Journal of Systems Engineering and Electronics
Vol. 35, No. 2, April 2024, pp.374 – 385

Real-time UAV path planning based on LSTM network

ZHANG Jiandong[1], GUO Yukun[1,2], ZHENG Lihui[1,3], YANG Qiming[1,*], SHI Guoqing[1], and WU Yong[1]

1. School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China;
2. The Flight Automatic Control Research Institute of AVIC, Xi'an 710065, China;
3. Military Representative Office of Marine Wuhan Bureau in Luoyang Area, Luoyang 471000, China

Abstract: To address the shortcomings of single-step decision making in the existing deep reinforcement learning based unmanned aerial vehicle (UAV) real-time path planning problem, a real-time UAV path planning algorithm based on long short-term memory (RPP-LSTM) network is proposed, which combines the memory characteristics of recurrent neural network (RNN) and the deep reinforcement learning algorithm. LSTM networks are used in this algorithm as Q-value networks for the deep Q network (DQN) algorithm, which makes the decision of the Q-value network has some memory. Thanks to LSTM network, the Q-value network can use the previous environmental information and action information which effectively avoids the problem of single-step decision considering only the current environment. Besides, the algorithm proposes a hierarchical reward and punishment function for the specific problem of UAV real-time path planning, so that the UAV can more reasonably perform path planning. Simulation verification shows that compared with the traditional feed-forward neural network (FNN) based UAV autonomous path planning algorithm, the RPP-LSTM proposed in this paper can adapt to more complex environments and has significantly improved robustness and accuracy when performing UAV real-time path planning.

Keywords: deep Q network, path planning, neural network, unmanned aerial vehicle (UAV), long short-term memory (LSTM).

DOI: 10.23919/JSEE.2023.000157

1. Introduction

Nowadays, unmanned aerial vehicles (UAVs) are widely used in daily life and excel in search, rescue, mapping, surveillance, and other fields [1−4]. As the main problem faced in UAV applications, the UAV path planning problem has also received more and more attention [5]. Real-time UAV path planning is an important factor in whether the UAV can complete its mission and conduct autonomous controlled flight, and its goal is to obtain a path that satisfies the requirements from the origin to the intended destination with unknown environmental information, and the path needs to satisfy constraints such as UAV maneuverability and endurance time [6].

The real-time path planning problem has been studied by many research results, like D* algorithm [7−10], lifelong planning A* (LPA*) algorithm [11−13], and D*Lite algorithm [14−16], but when high-dimensional obstacle data are detected from unknown environment, the traditional map-based algorithms have some limitations due to the difficulty of mapping. Moreover, when the unknown environment for real-time path planning is dynamic, it can cause the constructed maps to lose timeliness and become inaccurate [17].

The rise of deep reinforcement learning has provided a new way of thinking for path planning. In the past decade, the performance of deep reinforcement learning techniques has improved significantly and has been widely used in various fields such as autonomous driving [18,19], natural language processing [20,21], and computer vision [22,23]. It combines the powerful perceptual and representational capabilities of deep neural networks for processing high-dimensional decision information with the decision learning capabilities of reinforcement learning through trial and error, and has excellent performance in solving complex problems [24]. The neural networks are trained by deep reinforcement learning algorithms and the trained neural networks are used for real-time path planning to obtain the best path. This method does not depend on map mapping and is able to perform path planning even without obstacle maps, which can effectively overcome the problems of traditional algorithms [25].

Manuscript received September 15, 2022.
*Corresponding author.
This work was supported by the Natural Science Basic Research Program of Shaanxi (2022JQ-593).

2. Related work

Currently, deep learning and reinforcement learning have been applied in the field of path planning. In 2020, Venturini et al. [26] proposed a path planning method based on a deep reinforcement learning that was able to successfully perform path planning on a square cell map. In 2021, another simulation experiment based on a real map was conducted [27], but the success of the method was very dependent on the initialization. A deep reinforcement learning method for UAV path planning in dynamic environments with potential enemy threats was proposed in the [28], which uses a series of situational map neural networks to obtain the Q-value of the action set to plan the path, but the action set direction of this method is fixed and does not consider the UAV motion constraint, and the obtained path may not be suitable for UAV flight.

Chen et al. [29] implemented an UAV obstacle avoidance model based on feed-forward neural network (FNN), in which the UAV can effectively avoid obstacles and plan a reasonable path using the trained FNN. However, the FNN only considers the current environmental input, and path planning as a complex decision problem is not accurate enough to make a single-step decision only by relying on the current environmental situation.

A local path planning method for mobile robots based on long short-term memory (LSTM) networks and reinforcement learning was implemented in [30]. The method combines LSTM network with reinforcement learning algorithms to solve the problem of local deadlock and path redundancy in the robot's planning process in unknown and complex environments, and the method is also able to improve the success rate of path planning and optimize the path length. However, the path planning problem in dynamic environments is not mature enough and has certain defects.

The above literature show that there are still many shortcomings in the research of UAV path planning based on deep reinforcement learning. The existing research is mainly based on FNN, but FNN can only use the current environmental information and cannot use the previously explored environment information, which will make the intelligent body make judgment with insufficient information and lead to problems such as low accuracy or poor robustness of path planning. The memory function of LSTM neural networks is suitable for complex multi-step decision problems such as UAV path planning [31], but the current LSTM network-based UAV real-time path planning is not comprehensive enough. In this paper, a real-time UAV path planning algorithm based on LSTM (RPP-LSTM) is proposed to address the current problems of deep reinforcement learning for UAV real-time path planning.

The algorithm introduces the LSTM network as the Q-value network of the DQN algorithm, which enables the intelligent body to effectively use the historical information when making decisions and solves the problem that the FNN decision only considers the current environment. In addition to the introduction of the LSTM network, a hierarchical reward and punishment function is also proposed in order to adapt the DQN model based on the LSTM network. Because of these improvements, the RPP-LSTM effectively improves the accuracy and robustness of the real-time UAV path planning model based on the deep reinforcement learning algorithm.

The way this paper develops is shown as follows: Firstly, we establish the UAV real-time path planning model. Secondly, we construct the Markov decision process based on the UAV motion model and the motion scenario, and get the DQN algorithm model based on the Markov decision process. Then we combine the DQN algorithm with the LSTM network to get the RPP-LSTM. Finally, through simulation experiments and result analysis, we verify the feasibility and validity of the proposed algorithm through simulation experiments and result analysis.

3. Modeling the UAV path planning problem

Based on the UAV motion process and motion constraints, the UAV motion model is established. Then combine the real-time path planning constraints to get the UAV path planning problem model.

3.1 UAV motion model

In this paper, assuming that the UAV flight altitude is a certain constant, the motion of the UAV can be reduced to a two-dimensional motion in the X-Y plane. Based on this assumption, the two-dimensional fixed coordinate system (x_t, y_t) is used to represent the absolute position of the UAV at the moment of t. ψ_t is used to represent the yaw angle of the UAV's current trajectory, so $[x_t, y_t, \psi_t]$ can represent the current position state of the UAV.

