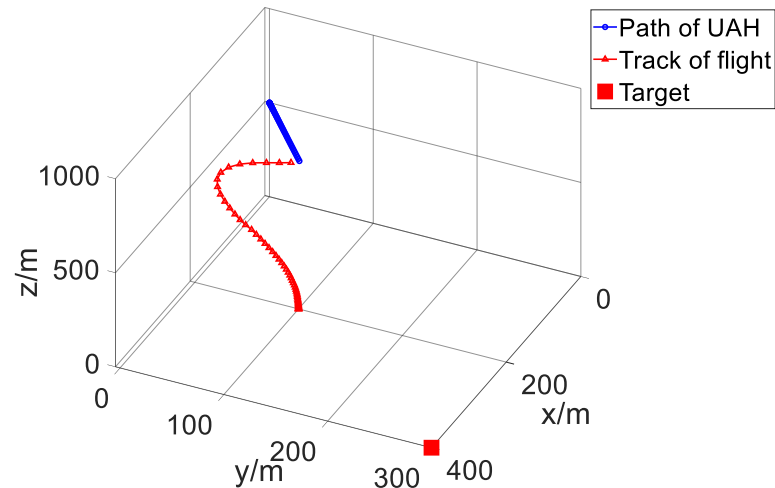


Figure 10. Path-planning result of the UAH, and the track of flight, in Case1.

The variations in the distance from the UAH to the target, and the distance from the flight to the UAH can be seen in Figure 11. It can be clearly illustrated that the distance from the flight to the UAH decreases much faster than the distance from the UAH to the target, and soon reaches the explosion distance of the flight, in less than 4 s. When the UAH is hit by the flight, the distance from the UAH to the target is 636.8 m, far beyond the working range of the UAH.

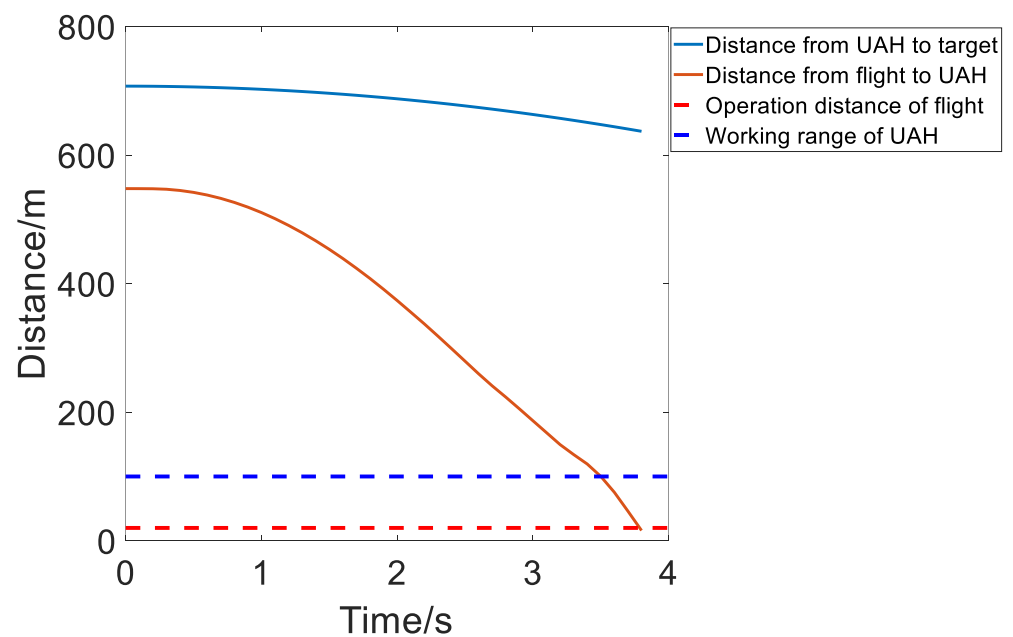


Figure 11. Distance from the UAH to the target, and distance from the flight to the UAH, in Case1.

The variation in the detection probability can be seen from Figure 12, and it is obvious that the detection probabilities rise above 50% at most of the time spots, and that the UAH can be easily tracked by radar and the flight. Thus, the path-planning result of the UAH

in Case1 can lead to a complete failure in the operation mission, without the effect of the stealth element in the reward function.

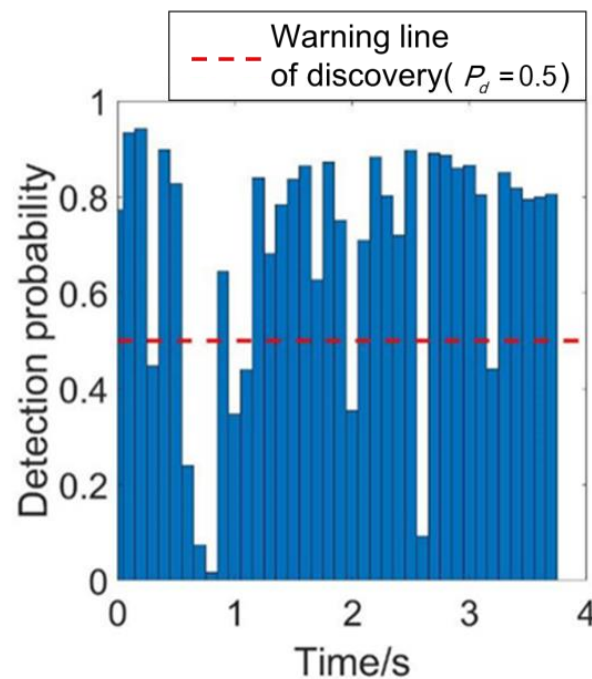


Figure 12. Variation in the detection probability in Case1.

