TYPE Original Research
PUBLISHED 24 January 2023
DOI 10.3389/fnbot.2022.1076338

Trans-UTPA: PSO and MADDPG based multi-UAVs trajectory planning algorithm for emergency communication

Jie Li[1], Shuang Cao[2], Xianjie Liu[1], Ruiyun Yu[2] and Xingwei Wang[1]*

[1]School of Computer Science and Engineering, Northeastern University, Shenyang, Liaoning, China, [2]School of Software, Northeastern University, Shenyang, Liaoning, China

Check for updates

OPEN ACCESS

EDITED BY
Shangce Gao,
University of Toyama, Japan

REVIEWED BY
Xianda Chen,
The Pennsylvania State University (PSU),
United States
Ting Jin,
Nanjing Forestry University, China

*CORRESPONDENCE
Xingwei Wang
✉ wangxw@mail.neu.edu.cn

RECEIVED 21 October 2022
ACCEPTED 31 December 2022
PUBLISHED 24 January 2023

CITATION
Li J, Cao S, Liu X, Yu R and Wang X (2023) Trans-UTPA: PSO and MADDPG based multi-UAVs trajectory planning algorithm for emergency communication. *Front. Neurorobot.* 16:1076338. doi: 10.3389/fnbot.2022.1076338

Communication infrastructure is damaged by disasters and it is difficult to support communication services in affected areas. UAVs play an important role in the emergency communication system. Due to the limited airborne energy of a UAV, it is a critical technical issue to effectively design flight routes to complete rescue missions. We fully consider the distribution of the rescue area, the type of mission, and the flight characteristics of the UAV. Firstly, according to the distribution of the crowd, the PSO algorithm is used to cluster the target-POI of the task area, and the neural collaborative filtering algorithm is used to prioritize the target-POI. Then we also design a Trans-UTPA algorithm. Based on MAPPO 's policy network and value function, we introduce transformer model to make Trans-UTPA's policy learning have no action space limitation and can be multi-task parallel, which improves the efficiency and generalization of sample processing. In a three-dimensional space, the UAV selects the emergency task to be performed (data acquisition and networking communication) based on strategic learning of state information (location information, energy consumption information, etc.) and action information (horizontal flight, ascent, and descent), and then designs the UAV flight path based on the maximization of the global value function. The experimental results show that the performance of the Trans-UTPA algorithm is further improved compared with the USCTP algorithm in terms of the success rate of each UAV reaching the target position, the number of collisions, and the average reward of the algorithm. Among them, the average reward of the algorithm exceeds the USCTP algorithm by 13%, and the number of collisions is reduced by 60%. Compared with the heuristic algorithm, it can cover more target-POIs, and has less energy consumption than the heuristic algorithm.

KEYWORDS
multi-UAVs collaboration, PSO, trajectory planning, energy consumption, multi-agent reinforcement learning, transformer

1. Introduction

UAVs play an increasingly important role in emergency communication networks. Due to natural disasters, communication infrastructure cannot work properly. Rescue missions need to be fast and agile. The use of multiple UAVs to form an air UAV group self-organizing network can achieve low-latency and high-reliability air-ground coordinated transmission between the UAV group and the ground intelligent terminal equipment (Wang et al., 2021). With the advent of the 6G era, UAVs are expected to offer additional new services such as real-time image transmission, caching and multicast, data dissemination or collection (Wu and Zhang, 2018; Wu et al., 2018) mobile relay and edge computing (Li et al., 2015; Yang et al., 2019; Ma et al., 2021c), and wireless power transmission.

In emergency scenarios, UAV clusters often undertake complex tasks with multiple targets and nodes. The core goal of multi-UAVs task allocation is how to achieve the efficient use of each UAV under the premise of ensuring the completion of the overall task, that is, the overall task allocation balance. The machine learning method enhanced by evolutionary computation (ECML) combines the advantages of ML and EC, and has strong potential. In particular, ECML (Zhang et al., 2022) also has strong search ability, which can greatly reduce the computational cost of cluster node analysis. In the actual construction of UAV cluster task allocation algorithm, genetic algorithm, particle swarm optimization algorithm (Lipare et al., 2021; Krishna et al., 2022; Pu et al., 2022) and ant colony algorithm are often introduced, and bionic intelligence is used to realize better cooperation between UAV cluster individuals, so as to achieve the overall task goal with the best effect (Wangsheng et al., 2021).