The change in the UAV position state quantity is then controlled by the UAV control variables $[\omega_t, v_t, \Delta t]$, including the UAV's trajectory yaw rate ω_t at the moment of t, the flight speed v_t and the time step Δt. Based on this variable, the equation for the increment of

the position state variable is as follows:

$$\begin{cases}\Delta\psi_t=\omega_t\Delta t\\ \Delta x_t=v_t[\sin(\Delta\psi_t+\psi_t)-\sin\psi_t]/\omega_t,\ \ \omega_t\neq 0\\ \Delta y_t=v_t[\cos(\Delta\psi_t+\psi_t)-\cos\psi_t]/\omega_t,\ \ \omega_t\neq 0\end{cases}. \tag{1}$$

Then the position state variable at the next moment is

$$\begin{bmatrix}x_{t+1}\\ y_{t+1}\\ \psi_{t+1}\end{bmatrix}=\begin{bmatrix}x_t\\ y_t\\ \psi_t\end{bmatrix}+\begin{bmatrix}\Delta x_t\\ \Delta y_t\\ \Delta\psi_t\end{bmatrix}. \tag{2}$$

In (2), $t+1=t+\Delta t$ indicates the next decision time.

Due to the limitations of their own physical conditions, UAVs have some limiting properties during flight, and if the UAV operating parameters exceed these limiting property values, the safety of the UAV itself decreases significantly, so the following constraints exist for UAVs in flight.

(i) Maximum cornering angle constraint

Assuming that the UAV flies at a constant speed V and its sampling time is ΔT, then the step length L can be approximated as $L=V\cdot\Delta T$. The maximum lateral overload of the UAV is $N_{y\max}$ and the track deflection angle is χ. Then the UAV makes a horizontal turn as shown in Fig. 1.

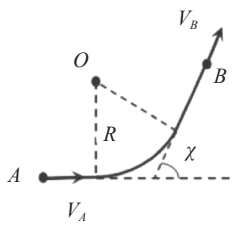

Fig. 1 Aircraft turning angle diagram

Point A is the current track point, V_A is the current velocity direction, B is the next track point, and V_B is the velocity direction of the next track point.

Due to the limitation of the maximum lateral overload, the UAV horizontal turn in a certain step of the existence of the maximum horizontal turn angle $\chi_{\max}$.

$$\chi_{\max}=2\arcsin\left(\frac{L}{2R_{\min}}\right) \tag{3}$$

where

$$R_{\min}=\frac{V^2}{g\sqrt{N_{y\max}^2-1}}. \tag{4}$$

(ii) Maximum flight distance constraint

Assuming that a path is planned with n path points, the maximum flight distance constraint is obtained by introducing the error amount Δ_L taking into account the aircraft engine performance. This is shown in

$$L+\Delta_L=\sum_{i=1}^{n}L_i+\Delta_L<L_{\max} \tag{5}$$

where L is the total path length, Δ_L is the amount of travel error, L_i is the distance between path point i and path point i+1, and $L_{\max}$ is the maximum flight distance.

Based on the above UAV motion model, the motion process of the UAV for each decision can be obtained as shown in Fig. 2.

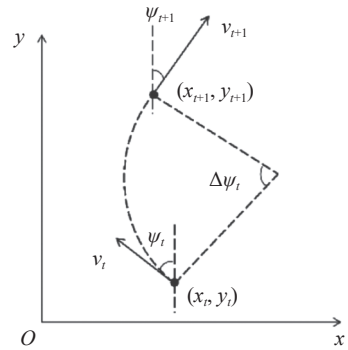

Fig. 2 UAV decision making process

$[x_t, y_t, v_t, \psi_t]$ is the motion parameter of the current path point, $[x_{t+1}, y_{t+1}, v_{t+1}, \psi_{t+1}]$ is the motion parameter of the next path point, and the arc between the two path points is the actual track.

3.2 Path planning evaluation indicators

In the process of real-time path planning, the UAV will encounter various types of obstacles. The UAV needs to make decisions based on the current position state and environment, so that the UAV can successfully avoid obstacles, reach the target point and complete the task. For different decision making methods, the final paths obtained are often different. In order to judge the merit of the UAV path, it can be judged based on the following indicators.

(i) Length of path

Path length is the most obvious indicator to judge the merit of a path. For a path, the shorter the path length is, the better the path is, provided that it can reach the target point smoothly. The equation for calculating the path length L_p is as follows:

$$L_p=\sum_{k=1}^{n}v_{tk}\Delta t_k \tag{6}$$

where n is the total step length of the planned path. v_{tk} is the velocity of the UAV of moment k. Δt_k is the time step of moment k.

(ii) Path smoothness

The path smoothness expresses the adjustment of the trajectory yaw angle during the overall process of the UAV path from the starting point to the target point. In this paper, the variance σ_ψ^2 of the trajectory yaw angle is used to determine this metric.

$$\sigma_\psi^2 = \frac{\sum_{i=1}^{n}(\psi_i - \bar{\psi})}{n-1} \tag{7}$$

For a path, a smaller σ_ψ^2 indicates a smaller change in the yaw angle of the path trajectory and a smoother UAV path. When σ_ψ^2 is 0, the path is a straight line.

(iii) Minimum distance between the path point and the obstacle

In the case of two paths with the same trend and little difference in length, the minimum distance between the path point and the obstacle is the key indicator to judge the superiority of the two paths. Under the premise of ensuring the safety distance, when the distance between the path point and the obstacle is smaller, it shows that the higher the avoidance accuracy of the path. The equation for calculating the minimum distance between the path point and the obstacle is

$$d_{\min} = \min(d_{t1}, d_{t2}, \cdots, d_{tn}) \tag{8}$$

where d_{ti} is the minimum distance between path point i and the obstacle.

UAVs are generally equipped with a variety of sensors, including cameras, radar, and ultrasonic rangefinders, in order to obtain the current environmental conditions during path planning. The environmental information is obtained based on the data measured by these sensors. Since the sensors such as camera and radar have their own limitations, this paper adopts the distance information measured by ultrasonic rangefinder as the current environment information.

The distance information of ultrasonic range finder is shown in Fig. 3, when the UAV detects the surrounding environment information d_i at the moment t, it can get the distance information of obstacles in different directions. When the obstacle distance is greater than the maximum distance $d_{\max}$ of the ultrasonic detector, $d_i = d_{\max}$. When d_i is small, it means that the UAV has approached the obstacle in that direction.