5.2.2. Result Discussions of Case2

Figure 13 shows the path of the UAH, as well as the track of flight in Case2. Compared with Case1, the UAH appears to travel farther, and approaches the target more closely. The path of the UAH also seems not straight to the target, meaning that the UAH has operated several evasive maneuvers. Furthermore, the complexity of the track of flight increases, indicating that it might be more difficult for the flight in Case2 to track the UAH than that in Case1. However, the UAH still cannot avoid being shot down by the flight. The UAH begins to be tracked by the flight at the time of $t = 1.4$ s, and is finally shot down at the time of $t = 8.2$ s.

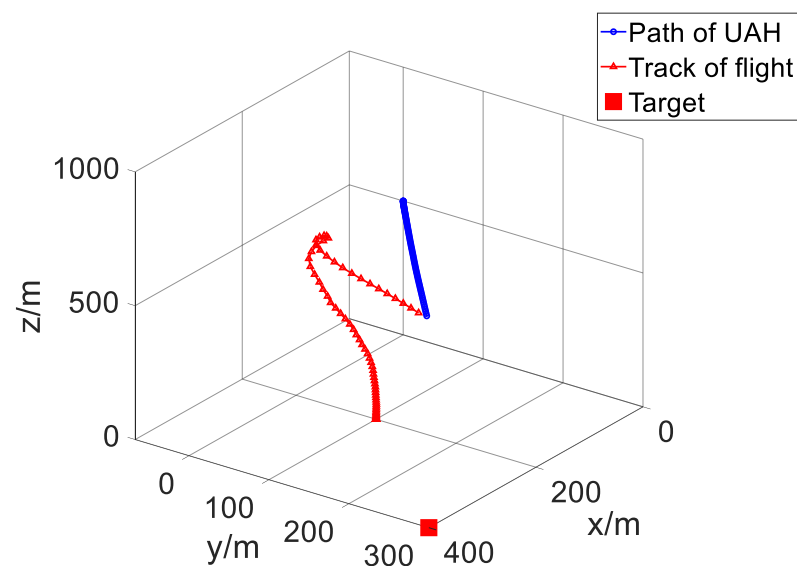


Figure 13. Path-planning result of the UAH, and the track of flight, in Case2.

Figure 14 shows the variations in the distance from the UAH to the target, and the distance from the flight to the UAH, and the variation in the detection probability can be seen from Figure 15. The distance from the flight to the UAH begins to decrease fast at the time spot of 1.4 s, almost the same time as when the flight launches. However, the distance from the flight to the UAH starts to increase a little at about $t = 5$ s, and varies randomly until the time spot of about 7 s. What is more, it can be seen from Figure 15 that the detection probability remains above 50% after the time spot of 5.4 s, but appears below 50% at most of the time spots before this time spot. Combined with the variation in the detection probability, it can be easily obtained that the flight usually fails to track the UAH before $t = 5.4$ s, and the decrease in the distance from the flight to the UAH cannot represent the successful tracking of the UAH. After $t = 5.4$ s, the UAH can be easily detected, and the flight finds itself at a wrong position and in a wrong direction. Then, the flight begins to correct its velocity, and starts to track the UAH again, with a high detection probability, and finally succeeds in shooting down the UAH.

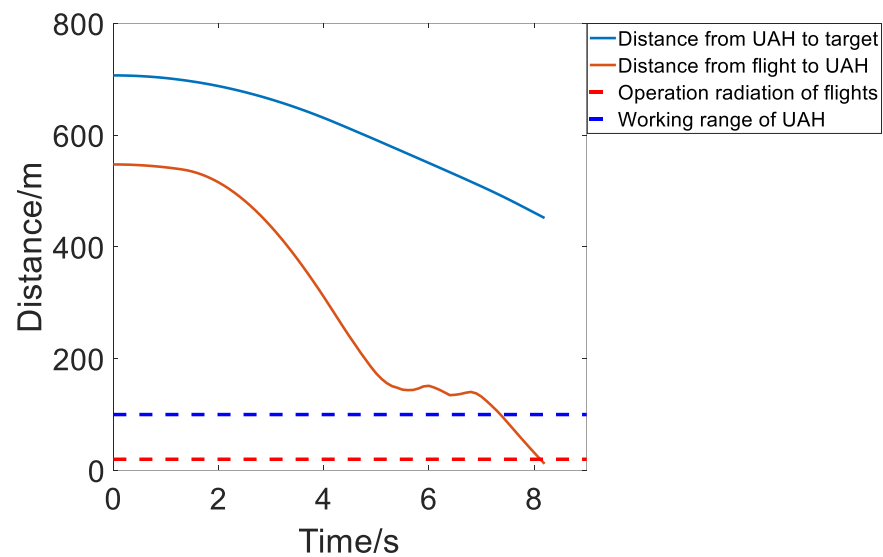


Figure 14. Distance from the UAH to the target, and distance from the flight to the UAH, in Case2.

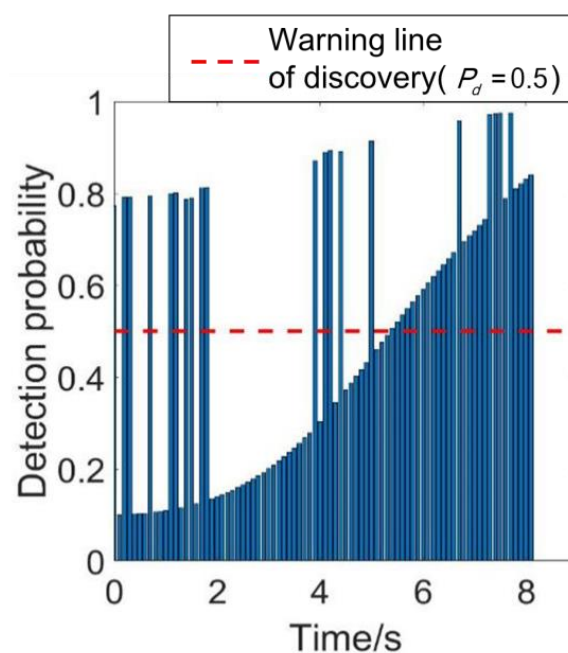


Figure 15. Variation in the detection probability in Case2.