Effectively designing the flight path (Jin and Yang, 2021) of UAVs can improve the working efficiency of UAVs. At present, the existing research on single UAV path planning has been very mature. Classical algorithms such as A * algorithm, Dijkstra algorithm, wavefront algorithm and fast exploration random tree algorithm have been widely used in UAVs. With the latest development of machine learning, the research and development of path planning algorithms based on machine learning has been growing rapidly, such as value iterative network, gated neural network and other path planning algorithms. In recent years, with the use of collaborative mission scenarios become widespread, the demand for UAV swarm collaborative path planning (Yao et al., 2016; Jin et al., 2021; Ma et al., 2021a) has become more urgent. The path planning of UAV cluster not only needs to consider the flight distance and energy consumption of single UAV, but also needs to evaluate the safety and cooperation ability of multiple UAVs to ensure that the UAV cluster can perform tasks safely and efficiently. Multi-agent reinforcement learning is the key to solving this problem.

The existing research on multi-agent motion planning problems can be roughly divided into two categories, centralized (Tang et al., 2018) and decentralized methods (Ma et al., 2021b). The centralized method defines the motion planning problem as an optimization problem, where the position, velocity, and target position information of all agents are available. The goal is to guide all agents toward their desired positions, avoid collisions (Tian et al., 2022), and minimize targets (such as energy or time). The decentralized algorithms can be appropriately extended because they allocate the computational effort to multiple agents. They are also very robust to interference when performing real-time calculations.

Although UAVs communication has rich application value, considering the limited airborne energy of a UAV and the need to simultaneously provide energy for propulsion and communication, achieving green UAV communication flight is a key challenge. At present, energy-saving UAVs communication (Liu et al., 2022) can be divided into three categories: (i) given communication requirements, minimize energy consumption; (ii) given total energy/power budget to maximize performance gain; (iii) Maximize energy efficiency. One way to improve energy efficiency is to reduce path loss.

Therefore, we design a PSO-based Trans-UTPA UAVs path planning algorithm. UAV path planning for multi-UAVs cooperative communication and data acquisition is studied. The contributions of this article are as follows:

(1) We propose a PSO-based global optimal target-POI clustering model. We regard each affected person as a particle. Assuming that the particle velocity is constant, we calculate the average distance and standard deviation based on the position of the particle and the position of the cluster center, and minimize the average distance and standard deviation to design the fitness function. All particles change their positions according to the maximum value of the global adaptive function value, so that all particles have clustering centers. We call the population location formed by all particles in a cluster center the target-POI region, where the UAV performs an emergency mission.

(2) We designed a green energy consumption calculation model. The UAV performs flight tasks in a three-dimensional environment, involving flight energy consumption, communication energy consumption and data acquisition energy consumption. When designing multi-UAV cooperative trajectory planning, we comprehensively consider the distance between each UAV and target-POI, the power of each UAV and the priority of target-POI, and select the appropriate UAV to perform flight tasks on target-POI to ensure that all UAVs have low energy consumption while completing flight tasks.

(3) We designed the Trans-UTPA multi-UAV cooperative path planning algorithm. Based on the MAPPO algorithm, we introduce Transformer mechanism to replace the traditional RNN structure for UAVs track sequential modeling. Multi-UAV performs strategy learning based on state information and action information in three-dimensional space, and selects the emergency task with the largest global value function in the evaluation network. The strategy learning of Trans-UTPA has no action space limitation and can be multi-task parallel, which improves the efficiency and generalization of sample processing.

(4) Through the simulation platform, the performance of the algorithm proposed in this paper is tested, and the Trans-UTPA is compared with the USCTP algorithm. The results show that the performance of the algorithm is better than that of the USCTP algorithm in terms of the success rate of the UAV reaching the target, the number of collisions of the UAV and the average reward of the algorithm. Compared with the UAV which executes A * algorithm, it consumes less energy and has certain advantages.

The organization of this chapter is as follows : The second part introduces the existing related research. The third part introduces the clustering recommendation based on PSO. The fourth part introduces the multi-UAVs cooperative path planning constraint model and the multi-UAVs cooperative path planning green energy consumption calculation model and Trans-UTPA algorithm. The fifth part reports the evaluation of the algorithm. We explain the simulation settings, and get the following result diagram through different parameter settings. In the sixth part, a conclusion and the next stage of work are given.

2. Related work

In emergency rescue scenarios, the crowds are scattered and the coverage of UAV services is limited. So we define an NP-hard problem: Given N points with a certain distance. We hope to select K cluster centers from the given vertices, and the remaining vertices and cluster centers complete the clustering. Minimize the distance from one vertex to other vertices in a cluster.