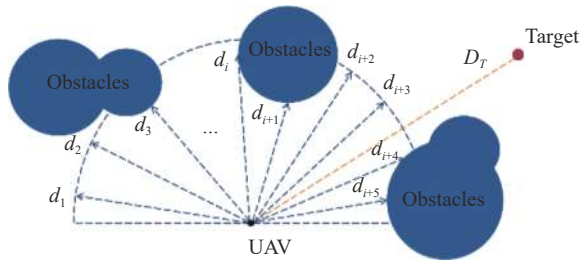

Fig. 3 Obstacle distance information schematic

4. UAV real-time path planning decision model

The UAV path planning model is transformed into an MDP, and the RPP-LSTM algorithm is proposed based on the DQN algorithm, using the memory property of the LSTM network.

4.1 Constructing Markov decision processes

The MDP is usually defined as an (S, A, ρ, f) quadruplet, where S is the set of all environmental states $\boldsymbol{s}_t$, $\boldsymbol{s}_t$ is the environmental state the agent is in at moment t, A is the set of all actions a_i the agent can make, $r_t \sim \rho(\boldsymbol{s}_t, a_i)$ is the immediate reward the agent receives for making an action a_i in environment $\boldsymbol{s}_t$, and $\boldsymbol{s}_{t+1} \sim f(\boldsymbol{s}_t, a_i)$ is the probability that the agent may make an action a_i at s_t to move to the next environmental state $\boldsymbol{s}_{t+1}$.

For the purpose of this paper, the state space of the path planning problem is

$$\boldsymbol{s}_t = [x_t, y_t, D, \psi, \alpha, \Delta\alpha, d_1, d_2, d_3, \cdots, d_9]. \tag{9}$$

(x_t, y_t) is the position coordinates of the current track point, D is the distance between the current track point and the target point, ψ is the yaw angle of the current track, α is the angle between the line of the UAV's current position and the target and the due north direction, and $\Delta\alpha$ is the difference between ψ and α. d_i is the distance of the obstacle within a certain angle in front of the UAV, and nine of the angle data are selected.

In this paper, the action set of the path planning problem is the difference in yaw angle of the UAV trajectory $\Delta\psi_i$, $0 \leqslant \Delta\psi_i \leqslant \chi_{\max}$.

Based on the obtained state set and the action set of the UAV path planning problem, an MDP can be constructed. The UAV starts from moment t, gets the surrounding environment state $\boldsymbol{s}_t$ by the current position information and ultrasonic rangefinder, and makes certain actions based on this environment state to reach the next path point. After reaching the next path point, the immediate reward value r_t is obtained according to the reward and punishment function, and then the above

behavior is repeated. The specific flow is shown in Fig. 4.

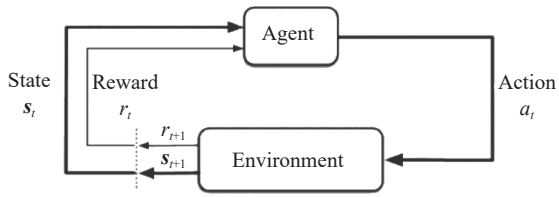

Fig. 4 MDP flow chart

For each action performed by the UAV, an immediate reward value is returned after interacting with the environment, but the Q value in the DQN algorithm is not simply a superposition of immediate reward values. Since the Markov decision process satisfies that the state of the next moment depends only on the state of the current moment and is independent of the previous state, when G_t is used to represent the reward value that the current tense has, then G_t is a superposition of the immediate reward r_t for all future moments.

$$G_t = r_{t+1} + \lambda r_{t+2} + \cdots = \sum_{k=0}^{\infty} \lambda^k r_{t+k+1} \tag{10}$$

where λ is the discount factor, which is generally less than 1. λ^k decreases as k increases, indicating that for G_t, the immediate reward of the current action is the most important, and the influence of the subsequent rewards gradually decreases. However, it is difficult to calculate the value of all rewards after the current moment, therefore, a value function needs to be introduced to evaluate the potential value of the current action. The equation of the value function is as follows:

$$Q_t(\boldsymbol{s}_t, a_t) = Q_t(\boldsymbol{s}_t, a_t) + \alpha(r_{t+1} + \lambda \max_a Q_t(\boldsymbol{s}_{t+1}, a_t)) - \alpha Q_t(\boldsymbol{s}_t, a_t) \tag{11}$$

where $Q_t(\boldsymbol{s}_t, a_t)$ at this point is the reward value for making the action a_t in the environment $\boldsymbol{s}_t$, and α is the learning rate.

4.2 DQN

The DQN algorithm is an algorithm that combines a neural network with a Q-learning algorithm. Using the powerful training ability of the Q-learning algorithm and the excellent fitting ability of the neural network, the DQN algorithm enables the system to learn from complex inputs and outputs.

The basic principle of the DQN algorithm is shown in Fig. 5.

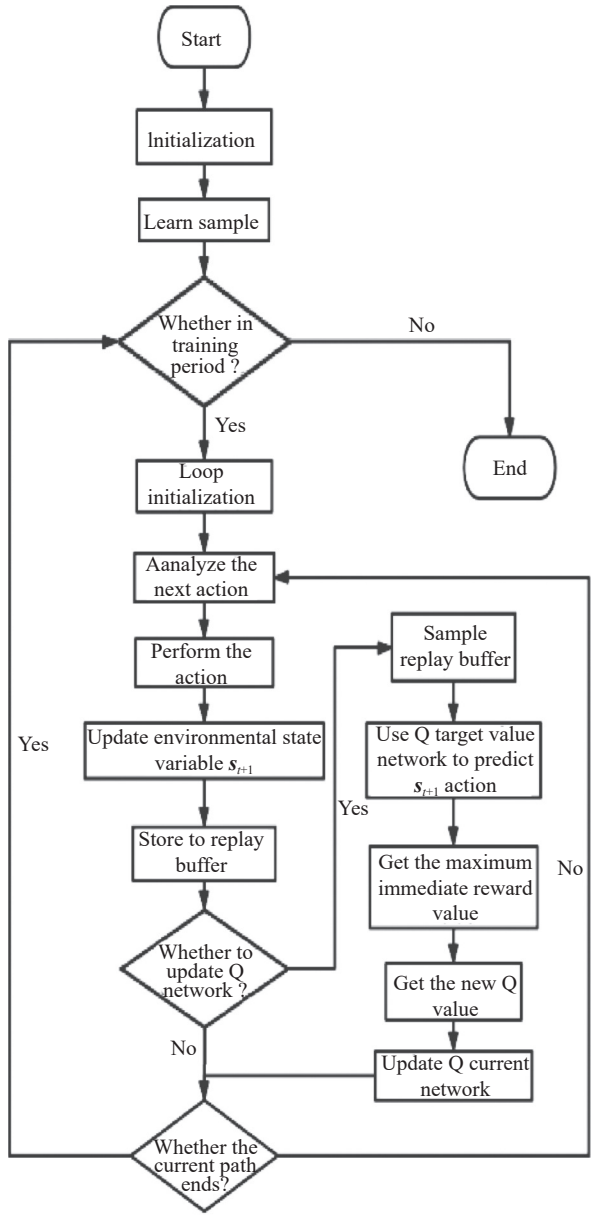

Fig. 5 DQN calculation process

Firstly, initialize the system, input the environment state quantity s into the current Q-value network, and get the action a that can get the maximum Q-value according to the action selection strategy. Transmit the execution action a to the environment system, get the next state environment state quantity $\boldsymbol{s}_{t+1}$ and the current reward r according to the environment information and the reward and punishment function, and then judge whether to terminate the current exploration by the current position. The obtained vector $(s, a, r, \boldsymbol{s}_{t+1})$ is stored in the playback memory unit during the exploration process; then, the playback memory unit is sampled after a certain period of time. The target value network uses the sampled vector $(s, a, r, \boldsymbol{s}_{t+1})$ to calculate the new a, and uses the stochas-

tic gradient descent algorithm to update the relevant parameters of the current value network to achieve the approximation of the complex function.