From the discussions above, it can be obtained that the weight of the stealth elements in Case2 ($w_2 = 0.25$) can only bring slight maneuvers to the UAH, and is too low for the UAH to evade the flight operation.

5.2.3. Result Discussions of Case3

The path of the UAH and the track of flight in Case3, shown in Figures 16–18, are the graphs of distance variations and variation in the detection probability. Compared with Case1 and Case2, the UAH can survive for a longer time, but it is shot down at a position rather far away from the target. Meanwhile, the flight does not launch until $t = 8.1$ s, and finally hits the UAH at $t = 11.4$ s.

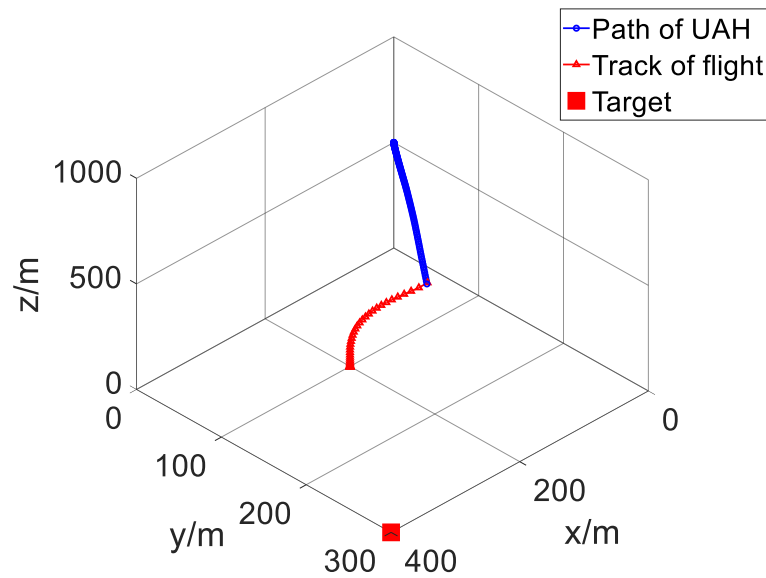


Figure 16. Path-planning result of the UAH, and track of flight, in Case3.

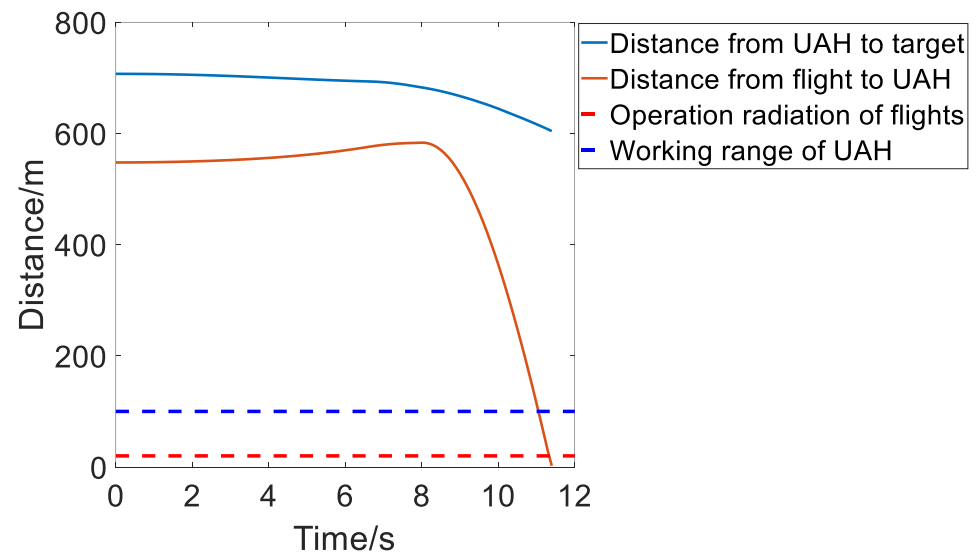


Figure 17. Distance from the UAH to the target, and distance from the flight to the UAH, in Case3.

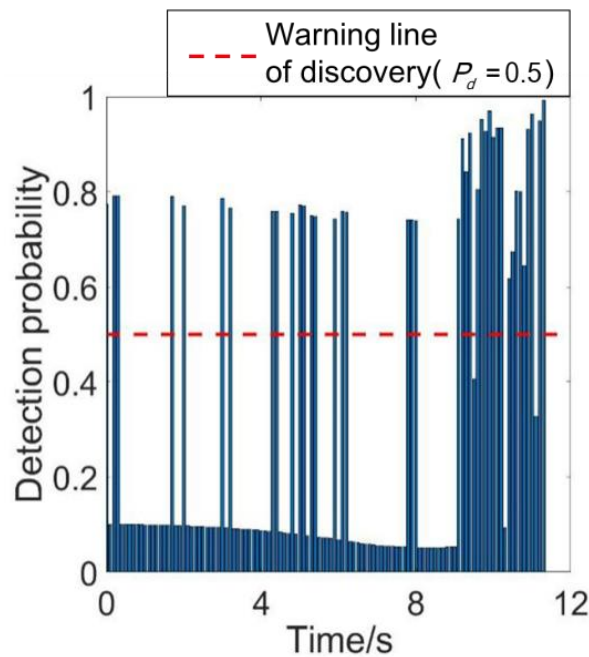


Figure 18. Variation in the detection probability in Case3.

Combining Figures 16–18, it can be concluded that the UAH in Case3 has been trying to evade radar detection since the path-planning simulation started. From Figure 19, it can be seen that the UAH tries to maneuver away from the radar by increasing its altitude. However, the maneuver away from the radar seems limited, for the reason that the weight of the distance element still equals 0.5, and the reward function might decrease significantly if the UAH maneuvers from the radar very fast, despite the increase in the UAH-target distance. Even worse, the weight values of $w_1 = 0.5$, $w_2 = 0.5$ cannot ensure a low detection probability, and the UAH reaches a distance rather close to the radar, as well as the flight launcher, when the flight launches. So, it is harder for the UAH to avoid the flight operation, and the flight hits the UAH very soon.

