drones

MDPI

Article

Research on UAV Trajectory Planning Algorithm Based on Adaptive Potential Field

Mingzhi Shao [1,2], Xin Liu [1,2,*], Changshi Xiao [1,2,3], Tengwen Zhang [1,2] and Haiwen Yuan [4]

[1] School of Ship and Port Engineering, Shandong Jiaotong University, Weihai 264209, China; 23226021@stu.sdjtu.edu.cn (M.S.); 23226029@stu.sdjtu.edu.cn (T.Z.)
[2] Weihai Institute of Marine Information Science and Technology, Weihai 264200, China
[3] School of Shipping, Wuhan University of Technology, Wuhan 430070, China
[4] School of Electrical and Information Technology, Wuhan Institute of Technology, Wuhan 430205, China; hw_yuan@whut.edu.cn
* Correspondence: 201061@sdjtu.edu.cn; Tel.: +86-133-6115-8427

Abstract: For multi-obstacle complex scenarios, the traditional artificial potential field method suffers from the defects of potential field imbalance, its capability to easily fall into the local minima, and encounter unreachable targets in complex navigation environments. Therefore, this paper proposes a three-dimensional adaptive potential field algorithm (SAPF) based on multi-agent reinforcement learning. First, in this paper, the gravitational function in the artificial potential field (APF) is modified to weaken the gravitational effect on the UAV in the region far away from the target point in order to reduce the risk of collision between the UAV and the obstacles during the moving process. Second, in the region close to the target point, this paper improves the artificial potential field function to ensure that the UAV can reach the target point smoothly and realize path convergence by considering the relative distance between the UAV's current position and the target point. Finally, for the characteristics of UAV trajectory planning, a 3D state space is designed based on the 3D coordinates of the UAV, the distance between the UAV and the nearest obstacle, and the distance between the UAV and the target point; an action space is designed based on the displacement increment of the UAV in the three coordinate axes; and the specific formulas for collision penalties and path optimization rewards are re-designed, which effectively avoids the UAV from entering the local minimal points. The experimental results show that the artificial potential field method designed with reinforcement learning can plan shorter paths and exhibit better planning results. In addition, the method is more adaptable in complex scenes and has better anti-interference.

Keywords: unmanned aircraft; trajectory planning; artificial potential field; reinforcement learning

check for updates

Academic Editor: Chao Huang

Received: 1 January 2025
Revised: 16 January 2025
Accepted: 20 January 2025
Published: 21 January 2025

Citation: Shao, M.; Liu, X.; Xiao, C.; Zhang, T.; Yuan, H. Research on UAV Trajectory Planning Algorithm Based on Adaptive Potential Field. *Drones* **2025**, *9*, 79. https://doi.org/10.3390/drones9020079

Copyright: © 2025 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

UAV technology is increasingly used in many fields, such as military reconnaissance [1], disaster rescue [2], logistics and distribution [3], and agricultural surveys [4], and has become an indispensable tool in modern society. Among them, UAV trajectory planning, as a core technology, is crucial for the mission execution efficiency, safety, and autonomous flight capability of UAVs in complex environments. An in-depth study of UAV trajectory planning to realize obstacle avoidance and ensure optimal trajectory is of great significance to promote the development of UAV technology and expand its applications.

UAV trajectory planning algorithms can be divided into two categories: traditional algorithms and intelligent optimization algorithms. Traditional algorithms include Dijkstra [5], A-star [6], the simulated annealing algorithm [7], and the artificial potential field method [8]; intelligent optimization algorithms include the genetic algorithm [9], ant colony algorithm [10], particle swarm algorithm [11], and artificial bee colony algorithm [12]. The artificial potential field method is widely used in UAV trajectory planning due to its advantages of high real-time performance and small computational volume. However, the artificial potential field method has problems, such as the difficult design of potential field function and parameter dependence on empirical adjustment. To solve these problems, researchers have made various improvements; Tang Yuyang [13] proposed using the Bi-A*PF algorithm, which optimizes the heuristic function by adopting a two-way search mechanism and setting the weight coefficients, and introduces an adaptive step strategy, which dynamically adjusts the bi-directional A* search step using the adjustment factor to achieve the comprehensive consideration of global path planning and real-time obstacle avoidance. Cao Xiaoyi [14] proposed a virtual control network-based UAV formation control algorithm, which realizes the real-time obstacle avoidance of UAVs and can return to the original path after successful obstacle avoidance by introducing the anchor point gravity as well as the target point gravity to improve the artificial potential field method. Bin Xian [15] proposed a method combining offline MPC global planning with the online improvement of the local planning of the artificial potential field method, which deals with the local minimum problem of the traditional artificial potential field method by introducing the regulating force and introducing the relative distance between the target and the UAV into the repulsive force function, which makes the UAV cluster have good dynamic obstacle avoidance capability. Hongzhen Guo [16–18] proposed a distributed event-triggered (ET) collision avoidance formation controller, which improved the robustness of QUAVs by employing the parameter adaptive method (PAM) and a dynamic ET mechanism to release the network load, realizing effective collision avoidance for UAVs.

With the development of artificial intelligence technology, reinforcement learning algorithms show the advantages of high environmental adaptability and flexibility in path planning, which can autonomously learn and optimize the complex decision-making process and adjust the strategy in real time to cope with the obstacles and risks. Lingyu Li [19] proposed a deep reinforcement learning (DRL)-based path planning strategy, which improves the action space of the deep Q-network using the artificial potential field and the reward function to achieve dynamic obstacle avoidance for ships, but this method cannot clearly explain the decision-making process and has a long training time and a high demand for data volume. Zixiang Wang [20] combined a deep Q network (DQN) with a proximal policy optimization (PPO) model to improve the robot's ability to navigate, but since the action space of the deep Q network (DQN) is limited to a discrete form, there is potential for further improvement in the quality of its path planning. In addition, with the continuous development of reinforcement learning technology, how to perform efficient path planning for UAV swarms has also become a hot research topic for scholars. Puente-Castro [21] developed a system based on reinforcement learning, which is capable of calculating the optimal flight paths for UAV swarms and can be used for field exploration, but this method is prone to falling into local optimal solutions, which means that the UAVs may find a flight path that appears to be optimal in a local area but is not actually globally optimal. Fan X [22] proposed a probability of containment (POC)-based search trajectory planning method for UAV swarms, which constructs an objective function for search trajectory optimization by introducing a probabilistic map of the target distribution in the a priori information and improves the differential evolution algorithm using the adaptive mutation operator to achieve a UAV swarm's full-coverage search of the mission area and fast search in limited

time, but this method is more complicated and it becomes more difficult to calculate the optimal flight path quickly in practical applications. To address these issues, reinforcement learning algorithms have been proposed to directly learn the potential field function of an artificial potential field. This approach abandons the tedious process of relying on manual settings and the adjustment of potential field parameters and adopts an end-to-end adaptive function learning method to learn and adapt the optimal repulsive potential field directly from the environmental data, which significantly improves the adaptability and generalization ability of the algorithm. At the same time, the powerful nonlinear fitting ability of neural networks is used to not only learn the specific potential field function but also implicitly optimize the parameter configuration in the potential field. However, there are two problems with the traditional artificial potential field algorithm: (1) in the region far away from the target point, the gravitational force is significantly enhanced while the repulsive force is almost negligible, and this force contrast may lead to the difficulty of the UAV to avoid obstacles during its autonomous flight, thus increasing the risk of collision; (2) when the UAV approaches the target point, the repulsive force will be significantly enhanced, while the influence of the gravitational force will be relatively weakened. This change in force may cause the UAV to be strongly repelled when approaching the target, thus making it difficult to stop securely at the target point. (3) The UAV in the traditional potential field algorithm is prone to enter the local minima.

Therefore, this paper proposes an SAPF algorithm based on multi-agent body reinforcement learning. In the region far away from the target point, the gravitational force is modified to weaken the gravitational effect on the points farther away from the target point, thus reducing the risk of collision; in the region closest to the target point, the repulsive potential field function is improved by taking into account the relative distance between the end of the trajectory and the target point, so as to ensure that the trajectory can reach the target point smoothly and achieve convergence. Finally, the UAV is prevented from entering the local minima by improving the potential field and appropriate reward and punishment mechanisms.

2. Study of Constraints on UAV Trajectory Planning

In the process of Unmanned Aerial Vehicle (UAV) trajectory planning, the resulting trajectory must strictly fit the intrinsic properties of the UAV to ensure that it can effectively track and execute the trajectory in real application scenarios. Therefore, the trajectory planning algorithm needs to follow a series of specific constraints, which cover the length of the trajectory, the global curvature of the trajectory, and the maximum curvature of the trajectory.

2.1. UAV Kinematic Constraints

(1) Trajectory length constraints.

Length is an intuitive and important criterion for assessing the merits of a route. The shorter the route, subject to all constraints, the less fuel the UAV will consume and the less time it will take to complete the entire route while maintaining a constant average flight speed. The flight path length L is defined as follows:

$$L = \sum_{w=2}^{W} |P_w - P_{w-1}| \tag{1}$$

where $|P_w - P_{w-1}| = \sqrt{(x_w - x_{w-1})^2 + (y_w - y_{w-1})^2 + (z_w - z_{w-1})^2}$ is the Euclidean distance between two neighboring track points in the *W-1* segment of the track.

(2) Trajectory global curvature constraints.

The global curvature of the route, *GS*, characterizes the smoothness of the route trajectory the smaller the *GS* is, the smoother the route is, and the planned route is easier to track using the UAV, as shown in the following equation:

$$GS = \frac{1}{W-2}\sum_{w=2}^{W-1}(\mu_1|\psi_w - \psi_{w-1}| + \mu_2|\gamma_w - \gamma_{w-1}|) \tag{2}$$

where, ψ_w and γ_w denote the deflection angle and inclination angle of the *W*th segment of the route, respectively, μ_1, $\mu_2 \in [0, 1]$ and $\mu_1 + \mu_2 = 1$, and *W* is the number of waypoints that constitute the whole route.

(3) Trajectory Maximum Curvature Constraints.

The maximum curvature of route LS reflects the trackability of the whole route; the fact that *LS* is too large means that there is a section of the route with a sudden change in the trajectory angle. At this time, the UAV needs to carry out a high-load maneuver response, and a high-quality route should minimize the occurrence of such a situation, i.e., reducing the *LS*. The formula is as follows:

$$LS = \max_{w\in\{2,\ldots,W-1\}}(\mu_1|\psi_w - \psi_{w-1}| + \mu_2|\gamma_w - \gamma_{w-1}|) \tag{3}$$

(4) Maximum turning angle constraint.