In order to solve the NP problem, a large number of clustering algorithms have emerged. The existing clustering methods are usually divided into density-based, hierarchical, graph decomposition and partitioning methods. The density-based algorithm is represented by DB-SCAN (Shinde et al., 2022) CGCA (Kowalski and Jeczmionek, 2022) and other similar methods. Hierarchical clustering methods can use two different strategies : top-down and bottom-up (also known as agglomerative clustering) (Kordos et al., 2022). Finally, the most popular is the partition method using k-means algorithm (Ma et al., 2021d; Sathyamoorthy et al., 2022), although these methods can achieve better clustering effect, because the emergency rescue pay attention to efficient and fast, so we use the PSO on the line clustering, because the PSO algorithm has a local optimal solution and global optimal solution, in order to avoid local convergence, we adjust the particle motion through the global optimal, in addition to PSO aggregation speed is also very fast.

When everyone has their own clustering population, designing a track for the UAV has become an important issue. A good route can improve the efficiency of the UAV. Dynamic path planning in unknown environments has always been a challenge in the current research field (Zhang et al., 2021). The dual Q network (DDQN) deep reinforcement learning proposed by DeepMind is applied to dynamic path planning in unknown environment. A reward-penalty function is set up. With the updating of the neural network and the increase of the probability of greedy rule, the local space searched by an agent is expanded. The results show that the reinforcement learning algorithm enhances the dynamic obstacle avoidance and local planning ability of the agent in the environment. We use multi-agent for multi-UAVs path planning. In addition to meeting obstacle avoidance and networking communication, it can also cover more clustering points and improve the utilization rate of UAV.

In addition, the energy of UAV is limited. How to maximize the efficiency of UAVs under limited energy is an important research content (Ahmed et al., 2016). Three algorithms are proposed to solve the multi-UAVs path planning problem in multiple 2D, barrier-free and discrete planes, while meeting the coverage requirements and minimum energy consumption. Discretize the space and provide a more realistic view of how the UAV travels along its path to ensure a collision-free trajectory. Li W. et al. (2022) proposed an improved probabilistic road map (IPRM) algorithm to solve the energy consumption problem of multi-UAVs path planning. The mathematical model and energy consumption model are established by simulating the real terrain environment. The sampling space of PRM algorithm is optimized to make the path clearer and improve the utilization of space and time.

Zhu et al. (2022) proposed a hexagonal region search (HAS) algorithm, which is combined with multi-agent deep Q network (DQN), called HAS-DQN. By limiting the total coverage of UAVs, HAS-DQN can effectively avoid collision with UAVs. Experiments show that HAS-DQN can effectively solve the path overlapping problem of multiple UAVs moving at the same cost in unknown environment.

In the algorithm design in this paper, we not only consider the energy consumption of UAV flight action, but also consider the energy consumption of UAVs communication and data acquisition. Combining the key degree of POI, the distance from the UAV to the mission area, and the remaining power of the UAV, a reward mechanism is designed to promote the UAV to make the optimal decision, that is, to maximize the global reward.

3. Clustering recommendation based on PSO

In the emergency communication scenario, the affected population is scattered, and the number of UAVs and endurance energy are limited. In order to enable UAVs to complete emergency rescue missions for emergency areas, we use genetic algorithms to complete the selection of initial seeds. Then, according to the PSO algorithm, the fitness function is calculated according to the location information of the crowd to find the optimal group to complete the crowd clustering, as shown in Figure 1.

3.1. Preliminaries

(1) target-POI definition: we define the region formed after clustering as target-POI.
(2) Coverage hypothesis: it is assumed that the size of the area formed after clustering is within the coverage of the UAV.
(3) The PSO speed hypothesis: the speed of the particle motion process is constant, with a constant retrieval speed, recorded as v.
(4) UAV to complete the ascending or descending flight energy consumption assumptions:the UAV ascending or descending flight operation energy consumption is constant, the amount of consumption and the number of execution.
(5) UAV data storage: it is assumed that the results of UAV data acquisition can be temporarily stored in UAV memory.

3.2. Initial seed selection based on genetic algorithm

The affected people are scattered in the affected area. To complete the clustering, we must first select the clustering center to complete the clustering partition. The selection of cluster centers is particularly important. Correct selection of cluster centers can improve work efficiency and reduce the consumption of some materials and human resources. Genetic algorithms (GAs) are inspired by the theory of nature, which is known for its global adaptability and robust search capabilities to capture good solutions (Sharma and Kaushik, 2021). Therefore, we adopt a genetic algorithm to complete the selection of initial seeds.