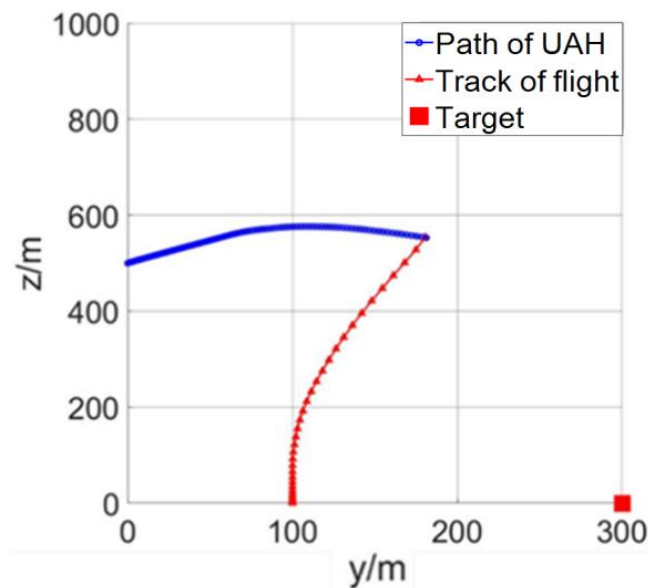


Figure 19. View of Figure 16 at the direction of [1,0,0].

5.2.4. Result Discussions of Case4

Figure 20 shows the path of the UAH, as well as the track of flight, in Case4, and the graphs of distance variations and the variation in the detection probability are shown in Figures 21 and 22. It can be seen from Figure 20 that the length of the UAH path appears much shorter. However, the survival time of the UAH is 13 s according to Figure 22, even longer than the survival time in Case3. Thus, it can be inferred that the UAH tries to keep away from the radar by slowing down the speed and, meanwhile, changing its attitude (yaw angle) to take advantage of its low RCS orientations. However, it seems that the UAH is still not willing to maneuver away from the radar at the cost of increasing the distance from the target. This can be explained as the effect of the weight of the distance $w_1 = 0.25$, which still performs as a limitation on UAH maneuvers. Therefore, with the weight of the stealth element $w_2 = 0.75$, the path-planning result of the UAH in Case4 can contribute to suppressing the detection probability, but the detection probability still raises above 50% with the UAH approaching the target, as well as the radar.

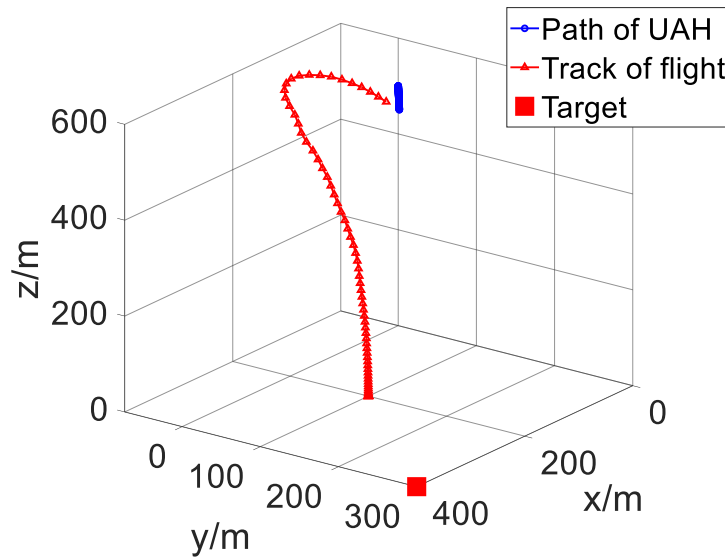


Figure 20. Path-planning result of the UAH, and track of flight, in Case4.

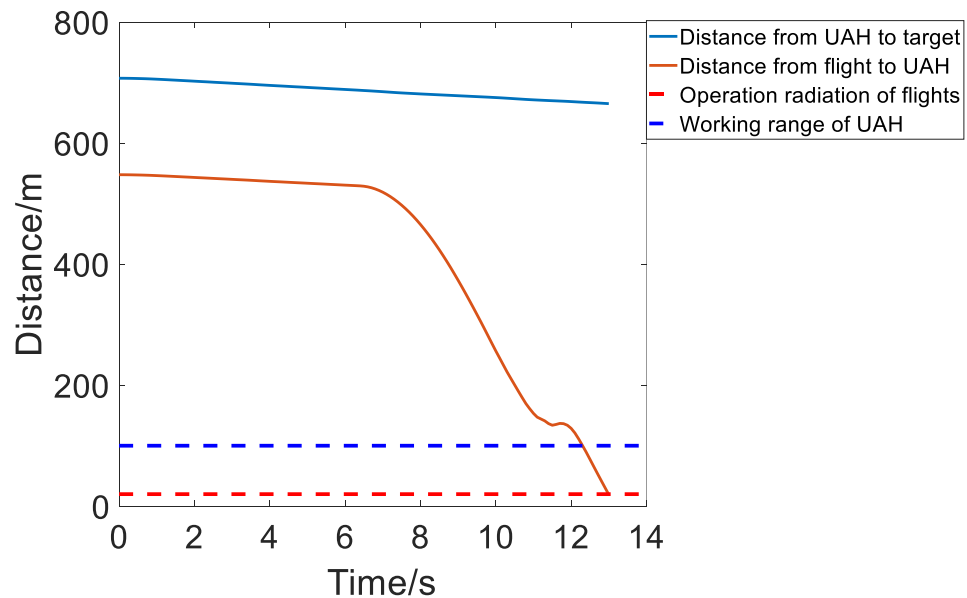


Figure 21. Distance from the UAH to the target, and distance from the flight to the UAH, in Case4.