As shown in Figure 1, it can be assumed that the actual turning angle of the UAV is α and the maximum turning angle under the constraints is α_{max}. The constraints are as follows:

$$\begin{cases} \alpha_i = \dfrac{A_i^T \cdot A_{i+1}}{\| A \| \cdot \| A_{i+1} \|} \\ -\alpha_{\max} \le \alpha_i \le \alpha_{\max} \end{cases} \tag{4}$$

where $A_w = (x_i - x_{i-1}, y_i - y_{i-1})$ is the projection vector of the *W*th segment voyage on the horizontal ground; A_w^T is the transpose matrix of A_w; α_{max} is the maximum turning angle; and α_w is the turning angle of the *W*th segment.

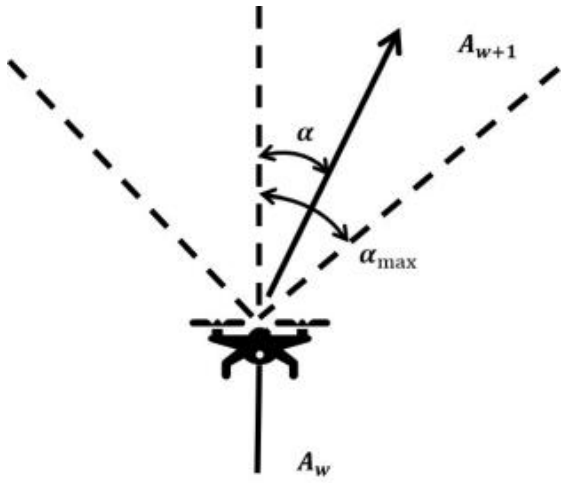


Figure 1. Turning angle constraint diagram.

(5) Flight altitude constraints.

Drones must be strictly controlled while flying within a specified altitude interval $[H_{min}, H_{max}]$. The formula is given as follows:

$$H_{min} < H < H_{max} \tag{5}$$

2.2. Modeling of the UAV Flight Environment

Assuming that A(x, y) is any point in the planning space, the expression for the two-dimensional planning space is as follows:

$$\{(x,y)|X_{\min} \le x \le X_{\max}, Y_{\min} \le y \le Y_{\max}\} \tag{6}$$

where $X_{\min}$, $X_{\max}, Y_{\min}, Y_{\max}$ represent the boundary value of the space.

This algorithm is based on continuous environment and action space construction, which allows the intelligent body to localize at any position within the environment and move in any direction in compliance with the action constraints. In order to simplify the modeling process, the vector diagram method is used. As shown in Figure 2, air defense positions (represented as circles with a radius equal to the maximum striking distance) and early warning detection radars (represented as circles with a radius equal to the maximum detection distance) in the plane are considered impenetrable circular obstacles. Irregularly shaped obstacles are similarly transformed into circular obstacles for processing by expanding outward and taking their smallest outer circle as the equivalent radius to ensure that the planned trajectory maintains a certain safety distance from potential threat sources.

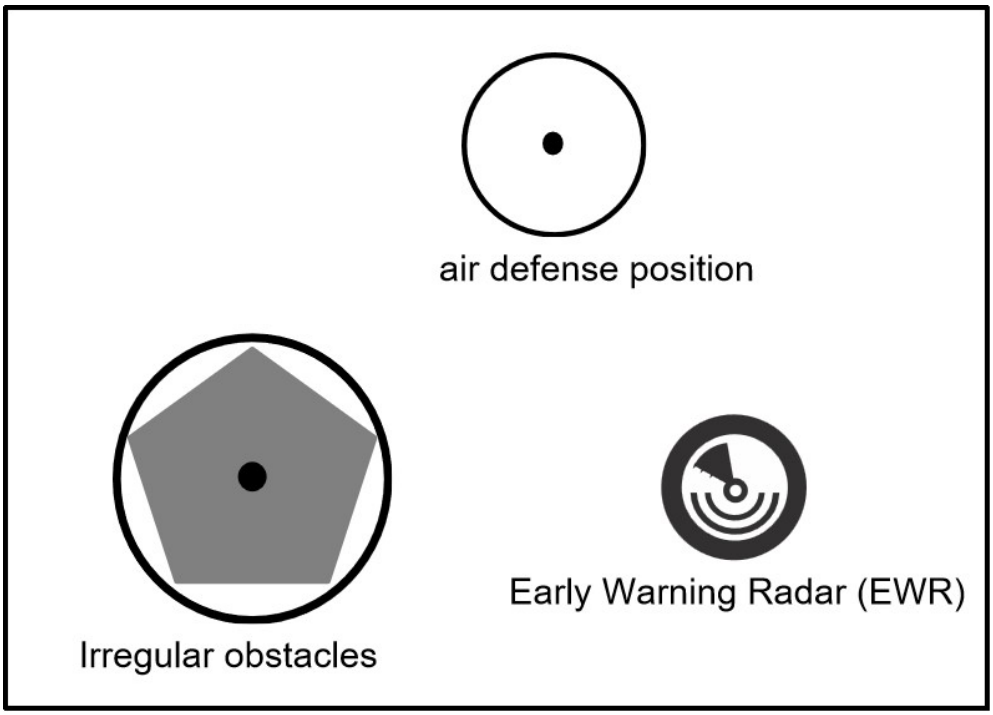


Figure 2. Vector environmental maps.

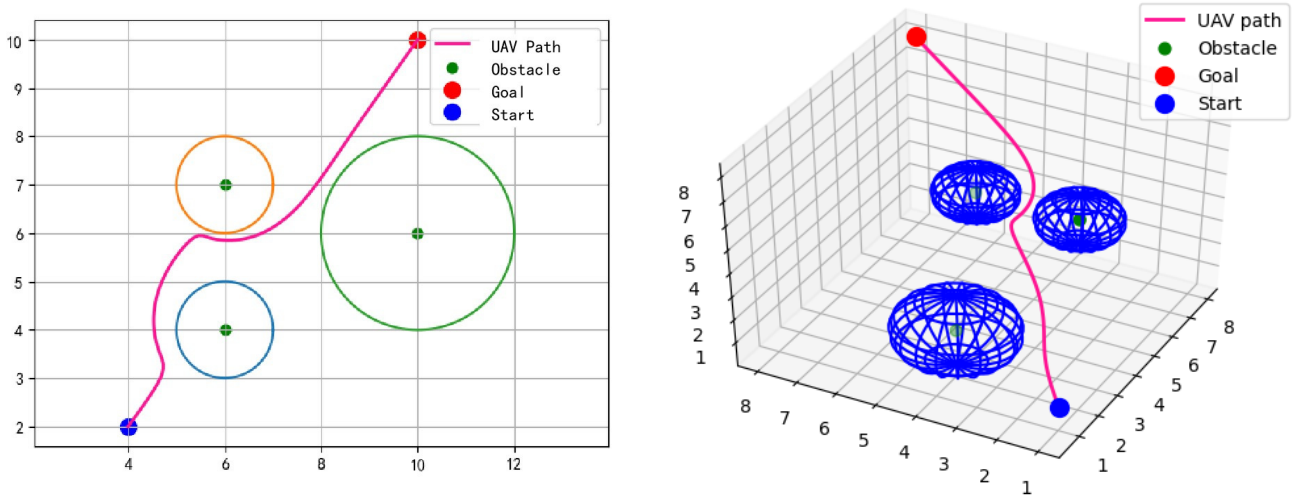
3. Principles of Artificial Potential Field Algorithms

This chapter introduces the principle of the artificial potential field algorithm, gravitational potential field, and repulsive potential field and improves it for the shortcomings of the artificial potential field.

3.1. Artificial Potential Field Method

APF is a commonly used path planning algorithm, which is mainly used for UAV path planning and obstacle avoidance, and the specific effect is shown in Figure 3. Its core concept lies in constructing a spatial potential field model in which any position is subject to the action of the potential field. The negative gradient direction of the potential

field indicates the movement orientation in path planning, i.e., the path extends along the direction of the decreasing value of the potential function. The essence of the path planning problem lies in finding a path from a start point to a goal point while satisfying all the predefined conditions. In the artificial potential field method, the goal point is set as the center of the gravitational field, and in an obstacle-free environment, the path gravitationally converges to the goal point. On the contrary, obstacles in the environment are considered the center of the repulsive field, which serves to ensure that the planned path avoids collision with the obstacles.



(**a**) The 2D schematic (**b**) The 3D schematic

Figure 3. APF path planning schematic.

Gravitational and Repulsive Potential Fields

In an artificial potential field algorithm, the target point is the only source of the gravitational field, which exerts a gravitational force on the points along the path, driving these points to converge towards the target point. A common method of calculating the gravitational field is the following:

$$U_{att}(P) = \frac{1}{2}\xi \parallel P - P_g \parallel^2 \tag{7}$$

In this algorithm, there exists a gravitational gain coefficient, ξ, and the specific coordinates of the target point P_g. According to the gravitational field model, when the Euclidean distance of a point P from the target point P_g in space increases, the gravitational force on the point increases accordingly. The specific shape of the gravitational field model is illustrated in Figure 4 [23], where the gravitational point is located at (5, 5), and it can be observed that the shape of the gravitational field bends as the gravitational coefficient increases. Guided by the gravitational field, any point in space moves in the direction that minimizes the potential energy function, which points exactly to the position of the target point, and the gravitational force at the end of the trajectory is as follows:

$$F_{att}(P) = -\nabla U_{att}(P) = \xi(P_g - P) \tag{8}$$

For each obstacle in space, a repulsive force field is introduced to adjust and optimize the original path. As the end of the path approaches the surface of the obstacle, the repulsive field is significantly enhanced, which is inversely proportional to the distance, until it reaches a threshold beyond which the repulsive force has a negligible effect, and

the path can then be maintained in a straight line towards the target point. Such a design aims to shorten the path length. Based on the above principle, a typical repulsive force field calculation method can be formulated as follows:

$$U_{rep}(\mathbf{P}) = \begin{cases} \frac{1}{2}\eta\left(\frac{1}{\|\boldsymbol{P}-\boldsymbol{P}_{obs}\|} - \frac{1}{\rho_0}\right)^2, & \| \boldsymbol{P}-\boldsymbol{P}_{obs} \| \le \rho_0 \\ 0, & \| \boldsymbol{P}-\boldsymbol{P}_{obs} \| > \rho_0 \end{cases} \tag{9}$$

where η is the repulsive force gain coefficient, $\mathbf{P}_{obs}$ is the obstacle center coordinate, ρ_0 is the repulsive force range, as shown by the repulsive force field model in Figure 5 [23], and the position of the repulsive point is (5, 5). The repulsive force field bulges in the vicinity of the repulsive point and increases with the increase in the repulsive force gain coefficient, and when the trajectory is close to the vicinity of the repulsive point, the opposite direction of the gradient drives the trajectory away from the repulsive point to complete the obstacle avoidance function. According to the formula of the repulsive force field, the repulsive force field in the neighborhood of the repulsive force point is infinite, and the repulsive force field in the neighborhood is set as a constant in the figure for the convenience of observation. Then, the repulsive force at the end of the trajectory is as follows:

$$F_{rep}(P) = \begin{cases} \eta\left(\frac{1}{\|\boldsymbol{P}-\boldsymbol{P}_{obs}\|} - \frac{1}{\rho_0}\right)\frac{1}{\|\boldsymbol{P}-\boldsymbol{P}_{obs}\|^2}\nabla \| P-\boldsymbol{P}_{obs} \|, & \| \boldsymbol{P}-\boldsymbol{P}_{obs} \| \le \rho_0 \\ 0, & \| \boldsymbol{P}-\boldsymbol{P}_{obs} \| > \rho_0 \end{cases} \tag{10}$$