First, the chromosome with K genes is set to K cluster centers, denoted as $X = \{X_1, X_2, X_3, ...X_k\}$, where X_k represents an n-dimensional vector, and the other interior points are denoted as $C = \{C_1, C_2, C_3, ..., C_i\}$. The latitude and longitude coordinates of C_i is (ψ_1, α_1), and the latitude and longitude coordinates of X_k is

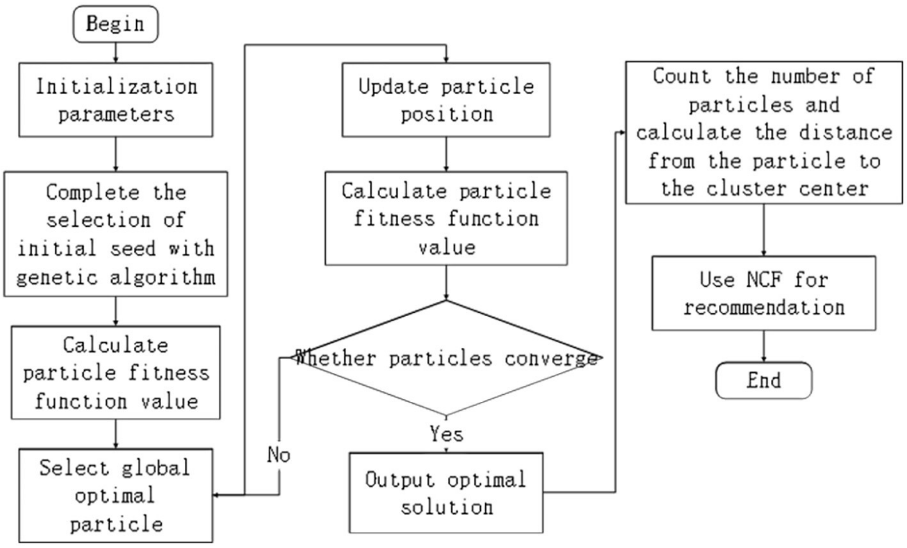
Begin
Initialization parameters
Complete the selection of initial seed with genetic algorithm
Calculate particle fitness function value
Select global optimal particle
No
Update particle position
Calculate particle fitness function value
Whether particles converge
Yes
Output optimal solution
Count the number of particles and calculate the distance from the particle to the cluster center
Use NCF for recommendation
End

FIGURE 1
Flowchart of PSO-based clustering recommendation algorithm.

(ψ_2, α_2).During evolution, the fitness function (Equation 1) is used to calculate the sum of the distances (Equation 2) (Wang et al., 2014) from all interior points to the cluster center.

$$\begin{aligned}\mathrm{Dist(C_i, X_j)} &= haversin\left(\frac{d}{R}\right)\\ &= haversin(\psi_2 - \psi_1) + cos(\psi_1)cos(\psi_2)haversin(\alpha_2 - \alpha_1)\end{aligned} \tag{1}$$

$$F(poi) = \sum_{X_j \in X} min(Dist(C_i, X_j)), 1 \le i \le k \tag{2}$$

where R represents the radius of the earth, and the value is 6371 km. φ represents the latitude, α represents the longitude.

When the sum of fitness $F(poi)$ reaches the minimum, the population tends to converge to an optimal chromosome (solution). Once the optimal cluster centers come out, we use them as initial seeds, denoted as X_{init}.

3.3. PSO-based clustering recommendation

PSO is a famous bionic optimization algorithm, which is an iterative search algorithm based on population. In PSO algorithm, all particles enter the search space to find the optimal solution. We hope to use the PSO search algorithm to divide the crowd in the rescue area so that the UAV can provide rescue services. In the rescue area, we regard each person as a particle, and design the fitness function according to the average distance and standard deviation between the particle and the cluster center. The next position of each particle is calculated by a fitness function. The larger the global fitness function value is, the particle moves to that position.

After the initial seed is determined, all particles in the clustering region move to the clustering center to complete the clustering operation. Suppose that the cluster region is composed of n particles, denoted by $Y = \{Y_1, Y_2, Y_3, ..., Y_n\}$, and the position of the cluster center is denoted by $M = \{M_1, M_2, M_3, ..., M_m\}$.