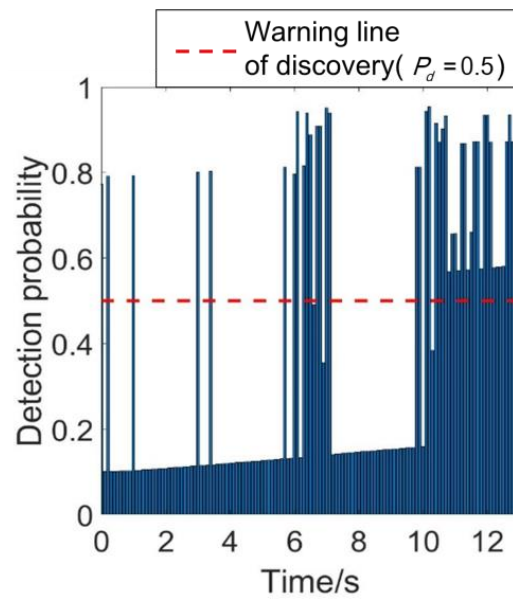


Figure 22. Variation in the detection probability in Case4.

5.2.5. Result Discussions of Case5

Figure 23 shows the path of the UAH, as well as the track of flight, in Case5, and the graphs of distance variations and the variation in the detection probability are shown in Figures 24 and 25. In Case5, the UAH successfully dodges the flight operation and, finally, reaches the target within the working range of the UAH. Combined with Figures 23 and 25, it can be inferred that the UAH tries to keep away from radar detection at any cost, even despite the significant increase in the distance from the target. With a low detection probability and long distance, it is very difficult for the flight to track and reach the UAH, and then the flight runs out of energy and finally crashes into the ground. Although the UAH has successfully approached the target and has completed the operation mission, the length of the path seems too long, and the cost of time reaches approximately 150 s. If the operation mission is an emergency mission that must to be accomplished within 2 min, the UAH might not operation the target in time, and the mission might be a failure in some sense.

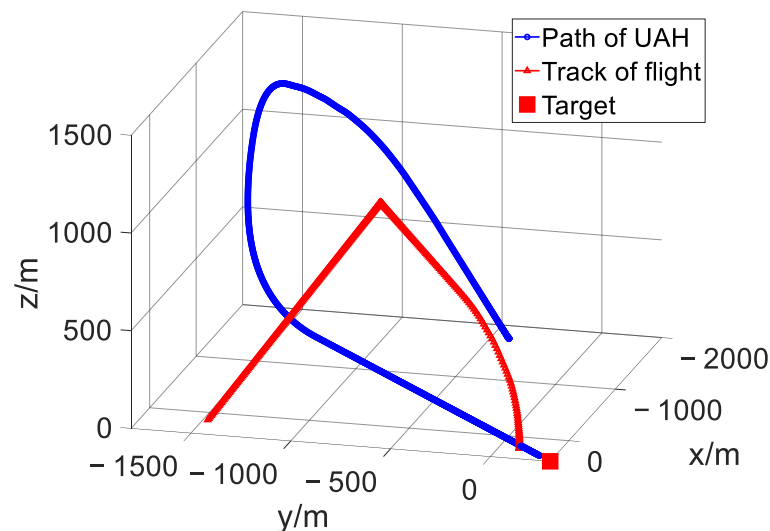


Figure 23. Path-planning result of the UAH and the track of flight in Case5.

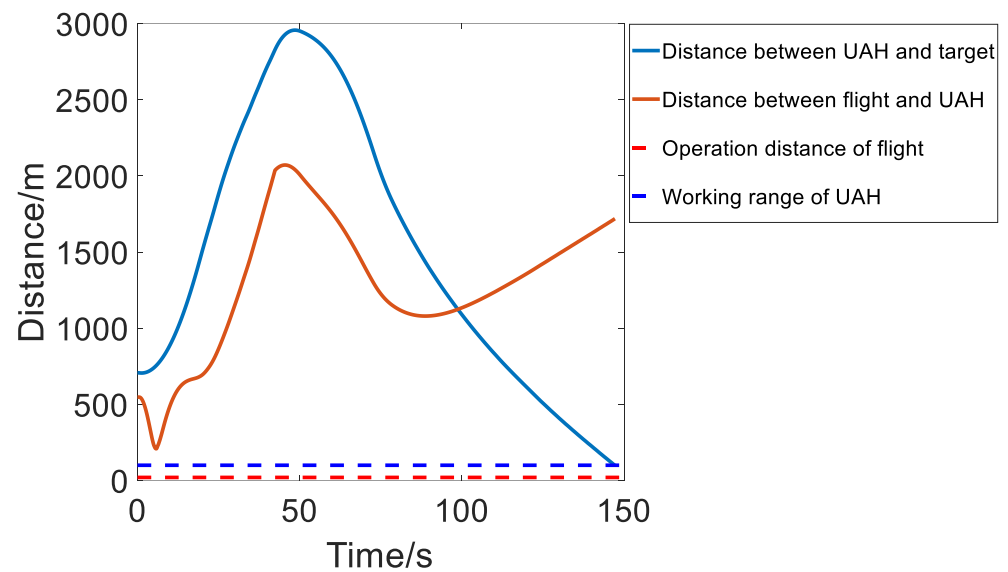


Figure 24. Distance from the UAH to the target, and distance from the flight to the UAH, in Case5.

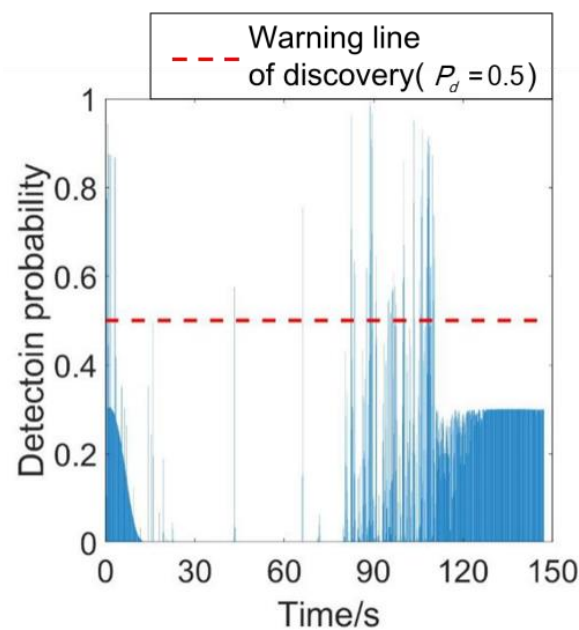


Figure 25. Variation in the detection probability in Case5.