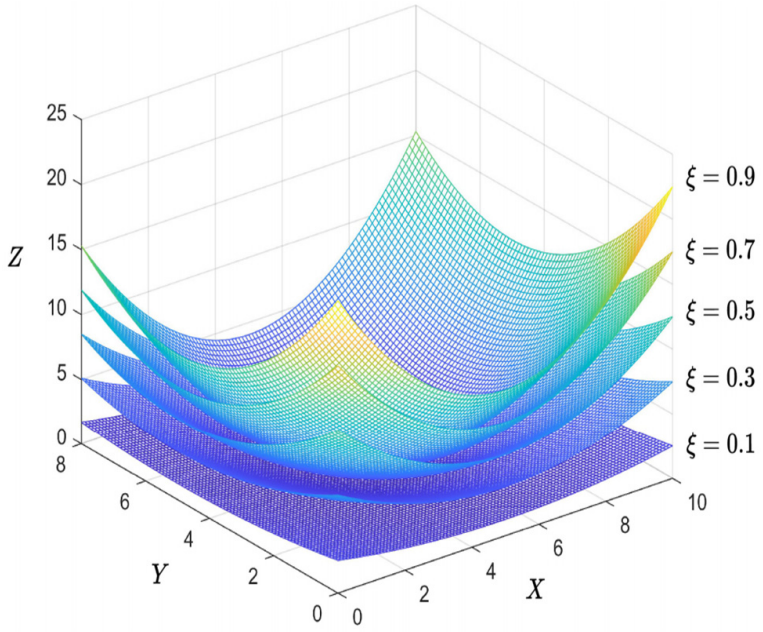


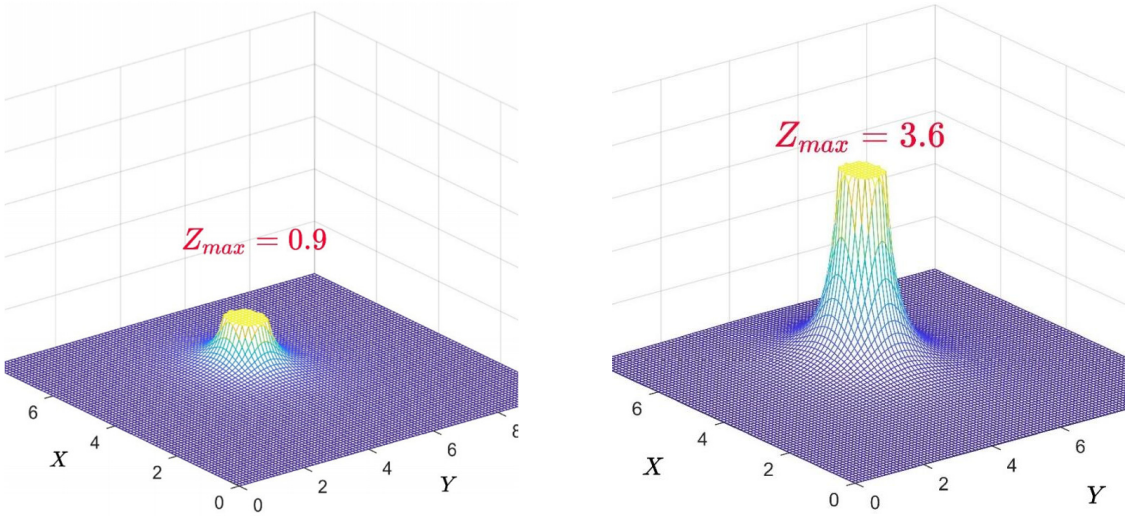
Figure 4. Schematic modeling of the gravitational field for different gravitational gain coefficients.

3.2. Improved Artificial Potential Field Method

Currently, traditional artificial potential field methods face the following challenges:

(1) In regions far from the target point, the gravitational force is significantly enhanced compared to the negligible effect of the repulsive force, which may result in the UAV colliding with obstacles during its movement.

(2) In regions close to the target point, the repulsive force is significantly enhanced, compared to the negligible effect of the gravitational force, which may result in the UAV failing to reach the target point.



(**a**) $\eta = 0.5$ (**b**) $\eta = 2$

Figure 5. Schematic of the gravitational field model with different repulsive gain coefficients.

A solution is provided as follows:

To address the first problem, this study weakened the gravitational effect on points farther away from the target point by correcting the gravitational force; then the corrected gravitational field equation can be calculated as follows:

$$U_{att}(\mathbf{P}) = \begin{cases} \frac{1}{2}\xi \| \boldsymbol{P} - \boldsymbol{P}_g \|^2, & \| \boldsymbol{P} - \boldsymbol{P}_{obs} \| \le d^*_{goal} \\ d^*_{goal}\xi \| \boldsymbol{P} - \boldsymbol{P}_g \| - \frac{1}{2}\xi {d^*_{goal}}^2, & \| \boldsymbol{P} - \boldsymbol{P}_{obs} \| > d^*_{goal} \end{cases} \tag{11}$$

where d^*_{goal} is the radius of the gravitational field weakening, and the gravitational field is weakened when the distance between the endpoint of the trajectory and the target point is greater than it. At this time, the gravitational force is as follows:

$$F_{att}(\mathbf{P}) = \begin{cases} \xi(\boldsymbol{P}_g - \boldsymbol{P}), & \| \boldsymbol{P} - \boldsymbol{P}_{obs} \| \le d^*_{goal} \\ \frac{d^*_{goal}\xi(\boldsymbol{P}_g - \boldsymbol{P})}{\|\boldsymbol{P} - \boldsymbol{P}_g\|}, & \| \boldsymbol{P} - \boldsymbol{P}_{obs} \| > d^*_{goal} \end{cases} \tag{12}$$

From the above equation, it can be seen that when $\| \mathbf{P} - \mathbf{P}_{obs} \| > d^*_{\text{goal}}$, $\| \mathbf{P} - \mathbf{P}_g \|$ increases and $F_{att}(\mathbf{P})$ decreases, the magnitude of the gravitational force is weakened. Comparing the gravitational field before and after the correction demonstrated in Figure 6, it can be clearly seen that at a distance beyond a certain range from the center of the gravitational point, the corrected gravitational field (green mesh) exhibits weaker strength in comparison with the pre-corrected model (blue mesh), which proves the effectiveness of the correction algorithm.

For the second problem mentioned above, this study improved the repulsive potential field function by considering the relative distance between the end of the trajectory and the target point, where the repulsive field can be tabulated as follows:

$$U_{\text{rep}}(\mathbf{P}) = \begin{cases} \frac{\eta}{2}\left(\frac{1}{\|\boldsymbol{P} - \boldsymbol{P}_{obs}\|} - \frac{1}{\rho_0}\right)^2 \| \boldsymbol{P} - \boldsymbol{P}_g \|^n, & \| \boldsymbol{P} - \boldsymbol{P}_o \| \le \rho_0 \\ 0, & \| \boldsymbol{P} - \boldsymbol{P}_o \| > \rho_0 \end{cases} \tag{13}$$

Then, the modified repulsive force is the following:

$$F_{rep}(P) = \begin{cases} \boldsymbol{F}_{repl}\nabla\| \boldsymbol{P} - \boldsymbol{P}_{obs} \| + \boldsymbol{F}_{rep2}\nabla\| \boldsymbol{P} - \boldsymbol{P}_{g} \|, & \| \boldsymbol{P} - \boldsymbol{P}_{obs} \| \leq \rho_0 \\ 0, & \| \boldsymbol{P} - \boldsymbol{P}_{obs} \| > \rho_0 \end{cases} \tag{14}$$

Observing the modified repulsive force field in Figure 7, it can be noticed that the repulsive force field is tilted relative to the original state; it favors the direction of the target point. Closer to the target point, the strength of the repulsive force field gradually decreases. As a result, the influence of the repulsive force becomes weaker as the trajectory approaches the region between the target point and the obstacle, thus ensuring that the trajectory can reach the target point successfully and achieve convergence.

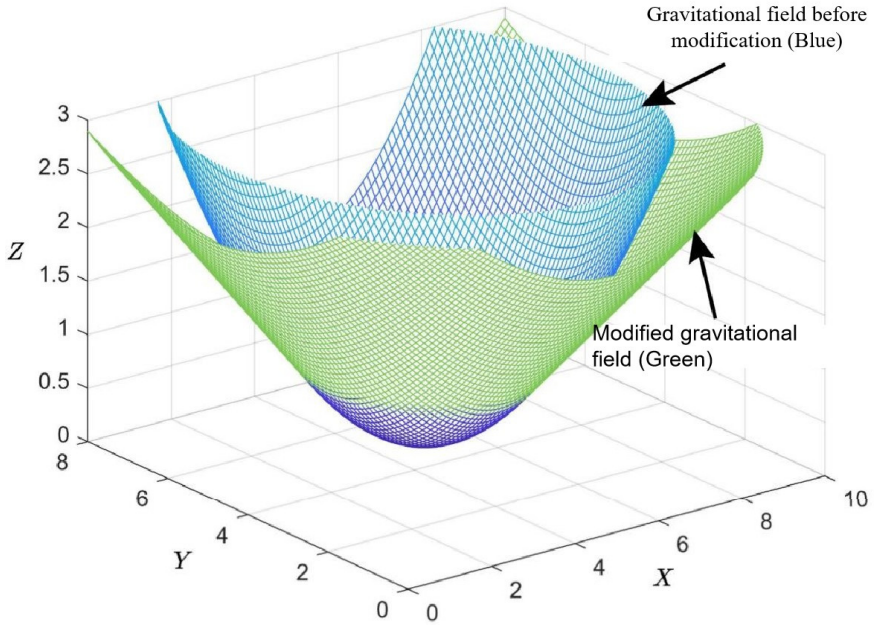


Figure 6. Schematic of the gravitational field model before and after correction.