The global optimal value is calculated by the social interaction of the particle in the group, and the best fitness value obtained by the particle in the population. In the search process, a particle not only needs to record its own personal best solution, but also records the overall solution of the cluster center selected by other particles, that is, the shortest reach to the cluster center. We designed a fitness function. The current position of the i th particle is denoted as P_i, and the average distance from all current particles to the cluster center M_m (Equation 3) and the standard deviation of all particle distances (Equation 4) are calculated.

$$minD = \frac{1}{n-m}\sum ||M_m - P_i||, i = 1, 2, 3, ..., n-m \tag{3}$$

$$minC = \sqrt{\frac{\sum_{i=1}^{n-m}(D - P_i)^2}{n-m}} \tag{4}$$

Then, the fitness function (Equation 5) for each current particle is designed based on the mean and standard deviation.

$$Fitness = k_1 D + k_2 C \tag{5}$$

Among them, k1 and k2 represent the weight coefficients between the average distance and the standard deviation, respectively. *Fitness* is the fitness function value.

All particles change their current position according to fitness and make the next action. The action rules are as follows (Equations 6, 7) (Wang et al., 2022):

$$fitness_i^{a+1} = fitness_i^a, ifFitness(P_i^{a+1}) \ge Fitness(fitness_i^a) \tag{6}$$

$$fitness_i^{a+1} = P_i^{a+1}, ifFitness(P_i^{a+1}) < Fitness(fitness_i^a) \tag{7}$$

where, P_i^{a+1} is the next position of the particle, $fitness_i^a$ represents the current best position of the particle, $fitness_i^{a+1}$ represents the new personal best value of the particle, and a is the number of iterations. $Fitness(P_i^a)$ refers to the fitness function value at the position of P_i^a.

$$globalworth_i^a = argMin^{n-m}_{i=1 Fitness(fitness_i^k)} \tag{8}$$

Using Equation (8), we can calculate the global optimum, where $globalworth_i^a$ is the global optimal value of particle i. Update your location by ensuring that the globalworth value is larger.

When all particles find their cluster centers, population clustering is completed. Since each particle runs to different cluster centers in the clustering process, the number of particles in each population is different. The number of particles in each population is recorded as $target - POI = \{T1, T2, T3, ..., Tm\}$,. Assuming that the initial release position of the UAV is (U_x, U_y), the distance from the UAV to the center(T_x^i, T_y^i) of the population Ti is calculated to be D_{gi} (Equation 9).

$$D_{gi} = \sqrt{(U_x - T_x^i)^2 + (U_y - T_y^i)^2} \tag{9}$$

Calculate the distance from all particles to the cluster center within population T_i, and record the maximum distance as R. Then the particle number N_{T_i} and the distance D_{gi} are converted into sparse vectors, and the neural collaborative filtering algorithm is used to complete the target-POI priority recommendation, indicating which area performs which flight task.

The values in the regional particle number feature N_{T_i} and the distance feature D_{gi} are converted into sparse vectors, which can be input into the neural network. Then, the input vector is multiplied with the embedding matrix in the embedding layer to obtain the embedded vector representation of the particle number feature and the distance feature. In the fusion layer, the dimension consistency of the particle number feature and the distance feature vector is completed. In the neural collaborative filtering layer, the vectors obtained from the pooling layer capture the nonlinear and high-order correlations between particles and regions through a hidden layer consisting of multiple fully connected layers. Finally, in the prediction layer, the output vector of the last layer is mapped to the final prediction result of target-POI. For target-POI with different priorities, UAVs perform different tasks. Target-POI with the highest priority performs task networking communication tasks, establishes a UAV network, and provides network communication for the crowd. The relatively simple target-POI performs the task of collecting relevant data.

4. Multi-UAVs path planning algorithm

4.1. Multi-UAVs cooperative path planning constraint model

In the process of emergency rescue, multiple UAVs complete the mission together. Relevant parameter statistics used in the multi-UAVs path planning constraint model are shown in Table 1.

TABLE 1 Statistics of relevant parameters used in multi-UAVs path planning constraint model.