It can be concluded from Case1 to Case5 that the selection of proper weight values for the reward function of the DQN algorithm for path planning is significantly important. If the weight of the stealth element is too low, the UAH will be easily detected by a radar with outstanding detection ability, and shot down by a flight with brilliant maneuverability. On the contrary, the time consumption might be excessive if the weight of the stealth element is too high. One thing to note is that the requirement of the weight values can change with changes in the environment. For example, the weight of the stealth element should be high when the flight is tracking the UAH, but the weight of the stealth element should be high when the flight no longer poses threats to the UAH. Therefore, a DQN algorithm with dynamic weights can be helpful, and that is the SDDW-DQN algorithm proposed in this paper.

5.2.6. Result Discussions of Case6 (Using the SDDW-DQN Algorithm)

The path of the UAH and the track of the flight in Case6, shown in Figures 26–28 show the graphs of the distance variations and the variation in the detection probability. The UAH in Case6 successfully reaches its working range after dodging the flight operation. Compared with Case5, the length of the path of the UAH in Case6 is much shorter, without the significant increase in the distance to the target. Moreover, the time consumption of the UAH in Case6 only equals 55 s, less than half of that in Case5. The reason why the path-planning result in Case6 has the advantage of both survivability and time cost can also be explained by Figure 28. It can be seen from Figure 28 that, although the detection probability is obviously higher than that in Case5, the detection probability in Case6 remains below 50% at most of the time spots before $t = 30$ s. The dynamic weights contribute to the real-time adjustment of the reward function, so that the UAH can suppress the detection probability at a low level and, meanwhile, avoid the significant cost of distance. After the flight operates self-destruction at $t = 31.2$ s, the weight of the stealth element becomes 0 at once, because the flight will no longer bring threats to the UAH, and the UAH should approach the target as soon as possible, despite being detected by radar. Therefore, the SDDW-DQN algorithm proposed in this paper can contribute to this.

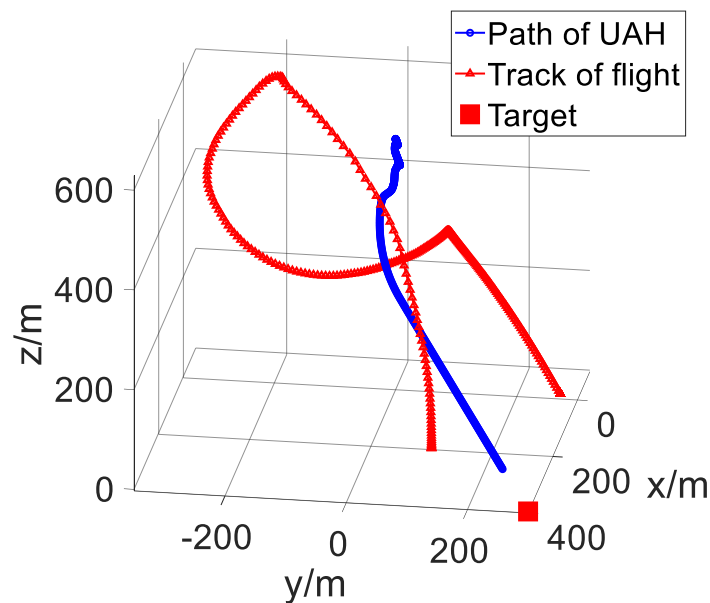


Figure 26. Path-planning result of the UAH, and path of the flight, in Case6.

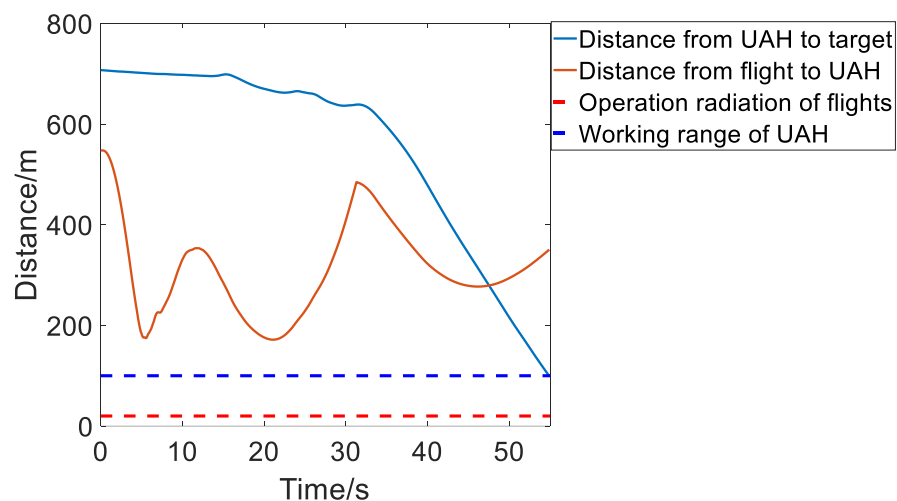


Figure 27. Distance from the UAH to the target, and distance from the flight to the UAH, in Case6.