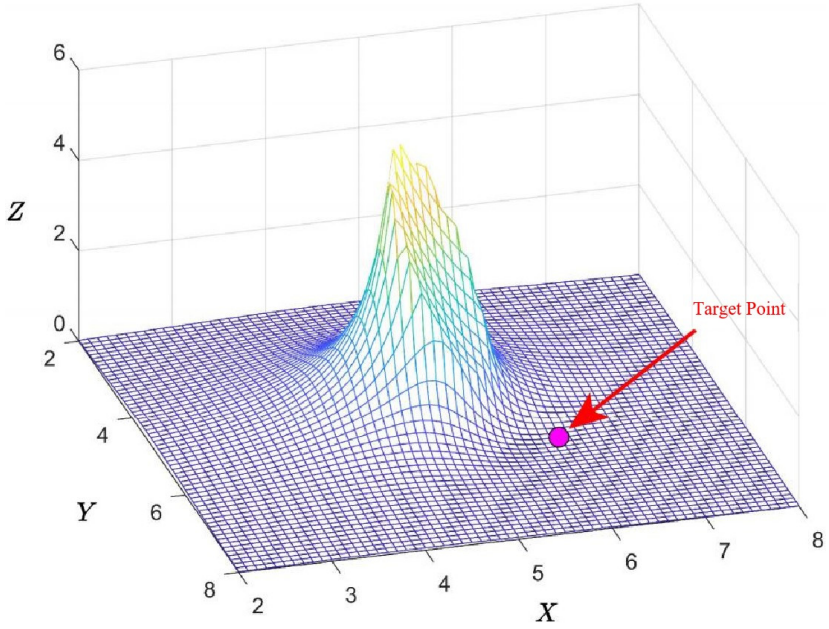


Figure 7. Schematic of the modified repulsive force field model.

3.3. Theory of Deep Deterministic Policy Gradient Algorithms

The deep deterministic policy gradient (DDPG) is a reinforcement learning algorithm that focuses on deterministic policy formulation. Unlike stochastic policy algorithms, DDPG does not learn the probability distribution of an action to be taken in a particular state but directly specifies the specific action to be taken in that state. This property gives the DDPG algorithm the advantage of fast convergence and high learning efficiency. However, a potential drawback of the deterministic strategy is that it is more deficient in environmental exploration compared to the stochastic strategy. To compensate for this shortcoming, it is often necessary to artificially add noise during the data collection phase to enhance the algorithm's exploration capability. Table 1 compares the advantages and disadvantages of the three strategies, PPO, DQN, and DDPG, compared to which PPO performs slightly less well in continuous control tasks, while DQN cannot be directly applied to the continuous action space tasks of UAVs. Therefore, DDPG has significant advantages in UAV path planning, especially in complex dynamic environments. After comparing the effectiveness of several mainstream reinforcement learning algorithms to solve the route planning problem, this study decided to adopt DDPG as the underlying algorithm, and the schematic diagram is shown in Figure 8.

Table 1. Comparison of PPO, DQN, and DDPG.

Metrics	PPO	DQN	DDPG
Algorithm type	Can handle continuous action space but outputs probability distributions with lower control accuracy.	Applicable only to discrete maneuver spaces and cannot directly handle continuous control tasks for UAVs.	Continuous motion space for precise control of UAVs.
Strategic output	Randomized strategy that outputs a probability distribution of actions.	Q-values for discrete actions, unable to generate continuous actions.	Deterministic strategy to directly generate precise action values.
Path smoothness	Paths may not be smooth enough and subject to randomization strategies.	Paths are discrete, may be discontinuous, and are not suitable for complex environments.	Generate smooth and continuous flight paths for complex environments.
High-dimensional state processing capability	Capable of handling high-dimensional state spaces, but may not be as efficient as DDPG in continuous control tasks.	Difficult to efficiently handle high-dimensional state spaces, suitable for low-dimensional discrete tasks.	Capable of efficiently handling high-dimensional state spaces.
Obstacle avoidance effect	Obstacle avoidance is better but may not be as accurate as DDPG in continuous control tasks.	Obstacle avoidance is poor, and it is difficult to handle continuous movements and complex environments.	Excellent performance in dynamic obstacle environments, enabling efficient obstacle avoidance.

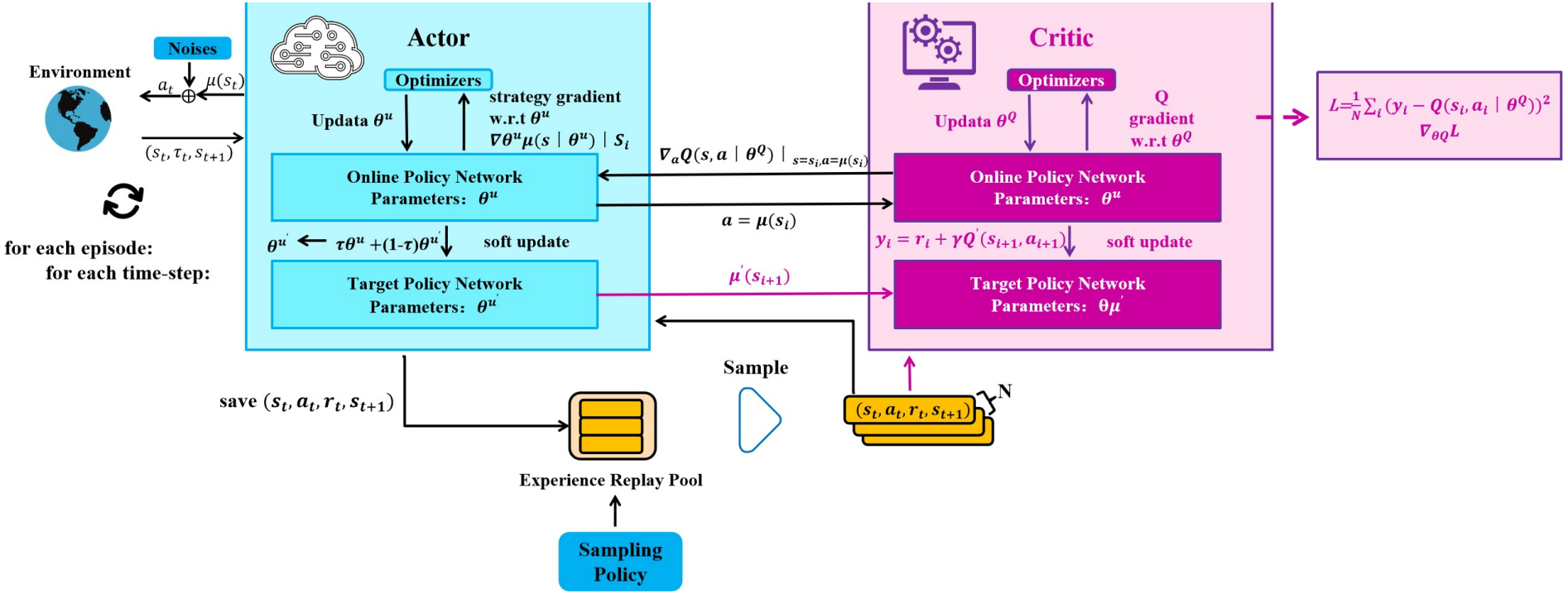


Figure 8. Schematic diagram of DDPG algorithm.

4. Stereo-Adaptive Artificial Potential Field Algorithm Based on Deep Reinforcement Learning

In this paper, obstacles are regarded as forms of reinforcement learning intelligence that teach us to control the size of the artificial potential field through data interacting with UAVs in the environment, which collaborate to generate superimposed potential fields to optimize UAV routes. Combining deep reinforcement learning and artificial potential field algorithms, this paper proposes SAPF. The repulsive potential field of SAPF changes with the position of the UAV, and the local minimum point moves along with it. During the training process, the penalty mechanism is triggered when falling into the local minimum, and this operation effectively avoids the local minimum problem of the traditional potential field algorithm.

4.1. Adaptive Artificial Potential Field Algorithm Framework

In the SAPF algorithmic architecture presented in this paper, the APF plays the role of the planning layer, which is responsible for determining the final planning route. The reinforcement learning model, on the other hand, forms the core of the learning and decision-making layer, which learns how to precisely regulate the size of the repulsive field by analyzing a large amount of data on the interaction between the intelligent body and the environment. When the UAV is at a certain position in 3D space, it acquires information about the relative position of obstacles and the relative position of the target through sensors such as radar. This information is then fed into a network of reinforcement learning strategies corresponding to the obstacles from which the network outputs the repulsive force gain coefficients that should be used for each obstacle in the current state of the environment. This process is repeated for all obstacles at moment t, generating a vector of repulsive gain coefficients, which is passed to the planning layer as a reinforcement learning action. The planning layer uses this information to compute the next waypoint for the UAV and feeds a reward value transferred from the current state to the algorithm, which guides further optimization of the algorithm. In this paper, three basic three-dimensional shapes were used to model obstacles: the sphere, cylinder, and cone, and the schematic diagram of the SAPF algorithm is shown in Figure 9.

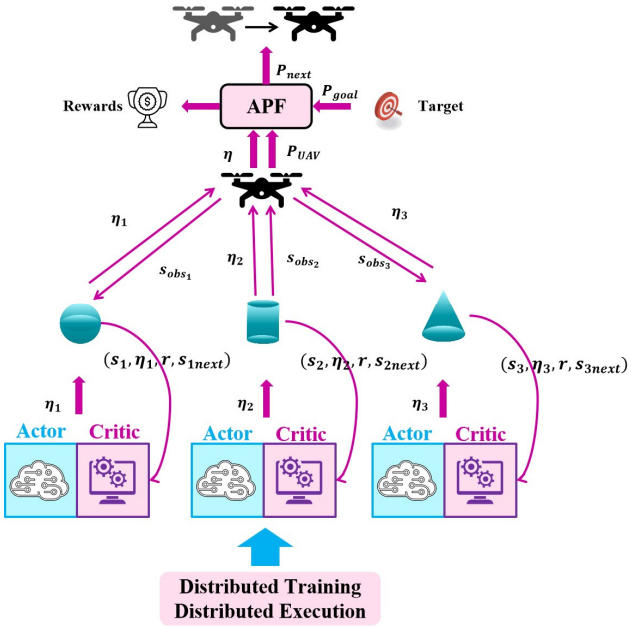


Figure 9. SAPF algorithm framework diagram.