Parameter	Interpretation
E	Onboard energy of a UAV
T_i=(T_i^x, T_i^y, T_i^z)	Location of target-POI i
O_i=(O_i^x, O_i^y, O_i^z)	The location of the i th obstacle
U_i^t=$(U_t^{Xi}, U_t^{Yi}, U_t^{Zi})$	Position of UAV i at time t
V_t	UAV flight speed at time t
$Elec_t$	Remaining power of UAV at time t
E=$\{(E_x^i, y, z) \cup (x, E_y^i, z) \cup (x, y, E_z^i) \| E_x^i = E_y^i = E_z^i = B\}$	Rescue area
B	B is the boundary value of the rescue area
V_{MAX}	Maximum Flight Speed of UAV
u	Number of flying UAVs
d_{co}	Communication distance Upper

Multiple UAVs complete the mission in a three-dimensional environment, need to pay attention to a variety of constraints: (1) Each UAV should fly in the rescue area; Equation (10) (2) In the process of emergency communication, the flight speed Vt of UAV is less than the maximum flight speed of UAV; Equation (11) (3) After the UAV reaches the mission area, it must have enough energy to complete the mission of the target-POI area, and the energy cannot be completely consumed; Equation (12) (4) Multiple access is prohibited in any target-POI zone; Equation (13) (5) Multiple UAVs cannot simultaneously appear in one area. Equation (14), where i and j denote the i-th and j-th UAVs, respectively;

$$U_i^t \in E \tag{10}$$

$$0 < V_t < V_{max} \tag{11}$$

$$Elec_t > 0 \tag{12}$$

$$Visit(T_i) = \begin{cases} 1 & \text{Not visited} \\ 0 & \text{visited} \end{cases} \tag{13}$$

$$(U_t^{Xi}, U_t^{Yi}, U_t^{Zi}) \neq (U_t^{Xj}, U_t^{Yj}, U_t^{Zj}) \tag{14}$$

The UAV's behavior during maneuvers should be prioritized as it provides smoothness during flight. The flight slope is defined as the maneuverability of the UAV during gliding and climbing. During the flight, the slope of the UAV is moved horizontally from one path point to another (Equation 15).

$$slope_i = \begin{cases} S & \text{if } slo_i \notin [tan(\alpha_{max}), tan(\beta_{max})] \\ 0 & \text{Otherwise} \end{cases} \tag{15}$$

where, $slope_i$ is the flying slope from one waypoint to the i-th waypoint; α_max and β_max represent the maximum tolerable gliding and climbing angles, and slo_i can be formulated, according to Equation (16) (Mughal et al., 2022), as follows:

$$slo_i = \frac{z_i - z_{i-1}}{||x_1 - x_{i-1}, y_i - y_{i-1}||_2} \tag{16}$$

where, slo_i is the flying slope taken by the UAV from the i-th waypoint (x_i, y_i, z_i). The UAV flight slope value $slope_i$ complies with constraint (Equation 17).

$$s \leq \frac{1}{2} \tag{17}$$

During the flight of the UAV, the maximum safe distance d_{safe} between the two adjacent UAVs should be guaranteed. The distance between the UAV i and the UAV j is calculated to be D_{ij} using Equation (18). When calculating the path of multiple UAVs, it is necessary to ensure that the UAVs are not too close. In order to maintain the safe distance between them, the limit can be expressed as follows (Equation 19):

$$D_{ij} = \sqrt{(U_t^{Xi} - U_t^{Xj})^2 + (U_t^{Yi} - U_t^{Yj})^2 + (U_t^{Zi} - U_t^{Zj})^2} \tag{18}$$

$$D_{ij} > d_{safe} \tag{19}$$

In addition, the UAV should also pay attention to the distance between the obstacle (Equation 20) to avoid collision between the UAV and the obstacle, so set the flight constraint between the UAV and the obstacle (Equation 21).

$$D_{obs} = \sqrt{(U_t^{Xi} - O_x^i)^2 + (U_t^{Yi} - O_y^j)^2 + (U_t^{Zi} - O_z^i)^2} \tag{20}$$

$$D_{obs} > d_{safe} \tag{21}$$

4.2. Green energy consumption calculation model for multi-unmanned cooperative path planning

Green energy consumption requires UAVs to consume the least energy and complete more missions during flight. We study the four basic operations of UAV flight operations: horizontal flight, ascent, descent, and hover. The horizontal flight energy consumption of UAV i at time t is recorded as $E_{level}^i(t)$, and the horizontal energy consumption calculation formula is Equation (22). The energy consumption of the UAV to complete a rise or fall is recorded as E_u and E_d, respectively.

$$E_{level}^i(t) = \sigma(c_1||v_t^i||^3 + \frac{c_2}{||v_t^i||}) \quad i = 1, 2, 3, ..., n \tag{22}$$

where σ represents the slot length, c_1 and c_2 are constants.