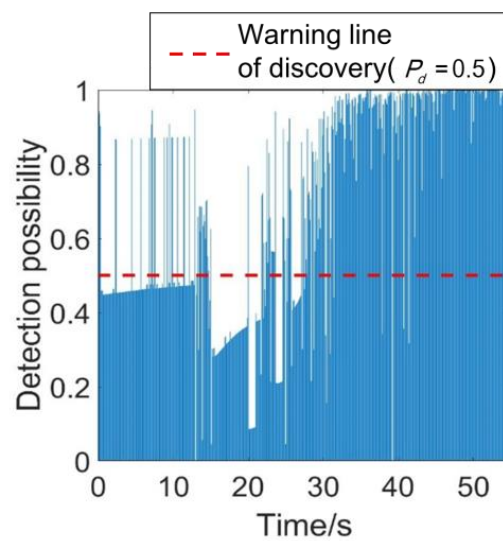


Figure 28. Variation in the detection probability in Case6.

To summarize, the advantages of the SDDW-DQN algorithm can be verified by comparing the cases above.

6. Conclusions

In this paper, a study on the path-planning problems of UAHs under the threat of radar detection and flight operation is carried out, and the stealth model of the UAH, and the dynamic models of the UAH and flight, are established. Aiming to obtain an optimized path, and realize evasion from radar detection and flight operation, a stealth-distance dynamic-weight Deep Q-Network (SDDW-DQN) algorithm is put forward for path planning in UAHs. During the establishment of the SDDW-DQN algorithm, a new reward function with dynamic weights related to detection possibility is designed. For the verification of the effectiveness of the SDDW-DQN algorithm for path-planning problems, path-planning simulations using the SDDW-DQN algorithm, as well as DQN algorithms with static weights, have been carried out in this paper. According to the simulation results, the following conclusions can be obtained:

- (1) The verification can be given from the simulation results in this paper that the SDDW-DQN algorithm has the ability of realizing path planning in UAHs. Under the threat of radar detection and flight operation, the UAH can evade these threats and, finally, reach the working range with the SDDW-DQN algorithm.
- (2) It is necessary for ensuring the survivability of UAHs that the stealth elements of UAHs are taking into consideration. The UAH can hardly survive the flight operation if the stealth weight w_2 is not high enough. Therefore, it is essential that proper values for the weights of the reward function are selected.
- (3) Dynamic weights in the reward function can be helpful for the UAH, to shorten the length of the path and raise the task execution efficiency, on the premise of evading flight operations. The path-planning result with the SDDW-DQN algorithm appears with a safer and shorter path of the UAH than that with the DQN algorithm with static weights.

In addition to the main conclusions above, some expanded conclusions can also be acquired by observing the path-planning result graphs, especially Figures 23 and 26.

- (4) In Figure 23, the UAH maintains stealth against detections for a long time, causing the flight to fly on inertia for too long, and finally run out of energy. Thus, the method of avoiding flight operations can be summarized as “blind”. The “blind” method can be realized via stealth design, and by maneuvering far enough away from detection threats.

- (5) In Figure 26, the UAH operates evasive maneuvers, and appears stealth, or exposed to radar now and then. The flight has to keep tracking the UAH by twisting its trajectory hard, leading to the increase in drag, and a more serious loss of energy. Therefore, another method of avoiding flight operations can be summarized as “exhaust”. The “exhaust” method can be realized by high-overloaded maneuvers or maneuvers with high stealth orientations facing detection threats at proper times.

For all the types of evasions from flight operations, “blind” and “exhaust” can comprise two broad approaches.

Author Contributions: Conceptualization, Z.W. and J.H.; writing—original draft preparation, Z.W.; writing—review and editing, J.H. and M.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 12004027.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- McEnroe, P.; Wang, S.; Liyanage, M. A Survey on the Convergence of Edge Computing and AI for UAVs: Opportunities and Challenges. *IEEE Internet Things J.* **2022**, *9*, 15435–15459. [\[CrossRef\]](#)
- Su, J.; Zhu, X.; Li, S.; Chen, W.-H. AI meets UAVs: A survey on AI empowered UAV perception systems for precision agriculture. *Neurocomputing* **2023**, *518*, 242–270. [\[CrossRef\]](#)
- Raja, A.; Njilla, L.; Yuan, J. Adversarial Attacks and Defenses Toward AI-Assisted UAV Infrastructure Inspection. *IEEE Internet Things J.* **2022**, *9*, 23379–23389. [\[CrossRef\]](#)
- Yin, X.; Peng, X.; Zhang, G.; Che, B.; Wang, C. Flight Control System Design and Autonomous Flight Control of Small-Scale Unmanned Helicopter Based on Nanosensors. *J. Nanoelectron. Optoelectron.* **2021**, *16*, 675–688. [\[CrossRef\]](#)
- Hoshu, A.A.; Wang, L.; Ansari, S.; Sattar, A.; Bilal, M.H.A. System Identification of Heterogeneous Multirotor Unmanned Aerial Vehicle. *Drones* **2022**, *6*, 309. [\[CrossRef\]](#)
- Gupta, P.; Pareek, B.; Singal, G.; Rao, D.V. Edge device based Military Vehicle Detection and Classification from UAV. *Multimedia Tools Appl.* **2021**, *81*, 19813–19834. [\[CrossRef\]](#)
- Wang, Z.; Henricks, Q.; Zhuang, M.; Pandey, A.; Sutkowsy, M.; Harter, B.; McCrink, M.; Gregory, J. Impact of Rotor-Airframe Orientation on the Aerodynamic and Aeroacoustic Characteristics of Small Unmanned Aerial Systems. *Drones* **2019**, *3*, 56. [\[CrossRef\]](#)
- Bhattacharya, P.; Gavrilova, M.L. Roadmap-Based Path Planning—Using the Voronoi Diagram for a Clearance-Based Shortest Path. *IEEE Robot. Autom. Mag.* **2008**, *15*, 58–66. [\[CrossRef\]](#)
- Chi, W.; Ding, X.; Wang, J.; Chen, G.; Sun, L. A Generalized Voronoi Diagram-Based Efficient Heuristic Path Planning Method for RRTs in Mobile Robots. *IEEE Trans. Ind. Electron.* **2021**, *69*, 4926–4937. [\[CrossRef\]](#)
- Song, C.; Cha, J.; Lee, M.; Kim, D.-S. Dynamic Voronoi Diagram for Moving Disks. *IEEE Trans. Vis. Comput. Graph.* **2021**, *27*, 2923–2940. [\[CrossRef\]](#)
- Alhassow, M.M.; Ata, O.; Atilla, D.C. Car-like Robot Path Planning Based on Voronoi and Q-Learning Algorithms. In Proceedings of the 2021 International Conference on Engineering and Emerging Technologies (ICEET), Istanbul, Turkey, 27–28 October 2021; pp. 591–594. [\[CrossRef\]](#)
- Zhang, T.; Wang, J.; Meng, M.Q.H. Generative Adversarial Network Based Heuristics for Sampling-Based Path Planning. *IEEE-CAA J. Autom. Sin.* **2021**, *9*, 64–74. [\[CrossRef\]](#)
- Ma, N.; Wang, J.; Liu, J.; Meng, M.Q.H. Conditional Generative Adversarial Networks for Optimal Path Planning. *IEEE Trans. Cogn. Dev. Syst.* **2022**, *14*, 662–671. [\[CrossRef\]](#)
- Zhu, H.; Wang, Y.; Ma, Z.; Li, X. A Comparative Study of Swarm Intelligence Algorithms for UCAV Path-Planning Problems. *Mathematics* **2021**, *9*, 171. [\[CrossRef\]](#)
- Du, N.; Zhou, Y.; Deng, W.; Luo, Q. Improved chimp optimization algorithm for three-dimensional path planning problem. *Multimedia Tools Appl.* **2022**, *81*, 27397–27422. [\[CrossRef\]](#)
- Hu, J.; Wang, L.; Hu, T.; Guo, C.; Wang, Y. Autonomous Maneuver Decision Making of Dual-UAV Cooperative Air Combat Based on Deep Reinforcement Learning. *Electronics* **2022**, *11*, 467. [\[CrossRef\]](#)
- Ren, J.; Huang, X.; Huang, R.N. Efficient Deep Reinforcement Learning for Optimal Path Planning. *Electronics* **2022**, *11*, 3628. [\[CrossRef\]](#)
- Chen, X.; Qi, Y.; Yin, Y.; Chen, Y.; Liu, L.; Chen, H. A Multi-Stage Deep Reinforcement Learning with Search-Based Optimization for Air–Ground Unmanned System Navigation. *Appl. Sci.* **2023**, *13*, 2244. [\[CrossRef\]](#)