4.2. Decision-Making Process Design

(1) State Space

In the framework of reinforcement learning applied to UAV path planning, the state space forms the basis of the model's policy decisions. In order to fully describe the environmental information of the UAV, the state space designed in this paper is given as follows:

3D coordinates of the UAV: the state variable (x_{UAV}, y_{UAV}, z_{UAV}) represents the 3D position of the UAV at the current moment. These coordinates are used to determine the exact position of the UAV in space and are the basic information for path planning.

Distance between the UAV and the nearest obstacle: the state variable d_{obs} represents the distance between the UAV and the nearest obstacle and is calculated as follows:

$$d_{\mathrm{obs}} = \sqrt{(x_{\mathrm{UAV}} - \Delta x_{\mathrm{obs}})^2 + (y_{\mathrm{UAV}} - \Delta y_{\mathrm{obs}})^2 + (z_{\mathrm{UAV}} - \Delta z_{\mathrm{obs}})^2} \tag{15}$$

Here, (Δx_{obs}, Δy_{obs}, Δz_{obs}) are the 3D coordinates of the nearest obstacle. This variable is used to evaluate the risk of collision between the UAV and the obstacle and is a key input for the obstacle avoidance decision.

The distance between the UAV and the target point: the state variable d_{goal} represents the distance between the UAV and the target point, calculated as follows:

$$d_{\mathrm{goal}} = \sqrt{\left(x_{\mathrm{UAV}} - \Delta x_{\mathrm{goal}}\right)^2 + \left(y_{\mathrm{UAV}} - \Delta y_{\mathrm{goal}}\right)^2 + \left(z_{\mathrm{UAV}} - \Delta z_{\mathrm{goal}}\right)^2} \tag{16}$$

Here, (Δx_{goal}, Δy_{goal}, Δz_{goal}) is the relative position of the target point to the UAV. This variable is used to evaluate the proximity of the UAV to the target point, which is an important basis for path optimization.

With the above state variables, the state space, s, can be represented as follows:

$$s = \left(x_{\mathrm{UAV}}, y_{UAV}, z_{\mathrm{UAV}}, d_{\mathrm{obs}}, d_{\mathrm{goal}}\right) \tag{17}$$

The equation shows the state, s, reinterpreted as information about the position of the UAV in 3D space relative to obstacles and the position of the UAV relative to the target point. Together, this information forms a comprehensive view of the strategy inputs. As shown by equations corresponding one-to-one with the UAV's position in space, there are no duplicate states; thus, the selection of s is reasonable.

(2) Action Space

The action space is a set representing all possible flight actions. The UAV selects an action from this set at each step to interact with the environment (i.e., the flight space) to achieve its path planning goal. The design of the maneuver space should be based on the motion characteristics of the UAV in 3D space, ensuring smooth and dynamically correct motion while keeping the design simple and efficient to meet real-time requirements. In this paper, we used the action space defined as the displacement increment of the UAV in 3D space, denoted as follows:

$$A = [\Delta x, \Delta y, \Delta z] \tag{18}$$

where Δx, Δy, Δz is in the range $[-d_{max}, d_{max}]$ and d_{max} is the maximum displacement.

(3) Reward mechanism

The reward mechanism of reinforcement learning is a crucial factor affecting the learning effect of the algorithm, and the reward value should reflect the key characteristics of the problem to be solved. The reward mechanism in reinforcement learning has a decisive impact on the learning effectiveness of the algorithm, and its design should be closely centered on the core features of the problem to be solved. In order to ensure the stability and efficiency of the learning process, each reward value needs to maintain the

consistency of the numerical range and minimize the variance of the reward in order to avoid drastic fluctuations in the cumulative reward. In the pursuit of high-performance trajectories, it is often required that the trajectories are both short and do not intersect with obstacles while maintaining a certain degree of smoothness. However, given that the underlying APF algorithm already ensures smoothness to a certain extent and usually satisfies the smoothness requirement when the trajectory is short, the reward mechanism was not designed in this study with special consideration of the smoothness constraint. Therefore, the following reward and penalty mechanisms were designed in this paper:

If the next track point P_{t+1} has an intersection with an obstacle, a penalty mechanism is implemented as follows:

$$Reward_1 = -\frac{\parallel P_{t+1} - O_{obs} \parallel}{R_{obs}} - C_o, if \parallel P_{t+1} - O_{obs} \parallel \leq R_{obs} \tag{19}$$

In the formula, this paper defines the intersection point of the trajectory and the obstacle as O_{obs}. For the sphere obstacle, the core point is the center of the sphere; for the cylinder and cone obstacle, the center of its horizontal cross-section circle is taken as the core point. At the same time, this paper sets the radius parameter R_{obs} of the obstacle for the sphere, where the radius is the actual radius of the sphere; for cylinders and cones, the radius of the horizontal cross-section circle is taken as the standard. C_o is a constant.

If the next track point P_{t+1} has no intersection with an obstacle, the reward value should tend to optimize the route length as follows:

$$Reward_2 = \begin{cases} -\frac{\|P_{t+1} - P_g\|}{\|P_0 - P_g\|}, if \parallel P_{t+1} - P_g \parallel > \varepsilon \\ -\frac{\|P_{t+1} - P_g\|}{\|P_0 - P_g\|} + C_1, if \parallel P_{t+1} - P_g \parallel \leq \varepsilon \end{cases} \tag{20}$$

In the formula, $\parallel P_{t+1} - P_g \parallel$ denotes the distance between the next track point and the target point; $\parallel P_0 - P_g \parallel$ denotes the distance from the starting point to the target point; and ε is a constant to determine whether the track point arrives near the target point. Then, the reward function is given as follows:

$$R_t(a_t \mid s_t) = \lambda_1 \cdot Reward_1 + \lambda_2 \cdot Reward_2 \tag{21}$$

Here, λ_1 and λ_2 are weighting factors.

5. Experimental Section

5.1. Experimental Parameters

In the experiments of 3D online trajectory planning, there are training and testing environments, which were trained and tested using the SAPF algorithm in this study, including the Fully Centralized DDPG algorithm (Fully Centralized DDPG), and the Fully Decentralized DDPG (Fully Decentralized DDPG), respectively. The experiments were conducted in an environment equipped with a 64-bit Windows 11 operating system with 16 GB of running memory, NVIDIA GeForce RTX 4060 GPU for accelerated computation, and a 13th-generation Intel Core i7-13700H processor with a clock frequency of 2.40 GHz. The software used was MATLAB 2024a.

The parameter settings for the planning layer in the SAPF algorithm are shown in Table 2; the hyperparameter settings for the learning and decision-making layer are shown in Table 3; and the reward parameter settings are shown in Table 4.

Table 2. SAPF algorithm APF for planning layer parameter settings.

Parameter Symbol	d^*_{goal}(km)	*threshold*	$r_0(km)$	$\dot{\chi}_{max}$
Set Point	5	0.2	5	10°
Parameter Symbol	$\dot{\gamma}_{max}$	γ_{max}	γ_{min}	step size
Set Point	10°	100°	-75°	0.2

Table 3. DDPG reinforcement learning model for hyperparameter settings.

Parameter Symbol	buffer size	net dim	γ	lr	τ
Set Point	10^6	2^8	0.99	$1e^{-3}$	$5e^{-3}$
Parameter Symbol	action noise	batch size	state dim	action dim	η_{max}
Set Point	0.1	128	6	1	3
Parameter Symbol	η_{min}				
Set Point	0.1				

Table 4. Reward function parameter setting.

Parameter Symbol	λ_1	λ_2	C_0	C_1
Set Point	1	1	1	3

5.2. Experimental Environment

The experiment is simulated in a 3D static obstacle environment, where the specific parameters of the obstacles are shown in Tables 5 and 6. The simulation is shown in Figure 10, and the size of the flight environment is (x, y, z) = (10 km, 10 km, 6 km), and the distance between neighboring grid points in the z-direction is 1 km.

Table 5. Spherical obstacle coordinates and shape parameter table.

	Obstacle Ball 1	**Obstacle Ball 2**	**Obstacle Ball 3**
Coordinates (center)	(2, 5, 2)	(4, 3, 2)	(8, 8, 2)
Radius of the ball	2	2	2

Table 6. Table of coordinates and shape parameters of cylindrical and conical obstacles.

	Cylinder (Geometry)	**Cone**
Coordinates (bottom center, km)	(4, 7, 0)	(8, 5, 0)
Bottom radius (km)	1	2
Height (km)	6	4

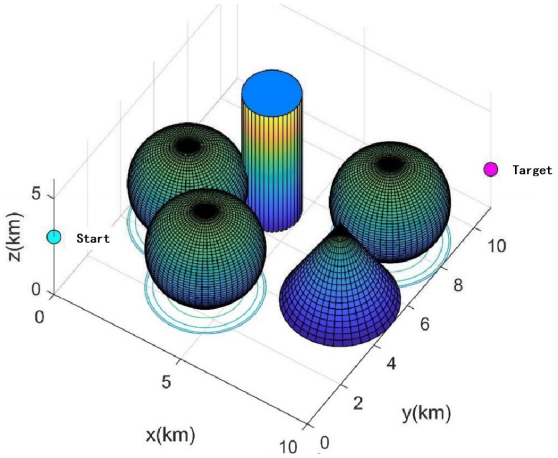


Figure 10. The 3D modeling experimental environment.

5.3. Training Experiments

In the experiment, 1000 rounds were trained in multiple obstacle environments using the Rapidly Exploring Random Trees (RRTs) algorithm, the A* algorithm, and the SAPF algorithm proposed in this paper, and the reward curves are shown in Figure 11.