From Fig. 5, we can see that the important factors affecting the results of the DQN algorithm are action selection strategy and reward and punishment function. For the UAV path planning problem, this paper proposes a matching action selection strategy and reward and punishment function.

(i) Action selection strategy

Generally speaking, the probability of exploration is higher and the probability of utilization is lower in the early stage of path planning, while the probability of exploration is lower and the probability of utilization is higher in the later stage. Pursuing accurate estimation in the early stage and maximizing the reward as much as possible in the late stage will be able to accomplish the path planning task better. In this paper, we choose to use the improved ε-greedy strategy to explore as much as possible in the early stage and utilize as much as possible in the late stage in order to improve the training efficiency of the neural network and achieve the largest possible average reward. The equation is as follows:

$$\varepsilon_{j+1}=\begin{cases}\varepsilon_j+\Delta\varepsilon\cdot e^{-\text{num\_episode}}, & \varepsilon_j<\varepsilon_{\max}\\ \varepsilon_{\max}, & \text{else}\end{cases}. \tag{12}$$

(ii) Reward and punishment function

The reward and punishment function in this paper is shown in

$$r=\begin{cases}r_{\text{obstacle}}, & |q_t(x,y)-q_0(x,y)|\leqslant D_{\text{safe}}\\ r_{\text{goal}}, & \left|q_t(x,y)-q_g(x,y)\right|\leqslant L\\ r_{\psi}, & |\psi_t-\alpha|\leqslant\psi_r\\ r_d, & d_t-d_{t-1}\leqslant 0\end{cases} \tag{13}$$

where $q_t(x,y)$ is the coordinates of the current UAV position. $q_0(x,y)$ is the coordinates of the obstacle position. $q_g(x,y)$ is the coordinates of the target point position. D_{safe} is the minimum safe distance. L is the step length. ψ_t is the yaw angle of the current trajectory. α is the angle between the line of the current position of the UAV and the target and the due north direction. ψ_r is a constant value, and let it be the reward trajectory angle. d_t and d_{t-1} are the distances between the UAV and the target at this moment and the previous moment, respectively.

When the distance between the obstacle and the UAV is less than the safe distance, it is considered that the UAV collides with the obstacle, and the reward and punishment function returns the collision reward value r_{obstacle}. When the distance between the target point and the UAV is less than the step size, it is considered that the UAV reaches the target point and the mission is completed, and the reward and punishment function returns the arrival reward value r_{goal}.

When the absolute value of the difference between angle α and the current track yaw angle ψ_t is less than the reward track angle ψ_r, the UAV track direction is considered correct and the reward function returns the track angle reward value r_{ψ}. When the UAV is closer to the target point position than the previous step, i.e., $d_t-d_{t-1}\leqslant 0$, the reward function returns the distance reward value r_d.

The priority of these four types of rewards is shown in Table 1.

Table 1 Priority of each reward

Reward type	Priority
Collision reward	1
Arrival reward	2
Track angle reward	3
Distance reward	4

4.3 LSTM

The DQN algorithm has obvious advantages in solving complex problems. Since the Q-network of the basic DQN algorithm is an FNN, its action decision only considers the current environment state, so there are certain limitations for multi-step decision problems. The UAV real-time path planning problem is a typical multi-step decision problem, in which not only the current environment state but also the previous environment state needs to be considered when making the current decision. Thus this paper introduces LSTM network to solve the problem.

LSTM networks, as a type of recurrent neural network (RNN), have a more refined information transfer mechanism than traditional RNN. Through a special network design, LSTM networks are able to change memory in a very precise way, applying a specialized learning mechanism to remember and update information, which helps to track information over a longer period of time and can remember historical information with very long time intervals. For the path planning problem in this paper, the LSTM network is theoretically well adapted and is used as the Q-network in the DQN algorithm.

The input of the LSTM network is the time series information. In this paper, the information of the previous path point and the information of this path point are used as the input of the time series $\boldsymbol{X}_{t-2}$, $\boldsymbol{X}_{t-1}$, $\boldsymbol{X}_t$ composed of the environmental state quantities $\boldsymbol{s}_{t-2}$, $\boldsymbol{s}_{t-1}$, $\boldsymbol{s}_t$, and the actions a_{t-2}, a_{t-1}, a_t. The output is the Q value corresponding to the action a_t. When a single-layer LSTM network is used for training, the specific network construction is shown in Fig. 6.