The flight operation of the UAV also involves hovering operation. We record the hovering time as t_{hover}, and the hovering energy consumed by the ith UAV in 1 min is recorded as $E_{hover}^i(t)$. Therefore, the calculation of the energy consumed by the i UAV in performing flight operations (Equation 23) is recorded as E_{fly}^i.

$$E_{fly}^i(t) = E_{level}^i(t)T_{fly} + P_{up} * E_{up}^i + P_{down} * E_{down}^i + t_{hover}E_{hover}^i(t) \tag{23}$$

The time consumed during the flight (excluding hover) is recorded as T_{fly},P_{up} is the number of rising, P_{down} is the number of falling.

In the multi-UAVs mission, if the distance between the UAVs is less than the maximum distance d_{com} that can be communicated between the UAVs, the UAVs communicate with each other to exchange information, and finally summarize it to a UAV. The information exchanged includes: the amount of existing energy, target-POI access, etc. In this article, we use the location with the highest demand as the network communication area, while other areas are engaged in data collection.

The UAV flight mission includes data acquisition and networking communication. So we have to calculate the UAV data acquisition energy consumption and network communication energy consumption. We refer to the location-critical communication model proposed in Qin et al. (2018). When the UAV accesses target-POI, the access information is transmitted to the nearby UAV, considering the energy consumption E_{tran}^{ij} of one transmission between UAV i and UAV j. The calculation formula (Equation 24) is as follows:

$$E_{tran}^{ij} = Bit * (D_{ij})^{\alpha} * E_m^B \tag{24}$$

Among them, Bit represents the size of the transmission information, E_m^B represents the energy consumed by transmitting 1 Mbit information within 1 km distance, and α represents the transmission loss index of the transmission medium.

In the three-dimensional environment, the UAV needs to rotate to complete the task of data acquisition. The rotation angle is also a factor that determines energy consumption. We use (Fu et al., 2018) to calculate the energy consumption caused by the rotation angle during the acquisition process as E_{corner}, and the formula is Equation (25):

$$E_{corner} = \sum_{i=1}^{n} w_n * Q \tag{25}$$

where w_n represents the angle of rotation of the nth target-POI UAV, and Q represents the energy consumed by one rotation. So UAV i access n target-POIs to complete the rotation of the fuselage and data collection tasks consumed energy recorded as $E_{collect}^i$, the formula is as follows (Equations 26, 27):

$$E_{collect}^i = P_{collect} * T_{target-POI}^j * n + E_{corner} \tag{26}$$

$$T_{target-POI}^j = \int_0^{2R_{max}} \frac{1}{V_t} \, \mathrm{d}r \tag{27}$$

After completing n target-POI missions in flight time T_{fly}, the residual energy of UAV i is calculated by Equation (28).

$$Elec_t = E - E_{fly}^i - \sum_{i=0}^{u-1} \sum_{j=1}^{u} E_{tran}^{ij} - E_{collect}^i \tag{28}$$

4.3. Trans-UTPA algorithm

4.3.1. Trans-UTPA algorithm framework

The architecture of the whole algorithm is shown in Figure 2. The individual value function of each UAV can be calculated