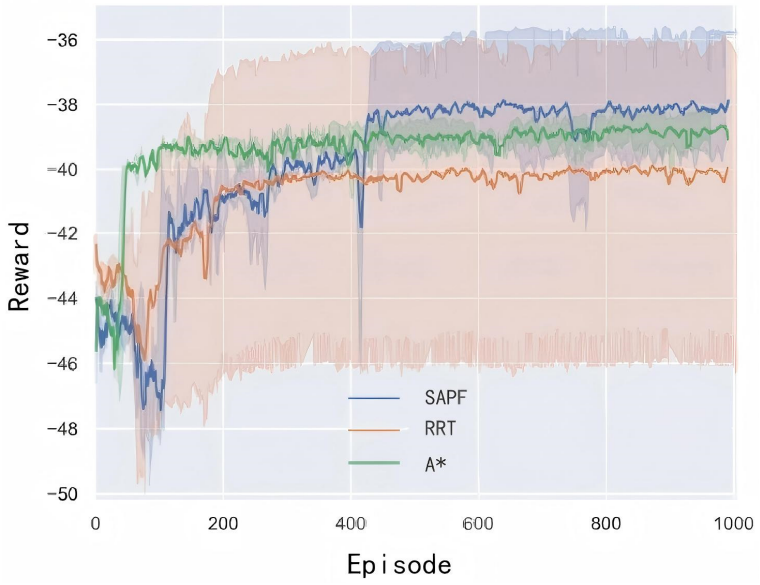


Figure 11. Training reward value change curve.

Convergence analysis: As can be seen from the reward curves in Figure 11, the SAPF algorithm gradually converges during training, indicating that the algorithm is able to learn the optimization strategy stably. In contrast, the reward curves of the RRT algorithm and the A* algorithm may fluctuate more or converge more slowly, indicating that they are less efficient at learning in complex environments.

Comparison of reward values: The SAPF algorithm generates significantly higher reward curve values than the RRT and A* algorithms, indicating that the SAPF algorithm is able to achieve higher cumulative rewards in the path planning task. Higher reward values usually mean better path planning results, such as shorter path lengths, fewer collisions, and higher task completion rates.

Algorithm effectiveness: The convergence and high reward values of the SAPF algorithm demonstrate its effectiveness in complex obstacle environments. Compared with RRT and A* algorithms, the SAPF algorithm is better able to adapt to dynamic and complex environments and generates better path-planning strategies.

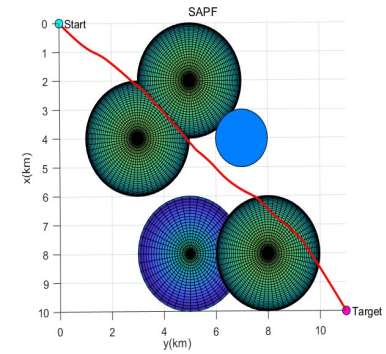
5.4. Static Environment Test Experiment

In order to better reflect its planning effect, this paper compared the SAPF algorithm with the A* algorithm, ACO algorithm, and RRT algorithm, respectively, and the experimental results are shown in Figures 12–14, with the top view of the experimental results on the left, and the three-dimensional image on the right.

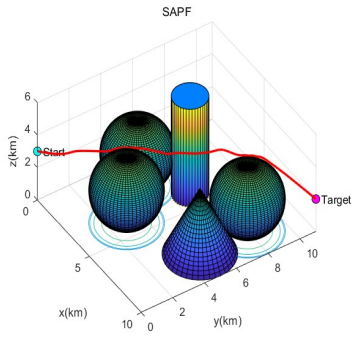
Figure 12 shows the comparison results of the SAPF and A* algorithm; it can be seen that SAPF can reach the target point with a shorter path length, and the path does not need to be smoothed, whereas the path of the A* algorithm is longer and the smoothed path produces intersections with the obstacles, which indicates that UAVs under the A* algorithm are more likely to collide with the obstacles.

Figure 13 shows the comparison results between the SAPF and ACO algorithm; it can be seen that the path of the ACO algorithm is more complex, even though the length of the path is still larger than the SAPF algorithm after the smoothing operation. It proves that

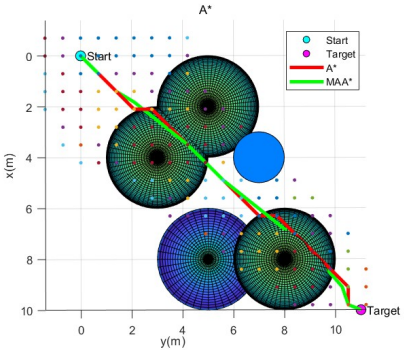
the UAV can reach the target point over a shorter time span, which proves the superiority of the SAPF algorithm.



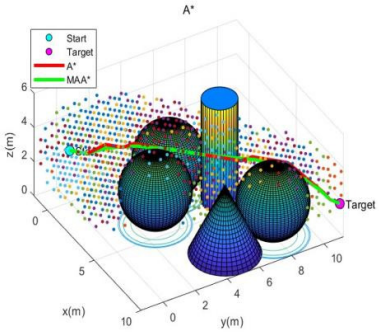
(**a**) SAPF top view



(**b**) SAPF 3D results map

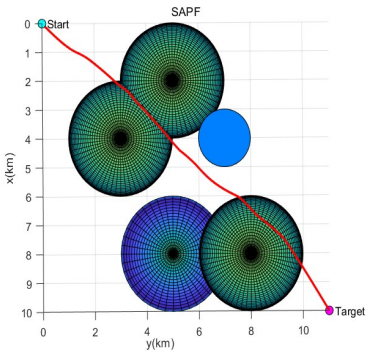


(**c**) A* top view

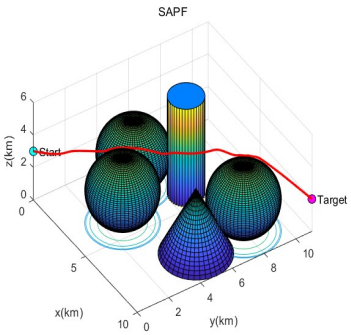


(**d**) A* 3D results map

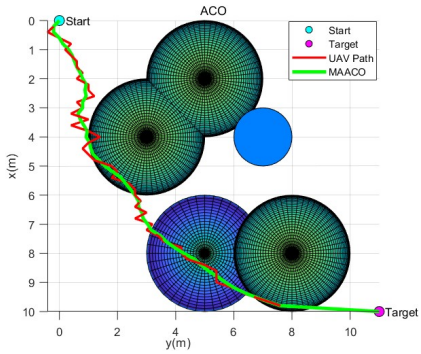
Figure 12. SAPF algorithm vs. A* algorithm results.



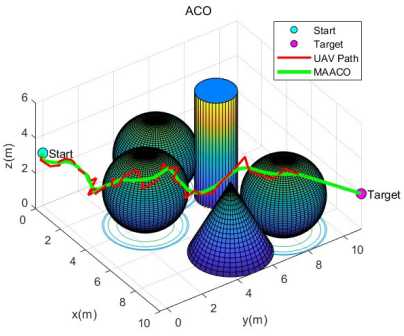
(**a**) SAPF top view



(**b**) SAPF 3D results map

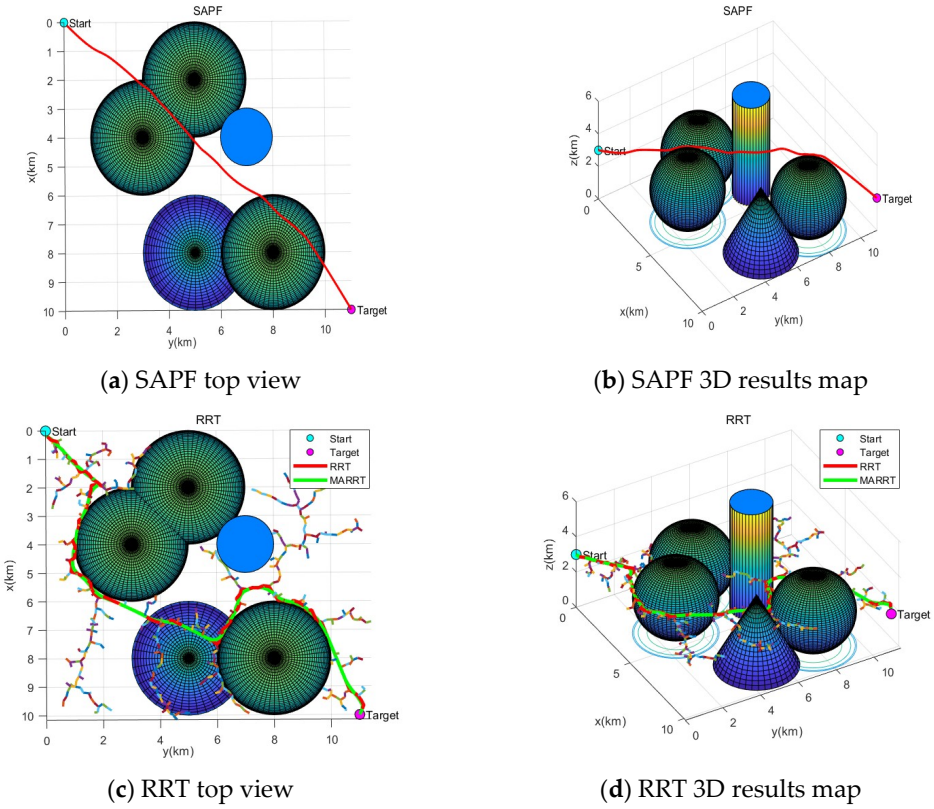


(**c**) ACO top view



(**d**) ACO 3D results map

Figure 13. SAPF algorithm vs. ACO algorithm results.



(**a**) SAPF top view (**b**) SAPF 3D results map

(**c**) RRT top view (**d**) RRT 3D results map

Figure 14. SAPF algorithm vs. RRT algorithm results.

Figure 14 shows the comparison between the SAPF and RRT algorithms, and it can be seen that the RRT algorithm has a "long way" result. When the UAV is involved in an emergency operation, this "long distance" result is not allowed, and the path of the RRT algorithm is close to the obstacles, which makes the UAV very prone to collisions with the obstacles under the influence of external factors (e.g., wind, rain, etc.). The SAPF algorithm, on the other hand, can reach the target point faster by virtue of its superiority, and the result is clearly better than the RRT algorithm.