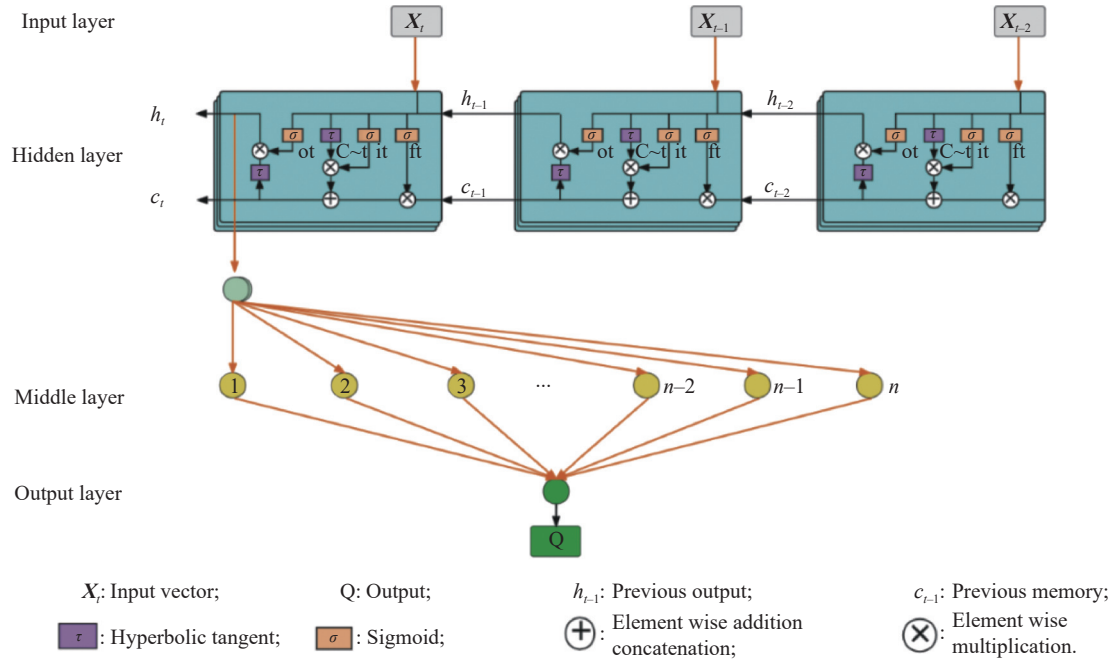

Fig. 6 LSTM network structure

Based on the MDP model, the state space can be obtained as (9). Therefore, the input time series of the Q-value network is a 15-dimensional vector, and the input layer of the Q-value network is 15×3 neurons. the parameters of the hidden layer of the Q-value network need to be tested in a large number of experiments to get the optimal results. In this paper, the experimental test results for the number of layers and the number of implicit layer LSTM networks show that a two-layer LSTM network with 40 LSTM networks in the first layer and 16 LSTM networks in the second layer are the best results. Since the output of the Q-value network is Q-value, the output layer is 1 neuron. The final Q-value network obtained is shown in Fig. 7.

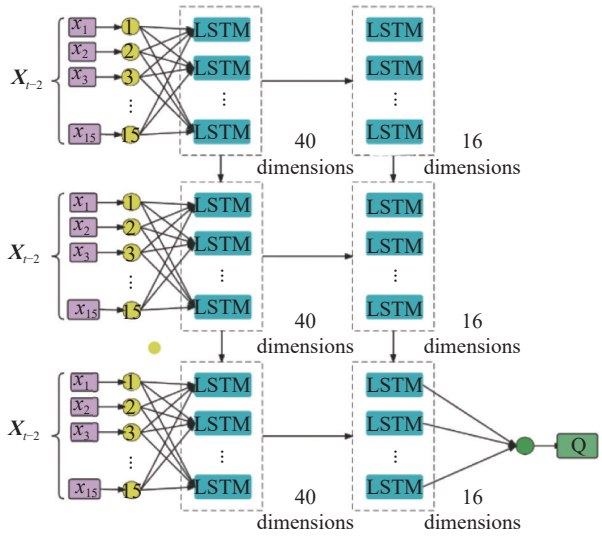

Fig. 7 Q-value network

The LSTM neural network is selected as the Q-value network and applied to the DQN algorithm to obtain the RPP-LSTM algorithm, and the detailed steps are shown in Algorithm 1.

Algorithm 1 RPP-LSTM algorithm

1. Initialize replay memory D to capacity N
2. Initialize action-value function Q with random weights
3. The Q uses expert samples for preliminary learning
4. For episode = 1 to M do
5. Receive an initial observation s_0
6. Initialize empty history h_0
7. For t=1 to T do
8. With probability ε select a random action a_t
9. Otherwise obtain avoid and acquire action from eval-network by
   Using ε-greedy policy
10. Execute action a_t in emulator and observe reward
11. Store transition $(\boldsymbol{s}_{t-2}, a_{t-2}, \boldsymbol{s}_{t-1}, a_{t-1}, \boldsymbol{s}_t, a_t, \boldsymbol{s}_{t+1}, r_t,)$ in D
12. Select a random number of historical tracks from D
13. Set

$$y_j = \begin{cases} r_j, & \text{for terminnal } s_{j+1} \\ r_j + \gamma \max_{a'} Q, & \text{for non-terminnal } s_{j+1} \end{cases}$$

14. Update target-network
15. End for
16. End for

5. Experiments

This section shows that the RPP-LSTM algorithm has better performance in the UAV path planning problem by comparing it with the traditional DQN algorithm based on FNN in different environment maps based on extensive simulation experiments.

5.1 Experimental settings

Set the size of the mission map for the simulation experiment as 20 km×20 km, assume that the UAV flies at a certain altitude, the speed v=200 m/s of the UAV, the discrete time interval ΔT is 1 s, and the maximum lateral overload of the UAV is 0.9. The calculated step length is about 200 m, the maximum trajectory yaw angle of the UAV is ±10°, and the minimum safety distance is 200 m. Choose the starting point of the UAV as (0,0,0.6) and the target point is (20,20,0.6) for training.

In addition to the above assumptions, the UAV ultrasonic rangefinder also needs to be able to detect the distance of obstacles at a certain angle directly in front of it. The maximum detection distance of the UAV radar is now set to two kilometers, the maximum detection angle is in front of the UAV, and obstacle distance sampling is performed at every interval.

The obstacles in the environmental space are regular geometric shapes, and the cylindrical obstacles are mainly selected in this paper. When the distance between UAV and obstacles is less than the minimum safe distance, the path planning is considered to fail.

Based on the above simulation conditions, the specific training parameters of the DQN algorithm are given as shown in Table 2.

Table 2 DQN parameter

Parameter	Priority
Collision reward	−10
Arrival reward	20
Track angle reward	5
Distance reward	5
γ	0.8
α	0.8
D_{safe}/km	0.2

5.2 Experiment I: path planning capability validation

This experiment is used to test whether there is a significant difference in the path planning ability of the two Q-value networks in the original network training environment, and to verify whether the RPP-LSTM algorithm has excellent path planning ability.

In the original network training environment there are 12 cylindrical obstacles with different radii. The starting coordinates of the UAV in this environment are (0,3,0.6) and the coordinates of the target point are (20,20,0.6), and the distance units in all experiments are kilometers. The direction from the starting point to the target point is taken as the initial velocity direction of the UAV, and the trained FNN and LSTM network are used for real-time UAV path planning, respectively, and the obtained results are shown in Fig. 8.