19. Wei, K.; Huang, K.; Wu, Y.; Li, Z.; He, H.; Zhang, J.; Chen, J.; Guo, S. High-Performance UAV Crowdsensing: A Deep Reinforcement Learning Approach. *IEEE Internet Things J.* **2022**, *9*, 18487–18499. [\[CrossRef\]](#)
20. Zheng, S.; Liu, H. Improved Multi-Agent Deep Deterministic Policy Gradient for Path Planning-Based Crowd Simulation. *IEEE Access* **2020**, *7*, 147755–147770. [\[CrossRef\]](#)
21. Li, D.; Yin, W.; Wong, W.E.; Jian, M.; Chau, M. Quality-Oriented Hybrid Path Planning Based on A* and Q-Learning for Unmanned Aerial Vehicle. *IEEE Access* **2021**, *10*, 7664–7674. [\[CrossRef\]](#)
22. Zhu, X.; Wang, L.; Li, Y.; Song, S.; Ma, S.; Yang, F.; Zhai, L. Path planning of multi-UAVs based on deep Q-network for energy-efficient data collection in UAVs-assisted IoT. *Veh. Commun.* **2022**, *36*, 100491. [\[CrossRef\]](#)
23. Xu, Y.; Wei, Y.; Jiang, K.; Wang, D.; Deng, H. Multiple UAVs Path Planning Based on Deep Reinforcement Learning in Communication Denial Environment. *Mathematics* **2023**, *11*, 405. [\[CrossRef\]](#)
24. Yao, J.; Li, X.; Zhang, Y.; Ji, J.; Wang, Y.; Liu, Y. Path Planning of Unmanned Helicopter in Complex Environment Based on Heuristic Deep Q-Network. *Int. J. Aerosp. Eng.* **2022**, *2022*, 1360956. [\[CrossRef\]](#)
25. Zhao, J.; Gan, Z.; Liang, J.; Wang, C.; Yue, K.; Li, W.; Li, Y.; Li, R. Path Planning Research of a UAV Base Station Searching for Disaster Victims' Location Information Based on Deep Reinforcement Learning. *Entropy* **2022**, *24*, 1767. [\[CrossRef\]](#)
26. Cheng, Y.; Zhang, W. Concise deep reinforcement learning obstacle avoidance for underactuated unmanned marine vessels. *Neurocomputing* **2017**, *272*, 63–73. [\[CrossRef\]](#)
27. Yao, J.; Li, X.; Zhang, Y.; Ji, J.; Wang, Y.; Liu, Y. Path Planning of Unmanned Helicopter in Complex Dynamic Environment Based on State-Coded Deep Q-Network. *Symmetry* **2022**, *14*, 856. [\[CrossRef\]](#)
28. Zhou, Z.; Huang, J.; Yi, M. Comprehensive optimization of aerodynamic noise and radar stealth for helicopter rotor based on Pareto solution. *Aerosp. Sci. Technol.* **2018**, *82–83*, 607–619. [\[CrossRef\]](#)
29. Zhou, Z.; Huang, J.; Wu, N. Acoustic and radar integrated stealth design for ducted tail rotor based on comprehensive optimization method. *Aerosp. Sci. Technol.* **2019**, *92*, 244–257. [\[CrossRef\]](#)
30. Zhang, Y.; Hu, M.; Lai, D.; Ma, G. Simulation on Aircraft Infrared Exposed Range and Detection Probability. *J. Syst. Simul.* **2016**, *28*, 441–448.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.