Figure 15 presents the performance differences between the different algorithms. Compared with the Rapid Randomized Exploration Tree (RRT) algorithm and the Ant Colony (ACO) algorithm, the SAPF demonstrates significant advantages in terms of path smoothing and distance optimization. Meanwhile, the A* algorithm and the SAPF algorithm are similar in terms of their path planning effectiveness; however, the paths generated by the A* algorithm need to be smoothed, which may lead to paths intersecting with obstacles, thus making the planned routes unfeasible. In addition, the A* algorithm's search speed slows down when dealing with situations containing a large number of discrete spatial nodes. In contrast, the SAPF algorithm does not need to perform path computation in advance and, thus, shows significant superiority in terms of algorithm execution efficiency.

In order to further demonstrate the effect of the four algorithms, the metrics such as the track length L, the global curvature of the route GS, and the maximum curvature of the route LS are calculated and shown in the form of bar charts and the results are shown in Table 7 and Figure 16. It can be seen that the SAPF algorithm proposed in this paper in different environments under the measure of different indicators of its planned flight path is significantly better than various traditional algorithms, proving the excellent performance of the SAPF algorithm in UAV obstacle avoidance.

Table 7. Results of indicator calculations.

	SAPF	A*	RRT	ACO
L (km)	15.37	16.23	23.35	28.34
GS (°)	2.80	13.95	16.50	46.57
LS (°)	11.00	47.66	55.57	92.73

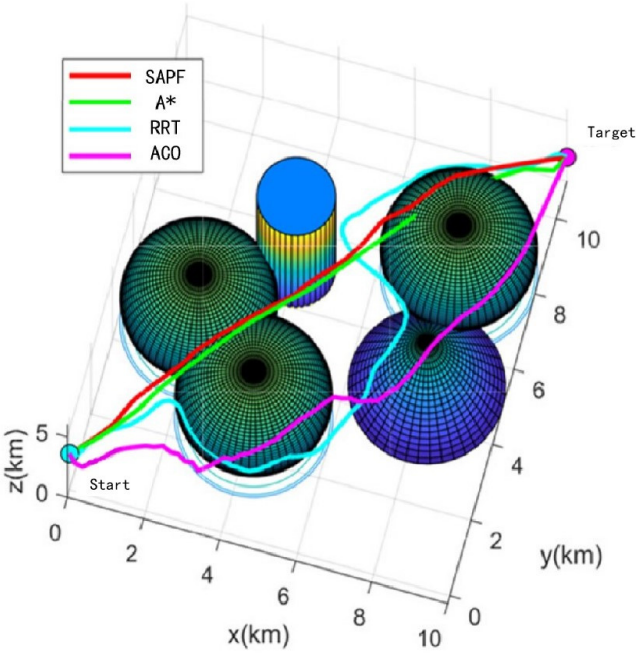


Figure 15. Comparison results of four algorithms: SAPF, A*, ACO, and RRT.

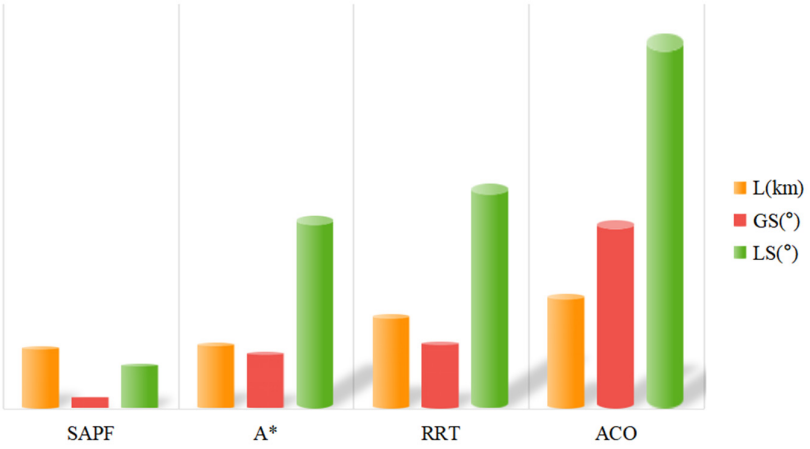
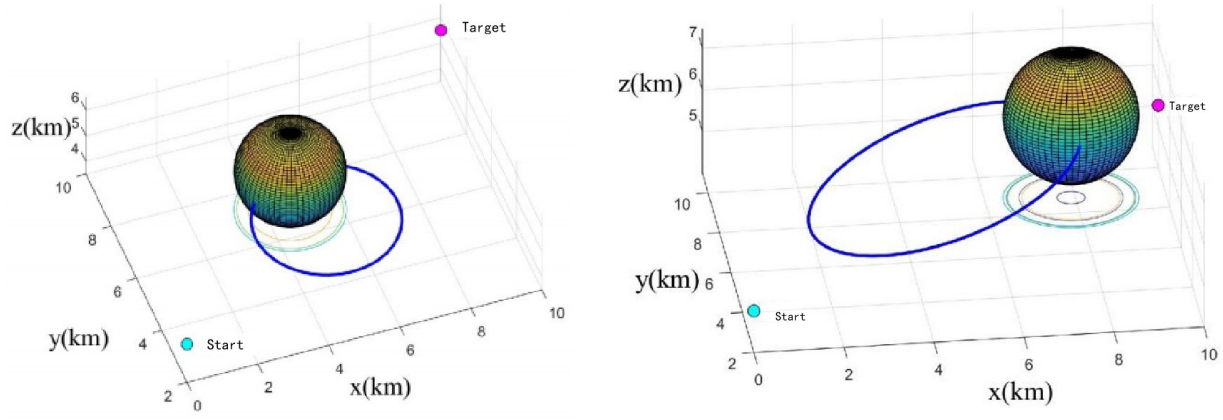


Figure 16. Indicator comparison charts.

5.5. Dynamic Environment Test Experiment

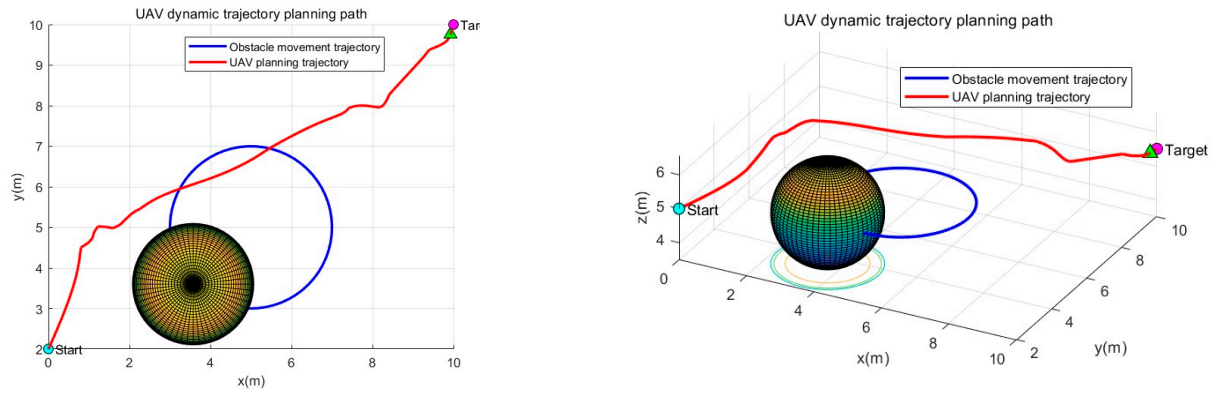
In order to further validate the effectiveness of the SAPF algorithm in dynamic obstacle environments, this study applied it to the trajectory planning task of a UAV. Figure 17 demonstrates the dynamic obstacle simulation environment used to train the UAV, where the blue curve represents the motion trajectory of the dynamic obstacle and the red curve represents the flight trajectory of the UAV under the planning of the SAPF algorithm. Figure 18 further shows the operation results of the SAPF algorithm in dynamic obstacle environments with different complexities. The experiments show that the SAPF algorithm exhibits excellent performance in multiple scenarios, not only planning a reasonable trajectory in real time but also ensuring the continuity and smoothness of the UAV flight path. In

addition, the algorithm is outstanding in terms of the collision avoidance effect, which can effectively deal with the random motion of dynamic obstacles and avoid the local minima problem and oscillation phenomenon commonly found in the traditional APF method.

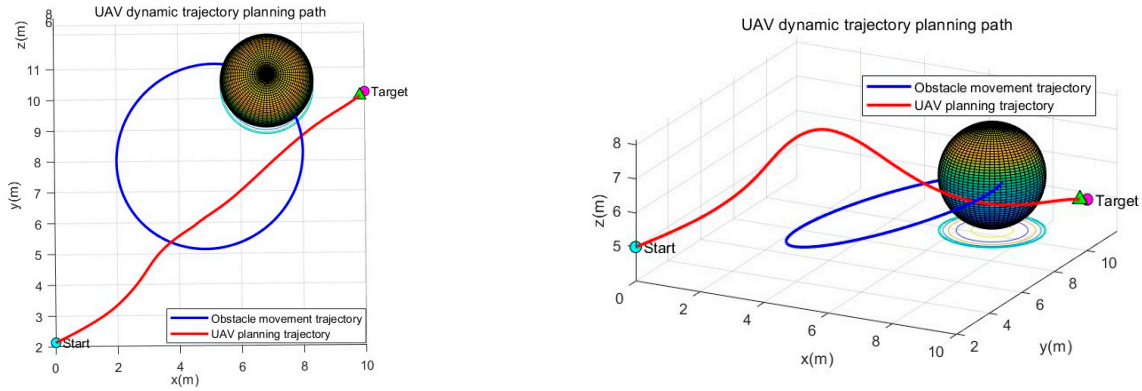


(**a**) Dynamic barrier environments 1 (**b**) Dynamic barrier environments 2

Figure 17. Dynamic barrier environments.



(**a**) Top view in dynamic obstacle environment 1 (**b**) The 3D maps in dynamic obstacle environments 1



(**c**) Top view in dynamic obstacle environment 1 (**d**) The 3D maps in dynamic obstacle environments 1

Figure 18. Performance results of the SAPF algorithm in different dynamic obstacle environments.

5.6. Analysis of Energy Consumption

In order to evaluate the performance of the improved APF algorithm in real UAV applications, energy consumption was quantitatively analyzed in this paper. The energy consumption of UAVs mainly includes power consumption and hovering consumption, which are modeled as follows:

Power consumption: the power consumption P_{motor} is related to the thrust T and the flight speed v and is calculated as follows:

$$P_{motor} = k \cdot T \cdot v \tag{22}$$

where k is the coefficient of power efficiency, which is taken as 0.8.

Hover consumption: the power consumption P_{hover} during hovering is related to the weight of the UAV, m, and the gravitational acceleration, g, which is calculated as follows:

$$P_{motor} = k \cdot T \cdot v \tag{23}$$

where a is the rotor efficiency, taken as 0.7.