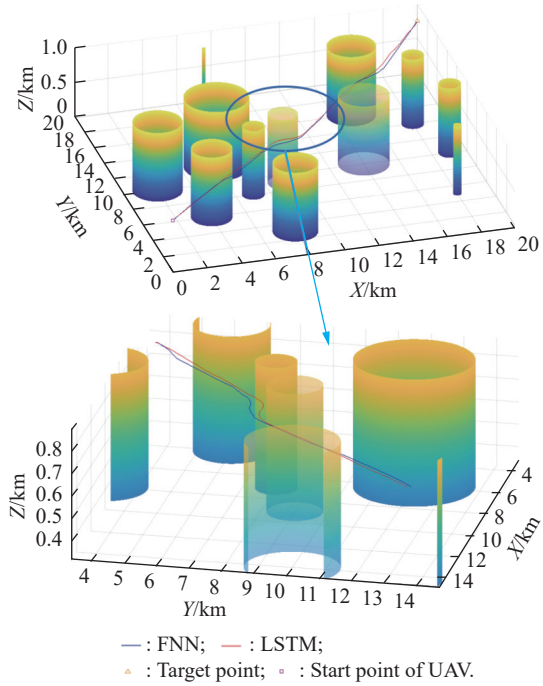

Fig. 8 Original environment path planning

In Fig. 8, the path planned based on FNN is assumed to be path 1, and the path planned based on LSTM network is path 2. It can be seen that the overall trend of path 1 and path 2 is the same, and both paths are relatively smooth and suitable for UAV flight. And both paths are sensitive to obstacle positions, and the distance between each path point and the obstacle is greater than the minimum safe distance, which can achieve effective obstacle avoidance and reach the target point safely. The experimental results show that both FNN and LSTM network can complete the path planning task and have good UAV path planning ability in the original training environment.

However, there are some differences between the two paths. Based on the formula of the path evaluation index, we can get the relevant evaluation of the two paths as shown in Table 3. It can be seen that the length of path 2 is smaller than the length of path 1, and the UAV can

reach the target point faster, saving the mission time and mission cost. The variance σ_{ψ}^2 of the track yaw angle of path 1 is also greater than the variance σ_{ψ}^2 of the track yaw angle of path 2, indicating that path 2 is smoother and requires less maneuverability for the UAV.

Table 3 Path data comparison 1

Evaluation indicator	FNN	LSTM
Path length/km	24.3	24.12
Trajectory yaw angle variance	738.85	188.58
Minimum distance to obstacle/m	430.7	354.4

Fig. 9 shows the curve of the minimum distance between the UAV and the obstacle as a function of time. The minimum distance between path 1 and the obstacle is 430.7 m, and the minimum distance between path 2 and the obstacle is 354.4 m, and path 1 as a whole is further away from the obstacle compared to path 2.

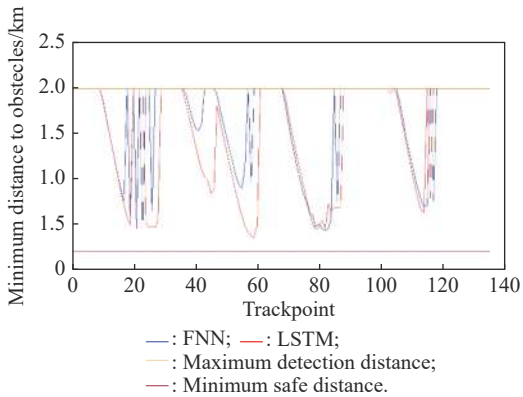

Fig. 9 Change in distance between path point and obstacle

From these indicators, it can be seen that although both paths successfully plan feasible paths, path 2 is shorter in length, requires less maneuverability from the UAV, and also has higher obstacle avoidance accuracy, which is better than path 1.

5.3 Experiment II: dynamic reprogramming capability

This experiment is designed to test the path planning ability of two neural networks for dynamic target points and is used to verify whether the dynamic replanning ability of RPP-LSTM algorithm is better.

The two networks are applied to the dynamic target point environment for path planning. In this scenario, the starting coordinates of the UAV are (0,3,0.6), the starting coordinates of the target point are (20,20,0.6), and the target point moves in a straight line along a certain direction. The trained FNN and LSTM network are used for UAV path planning respectively, and the results are shown in Fig. 10.

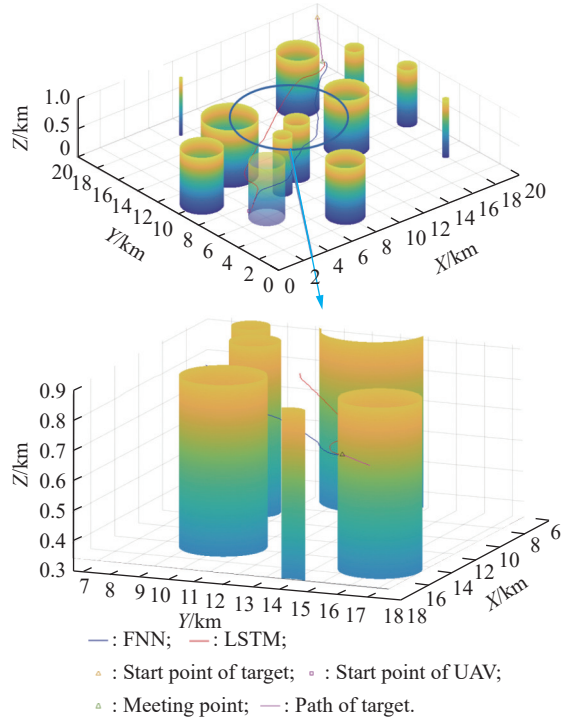

Fig. 10 Dynamic path planning

Assuming that the FNN-based UAV path is path 3 and the LSTM-based UAV path is path 4 in Fig. 10, it can be seen that for the same motion target, the path planned by the FNN is completely different from the path planned by the LSTM network. However, both paths can successfully complete the task and reach the target point, and there is no significant difference in advantages and disadvantages with the paths planned in the static scene. It shows that both FNN and LSTM network can accomplish the dynamic target point path planning task and have better dynamic target point path planning ability.

Based on the data in Table 4, in terms of path length, the two paths are of the same length with no obvious advantages and disadvantages, but there is still a certain difference in the variance of track yaw angle between them. The variance of the track yaw angle of path 3 is still larger than the variance of the track yaw angle difference of path 4, which indicates that the FNN still has the problem of large UAV flight angle change and high maneuverability requirement in dynamic target point planning.

Table 4 Path data comparison 2

Evaluation indicator	FNN	LSTM
Path length/km	21.24	21.24
Trajectory yaw angle variance	269.79	257.85

In terms of the minimum distance between the path and the obstacle, the two paths are not compared because they have different trends.

5.4 Experiment III: robustness testing

This experiment is designed to test the path planning ability of LSTM networks and FNN in unfamiliar environments, and is used to investigate whether the RPP-LSTM algorithm have better robustness in path planning problems.

Both are applied to an unfamiliar environment for path planning, which is still 12 cylindrical obstacles, but with random values of position and radius. The starting coordinates of the UAV in this environment are the point (0,3,0.6) and the coordinates of the target point are (20,16,0.6). The direction from the starting point to the target point is chosen as the initial velocity direction of the UAV. In the completely unfamiliar environment, the FNN and LSTM network obtained from the training were used for UAV path planning, and the results are shown in Fig. 11.