Total energy consumption: the total energy consumption E is the credit for power consumption and hover consumption as follows:

$$E = \int_0^t (P_{motor} + P_{hover}) dt \tag{24}$$

The parameter settings of the UAV in the experiment are shown in Table 8. The results of the flight are shown in Table 9. The experimental results show that the improved APF algorithm reduced the flight time and unnecessary maneuvers by optimizing the path planning, thus reducing the energy consumption by about 19%. This improvement significantly improved the energy efficiency of the UAV and extended its endurance, which is important for battery endurance and mission execution efficiency in practical applications.

Table 8. Parameterization of UAVs.

Parameter Symbol	m (kg)	v (m/s)	T (N)	g (m/s^2)	η	k
Set point	1.5	5	15	9.81	0.7	0.8

Table 9. Results of the flight.

Metrics	*APF*	*SAPF*
Total flight time/(s)	150	120
Total energy consumption/(J)	10,500	8500

6. Conclusions

This study focuses on deep reinforcement learning-based UAV route planning strategies and designed an adaptive route planning algorithm based on a reinforcement learning framework for static obstacle environments. First, an artificial potential field algorithm was introduced as a preliminary solution for the characteristics of static obstacle environments, and its basic principles were elaborated on in detail. In view of the limitations of the traditional artificial potential field algorithm, this study further proposed an improved artificial potential field algorithm and verified its effectiveness in obstacle avoidance route planning through MATLAB simulation. Secondly, this paper proposed an SAPF algorithm combined with a deep deterministic policy gradient reinforcement learning architecture. The algorithm is inspired by the artificial potential field algorithm, which regards obstacles as a form of intelligence in reinforcement learning, and through the mutual learning and collaboration mechanisms between forms of intelligence, the algorithm is able to dynamically adjust the "potential field" in the space, so as to optimize the UAV's route planning. In the Experimental section, this paper compared the proposed SAPF algorithm with the traditional path planning algorithms (including the A* algorithm, the fast random search

tree algorithm RRT, and the ant colony optimization algorithm ACO) in a static obstacle environment. The experimental results show that the SAPF algorithm demonstrates significant advantages in route planning.

UAV path planning technology has a wide range of potential applications in real-world formation flight scenarios and can significantly improve the efficiency of scenarios such as surveillance missions, search and rescue operations, and cargo transportation. In surveillance missions, optimized path planning algorithms enable UAV formations to cover the target area with optimal paths, reducing repeated flights and energy waste while improving data collection and transmission efficiency through dynamic obstacle avoidance and cooperative work. In search and rescue operations, the improved path planning technology enables UAV formations to respond quickly and adapt to complex environments, shortening response time and improving the success rate of search and rescue through multi-target searches and task allocation. In cargo transportation, the path planning algorithm reduces energy consumption, extends endurance through energy optimization and multi-UAV cooperative transportation and ensures the successful completion of transportation tasks through dynamic path adjustment. These improvements not only enhance the performance of UAV formation but also provide technical support for its application in more practical scenarios, demonstrating the important value of path planning technology for improving the efficiency of UAV formation flight.

In our future research, we will focus on the following specific and actionable directions: first, exploring multi-objective optimization path planning by introducing multi-objective optimization algorithms (e.g., NSGA-II) to simultaneously optimize the path length, energy efficiency, and mission completion time; second, we aim to research real-time path planning in dynamic environments using deep reinforcement learning (DRL) and sensor fusion to achieve the real-time sensing and dynamic adjustment of UAV formation and their adjustment to environmental changes; third, research on cooperative path planning for heterogeneous UAV formations should be carried out, solving the problem of task allocation and path coordination between different types of UAVs (e.g., fixed-wing and multi-rotor UAVs); fourth, energy-efficient path planning algorithms should be designed to optimize flight paths by combining the energy consumption model and wind field information to prolong the endurance time of UAVs; and lastly, the management of large-scale UAV formations should be explored with the development of a large-scale UAV formation management algorithm. UAV formation management and developed distributed path planning algorithms and communication protocols can support the cooperative work of large-scale formations. These research directions will further enhance the performance and application value of UAV formation and provide technical support for future practical applications.

Author Contributions: M.S.: software, validation, writing—original draft; X.L.: conceptualization, methodology, project administration, funding acquisition; T.Z.: writing—review and editing; H.Y. and C.X.: visualization, supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Shandong Province issued by the Science and Technology Department of Shandong Province under grant number: ZR2022QE201.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. Lee, M.; Choi, M.; Yang, T.; Kim, J.; Kim, J.; Kwon, O.; Cho, N. A Study on the Advancement of Intelligent Military Drones: Focusing on Reconnaissance Operations. *IEEE Access* **2024**, *12*, 55964–55975. [CrossRef]
2. Hewett, R.; Puangpontip, S. On Controlling Drones for Disaster Relief. *Procedia Comput. Sci.* **2022**, *207*, 3703–3712. [CrossRef]
3. Ghelichi, Z.; Gentili, M.; Mirchandani, P.B. Drone logistics for uncertain demand of disaster-impacted populations. *Transp. Res. Part C Emerg. Technol.* **2022**, *141*, 103735. [CrossRef]
4. Ghazali, M.H.M.; Azmin, A.; Rahiman, W. Drone implementation in precision agriculture—A survey. *Int. J. Emerg. Technol. Adv. Eng.* **2022**, *12*, 67–77. [CrossRef] [PubMed]
5. Husain, Z.; Al Zaabi, A.; Hildmann, H.; Saffre, F.; Ruta, D.; Isakovic, A.F. Search and rescue in a maze-like environment with ant and dijkstra algorithms. *Drones* **2022**, *6*, 273. [CrossRef]
6. Chen, Y. Multi-UAV avoiding barrier problem based on the improved artificial potential afield nd A-star algorithm. In Proceedings of the International Conference on Algorithms, Software Engineering, and Network Security, Nanchang, China, 26–28 April 2024; pp. 1–5.
7. Kyriakakis, N.A.; Stamadianos, T.; Marinaki, M.; Marinakis, Y. A Simulated Annealing Heuristic Approach for the Energy Minimizing Electric Vehicle Routing Problem with Drones. In *Disruptive Technologies and Optimization Towards Industry 4.0 Logistics*; Springer International Publishing: Cham, Switzerland, 2024; pp. 227–246.
8. Fan, X.; Guo, Y.; Liu, H.; Wei, B.; Lyu, W. Improved artificial potential field method applied for AUV path planning. *Math. Probl. Eng.* **2020**, *2020*, 6523158. [CrossRef]
9. Katoch, S.; Chauhan, S.S.; Kumar, V. A review on genetic algorithm: Past, present, and future. *Multimed. Tools Appl.* **2021**, *80*, 8091–8126. [CrossRef] [PubMed]
10. Cui, J.; Wu, L.; Huang, X.; Xu, D.; Liu, C.; Xiao, W. Multi-strategy adaptable ant colony optimization algorithm and its application in robot path planning. *Knowl. Based Syst.* **2024**, *288*, 111459. [CrossRef]
11. Zheng, L.; Yu, W.; Li, G.; Qin, G.; Luo, Y. Particle swarm algorithm path-planning method for mobile robots based on artificial potential fields. *Sensors* **2023**, *23*, 6082. [CrossRef] [PubMed]
12. Zhang, Z.; Fu, Y.; Gao, K.; Pan, Q.; Huang, M. A learning-driven multi-objective cooperative artificial bee colony algorithm for distributed flexible job shop scheduling problems with preventive maintenance and transportation operations. *Comput. Ind. Eng.* **2024**, *196*, 110484. [CrossRef]
13. Tang, Y.; Zheng, E.; Qiu, X. Three-Dimensional Trajectory Planning for Unmanned Aerial Vehicles Based on Optimized Bidirectional A* and Artificial Potential Field Method. *J. Air Force Eng. Univ.* **2024**, *25*, 69–75.
14. Chen, H.; Chen, H.; Qiang, L. Path Planning Method for Multi-UAV Formation Based on Improved Artificial Potential Field Method. *J. Comput. Appl.* **2024**, *32*, 414–420.
15. Xian, B.; Song, N. Multi-UAV Path Planning Based on Model Predictive Control and Improved Artificial Potential Field Method. *Control Decis.* **2024**, *39*, 2133–2141.
16. Guo, H.; Chen, M.; Shen, Y.; Lungu, M. Distributed Event-Triggered Collision Avoidance Formation Control for QUAVs With Disturbances Based on Virtual Tubes. *IEEE Trans. Ind. Electron.* **2025**, *72*, 1892–1903. [CrossRef]
17. Guo, H.; Mou, C.; Lungu, M.; Li, B. Distributed event-triggered collision avoidance coordinated control for QUAVs based on flexible virtual tubes. *Chin. J. Aeronaut.* **2024**, *38*, 103300. [CrossRef]
18. Guo, H.; Chen, M.; Shi, S.; Han, Z.; Lungu, M. Distributed Coordinated Control for QUAVs with Switching Formation Strategy. *IEEE Trans. Intell. Veh.* **2024**, 1–11. [CrossRef]
19. Li, L.; Wu, D.; Huang, Y.; Yuan, Z.M. A path planning strategy unified with a COLREGS collision avoidance function based on deep reinforcement learning and artificial potential field. *Appl. Ocean Res.* **2021**, *113*, 102759. [CrossRef]
20. Wang, Z.; Yan, H.; Wang, Y.; Xu, Z.; Wang, Z.; Wu, Z. Research on Autonomous Robots Navigation based on Reinforcement Learning. In Proceedings of the 2024 3rd International Conference on Robotics, Artificial Intelligence and Intelligent Control (RAIIC), Mianyang, China, 5–7 July 2024; pp. 78–81.
21. Puente-Castro, A.; Rivero, D.; Pazos, A.; Fernandez-Blanco, E. UAV swarm path planning with reinforcement learning for field prospecting. *Appl. Intell.* **2022**, *52*, 14101–14118. [CrossRef]
22. Fan, X.; Li, H.; Chen, Y.; Dong, D. UAV Swarm Search Path Planning Method Based on Probability of Containment. *Drones* **2024**, *8*, 132. [CrossRef]
23. Guo, J.; Xian, Y.; Ren, L.; Li, P. UAV Trajectory Planning Based on Adaptive Repulsive Potential Field. *J. Beijing Univ. Aeronaut. Astronaut.* **2024**.

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.