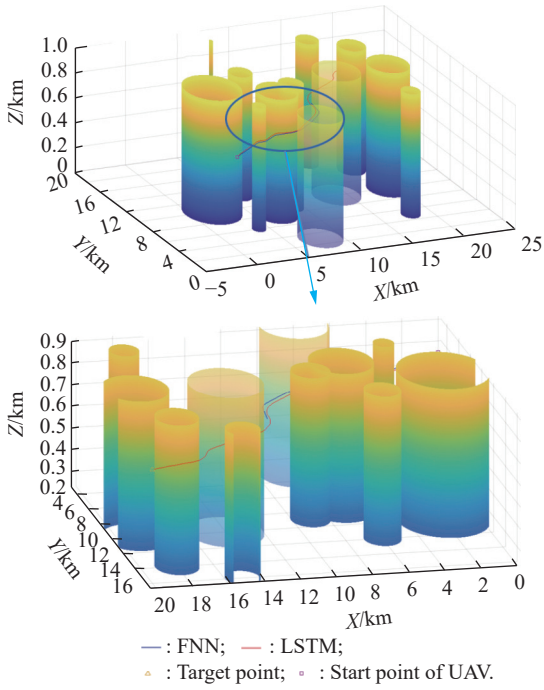

Fig. 11 Path planning in unfamiliar environments

In Fig. 11, it is assumed that the path planned based on FNN network is path 5 and the path planned based on LSTM network is path 6. It can be seen that the FNN network fails to complete the path planning task because it is too close to the obstacle, resulting in the path planning failure, and the LSTM network successfully completes the path planning task and reaches the target point. Two paths of the specific data are shown in Table 5.

Table 5 Path data comparison 3

Evaluation indicator	FNN	LSTM
Path length/km	Undone	25.2
Trajectory yaw angle variance	Undone	363.1
Minimum distance to obstacle/m	Undone	344.7

Although path 5 does not complete the task, the planned part shows that the FNN also has some environmental adaptability. The previous part of path 5 achieves effective obstacle avoidance in unfamiliar environments. However, compared with the LSTM network, the FNN is significantly less adaptable to the unfamiliar environment. The RPP-LSTM can use the environmental information to achieve effective obstacle avoidance and reach the target point with strong robustness.

In addition to the above path metrics evaluation, the path planning speed of the neural network is also considered in the path planning problem. In this paper, 25 paths are planned using two neural networks, and the average single-step time for planning each path is obtained, and the results are shown in Fig. 12.

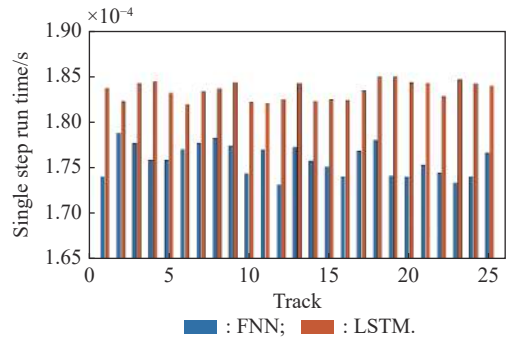

Fig. 12 Aircraft turning angle diagram

5.5 Experiment analyzing

Based on all the above simulation results and analysis, it can be seen that the trained LSTM network can perform UAV path planning very well, and can successfully plan feasible paths in both the original training environment and the unfamiliar environment, and has many advantages compared to the FNN.

The path planning problem is a complex problem that needs to take into account many factors. The FNN only considers the current environment and makes action selection, which can achieve effective obstacle avoidance and reach the target point in the training environment. However, because the action selection is only based on the current environment, the planned path does not consider the previous path planning, and its obstacle avoidance accuracy is not high enough, and it often chooses to avoid obstacles at a long distance when it

finds them, and the planned path length is also longer. Most importantly, its adaptability to complex environments is very poor, and its robustness is poor.

The RPP-LSTM is not only based on current environment information when making action selection, but also relies on historical information, which has memory for past path planning and can effectively use the previous environment and action information for path planning. Due to the participation of historical information, the RPP-LSTM has higher accuracy in obstacle avoidance, can grasp the minimum safety distance well, and does not blindly avoid obstacles at a long distance, and the planned path is often better. The application of historical information also enables the RPP-LSTM to better adapt to the complex environment and improve the robustness to complete the path planning task successfully.

6. Conclusions

In this paper, a deep reinforcement learning algorithm based on LSTM networks is proposed for solving the UAV real-time path planning problem. The UAV real-time path planning problem is a complex multi-step decision problem. In this paper, the DQN algorithm is used as a framework to combine the UAV motion constraints and path planning requirements to construct action selection strategies, reward and punishment functions, and then build Q-value functions based on LSTM networks. The LSTM networks have the ability of "temporal memory", which can effectively solve the shortcomings of the DQN algorithm in dealing with multi-step decision problems. The obtained DQN model is used for neural network training, and the final mature neural network is used for UAV real-time path planning. The experimental simulation results show that the RPP-LSTM is practical and feasible, and the obtained LSTM network has better path planning capability compared with the traditional FNN, and is also significantly better than the FNN in terms of dynamic replanning capability and robustness. Among the deep reinforcement learning methods, there are other algorithms besides the DQN algorithm, and in the next work, the LSTM network will be trained using different deep reinforcement learning methods for comparative study.

References

[1] AZMAT M, KUMMER S. Potential applications of unmanned ground and aerial vehicles to mitigate challenges of transport and logistics-related critical success factors in the humanitarian supply chain. Asian Journal of Sustainability and Social Responsibility, 2020, 5(1): 1–22.

[2] HOSSAIN M S, CHAITANYA K, BHATTACHARYA Y, et al. Integration of smart watch and geographic information system (GIS) to identify post-earthquake critical rescue area part. II. Analytical evaluation of the system. Progress in Disaster Science, 2021, 9: 100132.

[3] KHAN M T R, MUHAMMAD SAAD M, RU Y, et al. Aspects of unmanned aerial vehicles path planning: overview and applications. International Journal of Communication Systems, 2021, 34(10): e4827.

[4] YANG C H, TSAI M H, KANG S C, et al. UAV path planning method for digital terrain model reconstruction —a debris fan example. Automation in Construction, 2018, 93: 214–230.

[5] WANG G Q, ZHENG X Y, ZHAO H T, et al. Unmanned aerial vehicles path planning based on deep reinforcement learning. Proc. of the International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, 2019: 81−88.

[6] ZHENG Z, LIU Y, ZHANG X Y. The more obstacle information sharing, the more effective real-time path planning? Knowledge-Based Systems, 2016, 114: 36−46.

[7] STENTZ A. The focused d* algorithm for real-time replanning. Proc. of the International Joint Conference on Artificial Intelligence, 1995: 1652−1659.

[8] CHEN G, LIU D, WANG Y F, et al. Path planning method with obstacle avoidance for manipulators in dynamic environment. International Journal of Advanced Robotic Systems, 2018, 15(6): 1729881418820223.

[9] ZHANG Z, WU J, DAI J Y, et al. A novel real-time penetration path planning algorithm for stealth UAV in 3D complex dynamic environment. IEEE Access, 2020, 8: 122757–122771.

[10] HUANG H, HUANG P, ZHONG S, et al. Dynamic path planning based on improved D algorithms of Gaode map. Proc. of the IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference, 2019: 15−17.

[11] LIKHACHEV M, KOENIG S. A generalized framework for lifelong planning A* search. Proc. of the International Conference on Automated Planning and Scheduling, 2005: 5–10.

[12] OGATA K. A generic approach on how to formally specify and model check path finding algorithms: Dijkstra, A* and LPA. International Journal of Software Engineering and Knowledge Engineering, 2020, 30(10): 1481–1523.

[13] LIM J, OREN S, PANAGIOTIS T. Class-ordered LPA*: an incremental-search algorithm for weighted colored graphs. Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2021: 6907−6913.

[14] SVEN K, LIKHACHEV M. D* lite. Proc. of the 18th National Conference on Artificial Intelligence, 2002: 476–483.

[15] XIE K L, QIANG J, YANG H T. Research and optimization of d-start lite algorithm in track planning. IEEE Access, 2020, 8: 161920–161928.

[16] ZHU X H, YAN B, YUE Y. Path planning and collision avoidance in unknown environments for USVs based on an improved D* Lite. Applied Sciences, 2021, 11(17): 7863.

[17] LI J K, LIU Y. Deep reinforcement learning based adaptive real-time path planning for UAV. Proc. of the 8th International Conference on Dependable Systems and Their Applications, 2021: 522−530.

[18] HU Z J, GAO X G, WAN K F, et al. Relevant experience learning: a deep reinforcement learning method for UAV autonomous motion planning in complex unknown environments. Chinese Journal of Aeronautics, 2021, 34(12): 187–204.

[19] CANDELI A, DE TOMMASI G, LUI D G, et al. A deep deterministic policy gradient learning approach to missile autopilot design. IEEE Access, 2022, 10: 19685–19696.

[20] XIANG X C, FOO S. Recent advances in deep reinforcement learning applications for solving partially observable

markov decision processes (POMDP) problems: Part 1—fundamentals and applications in games, robotics and natural language processing. Machine Learning and Knowledge Extraction, 2021, 3(3): 554–581.
[21] YANG S M, YOO S Y, JEONG O R. DeNERT-KG: named entity and relation extraction model using DQN, knowledge graph, and BERT. Applied Sciences, 2020, 10(18): 6429.
[22] LE N, RATHOUR V S, YAMAZAKI K, et al. Deep reinforcement learning in computer vision: a comprehensive survey. Artificial Intelligence Review, 2022, 55: 2733–2819.
[23] RAHMAN S, SARKER S, HAQUE A K M, et al. Deep reinforcement learning: a new frontier in computer vision research. AHAD M A R, INOUE A, ed. Vision, sensing and analytics: integrative approaches. Cham: Springer International Publishing, 2021.
[24] AZAR A T, KOUBAA A, ALI MOHAMED N, et al. Drone deep reinforcement learning: a review. Electronics, 2021, 10(9): 999.
[25] TAI L, PAOLO G, LIU M. Virtual-to-real deep reinforcement learning: continuous control of mobile robots for mapless navigation. Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2017: 24−28.
[26] VENTURINI F, MASON F, PASE F, et al. Distributed reinforcement learning for flexible UAV swarm control with transfer learning capabilities. Proc. of the 6th ACM Workshop on Micro Aerial Vehicle Networks, Systems, and Applications, 2020: 1−6.
[27] VENTURINI F, MASON F, PASE F, et al. Distributed reinforcement learning for flexible and efficient uav swarm control. IEEE Trans. on Cognitive Communications and Networking, 2021, 7(3): 955–969.
[28] YAN C, XIANG X J, WANG C. Towards real-time path planning through deep reinforcement learning for a UAV in dynamic environments. Journal of Intelligent & Robotic Systems, 2020, 98(2): 297–309.
[29] CHEN X, AI Y D. Multi-UAV path planning based on improved neural network. Proc. of the Chinese Control and Decision Conference, 2018: 9−11.
[30] GUO N, LI C H, GAO T T, et al. A fusion method of local path planning for mobile robots based on LSTM neural network and reinforcement learning. Mathematical Problems in Engineering, 2021, 2021(10): 5524232.
[31] GUO N, LI C H, WANG D, et al. Local path planning of mobile robot based on long short-term memory neural network. Automatic Control and Computer Sciences, 2021, 55(1): 53–65.

Biographies

ZHANG Jiandong was born in 1974. He received both his M.S. and Ph.D. degrees in system engineering from Northwestern Polytechnical University. He is an associate professor at the Department of System and Control Engineering in Northwestern Polytechnical University, China. His research interests include modeling simulation and effectiveness evaluation of complex systems, development and design of integrated avionics system, and system measurement & test technologies.
E-mail: jdzhang@nwpu.edu.cn

GUO Yukun was born in 1999. He received his B.S. degree in detection, guidance and control technology from Northwestern Polytechnical University in Xi'an, China. He is currently working toward his M.S. degree in electronic science and technology from the School of Electronics and Information Technology at Northwestern Polytechnical University. His research interests include unmanned aerial vehicle path planning and deep reinforcement learning.
E-mail: 2020202124@mail.nwpu.edu.cn

ZHENG Lihui was born in 1988. He received his B.S. degree in fire command and control engineering from the Naval Aviation University in Yantai, China. He is currently pursuing his M.S. degree in the School of Electronics and Information Technology at Northwestern Polytechnical University. His research interests include advanced firepower and command and control theory, mission planning and combat flight software, artificial intelligence and multiunmanned system mission decision technology.
E-mail: lihuizheng@mail.nwpu.edu.cn

YANG Qiming was born in 1988. He received his M.S. degree from Northwestern Polytechnical University (NPU), Xi'an, China in 2013. He was awarded with a Ph.D. degree in electronic science and technology in 2020. He is an assistant researcher of NPU. His main research interests include artificial intelligence and its application on control and decision of unmanned aerial vehicle.
E-mail: yangqm@nwpu.edu.cn

SHI Guoqing was born in 1974. He received his M.S. and Ph.D. degrees in system engineering from Northwestern Polytechnical University. He is an associate professor at the Department of System and Control Engineering in Northwestern Polytechnical University, China. His research interests include integrated avionics system measurement & test technologies, development and design of embedded real-time systems, modeling simulation and effectiveness evaluation of complex systems.
E-mail: shiguoqing@nwpu.edu.cn

WU Yong was born in 1964. He received his B.S. degree in aeronautical fire control and M.S. degree in fire control from Northwestern Polytechnical University. He is a professor in the Department of Systems and Control Engineering, Northwestern Polytechnical University, China. His research interests include avionics integrated systems and simulation techniques, complex systems modeling, and simulation and effectiveness assessment.
E-mail: yongwu@nwpu.edu.cn