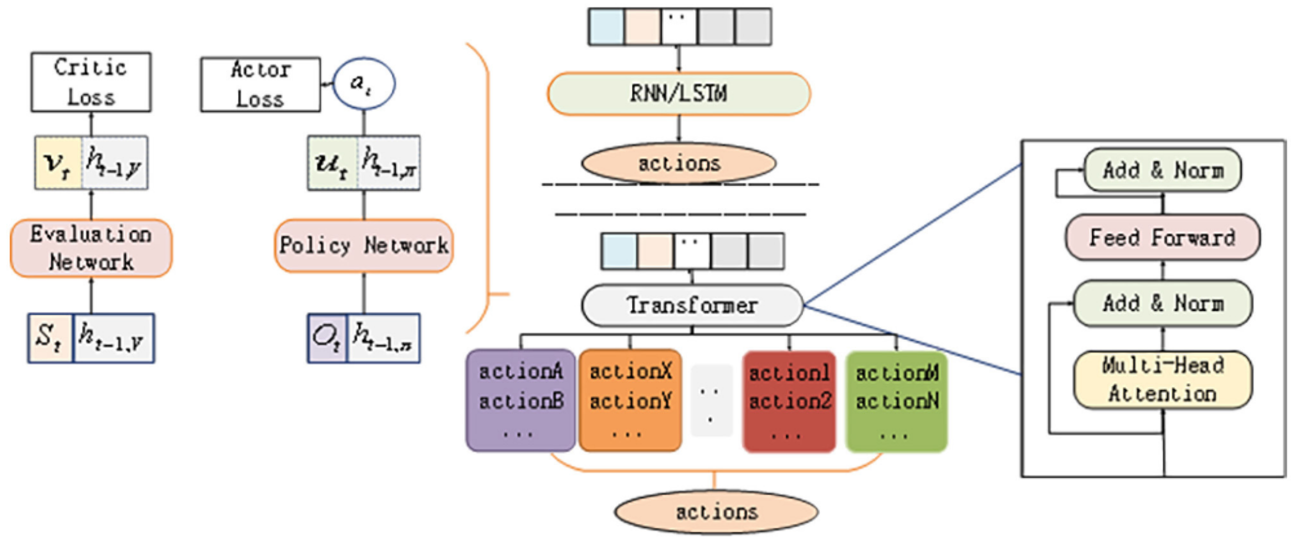

FIGURE 2
Trans-UTPA algorithm framework.

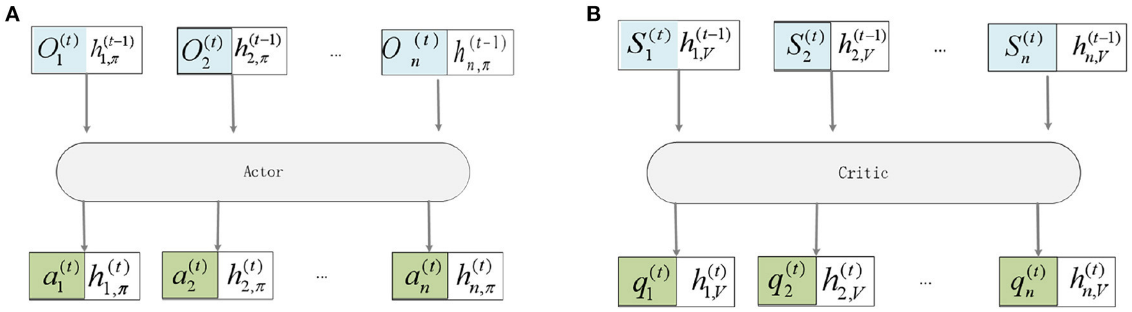
A
$O_1^{(t)}$ $h_{1,\pi}^{(t-1)}$
$O_2^{(t)}$ $h_{2,\pi}^{(t-1)}$
...
$O_n^{(t)}$ $h_{n,\pi}^{(t-1)}$
Actor
$a_1^{(t)}$ $h_{1,\pi}^{(t)}$
$a_2^{(t)}$ $h_{2,\pi}^{(t)}$
...
$a_n^{(t)}$ $h_{n,\pi}^{(t)}$
B
$S_1^{(t)}$ $h_{1,V}^{(t-1)}$
$S_2^{(t)}$ $h_{2,V}^{(t-1)}$
...
$S_n^{(t)}$ $h_{n,V}^{(t-1)}$
Critic
$q_1^{(t)}$ $h_{1,V}^{(t)}$
$q_2^{(t)}$ $h_{2,V}^{(t)}$
...
$q_n^{(t)}$ $h_{n,V}^{(t)}$

FIGURE 3
Embedded representation of observations. **(A)** This is Actor framework. **(B)** This is Critic framework.

using the self-attention mechanism. Since Trans-UTPA is composed of two independent networks, namely the policy network $\pi_\theta(a_i|o_i)$ containing parameter θ and the evaluation network $V_\phi(s)$ containing parameter ϕ, these two networks are processed separately. All UAVs share the critic network, and each UAV has its own actor-critc network. First, the collected trajectory is recorded as $\tau = \tau_1, \tau_2, ..., \tau_n$,$n$ represents the number of UAVs, and the trajectory of a single UAV i is recorded as $\tau_i = \{s_i^{(t)}, o_i^{(t)}, a_i^{(t)}\}$.

4.3.2. Design of strategy function and value function based on Transformer

The state information $S_{i,1}^{(t)}, .., S_{i,m}^{(t)}$ of each UAV at time t is input into an embedded layer for representation, and the state value $S_i^{(t)}$ of the UAV in the environment at time t is obtained (Equations 29, 30):

$$O_i^{(t)} = E(o_{i,1}^{(t)}), E(o_{i,2}^{(t)}), ..., E(o_{i,m}^{(t)}) \tag{29}$$

$$S_i^{(t)} = E(s_{i,1}^{(t)}), E(s_{i,2}^{(t)}), ..., E(s_{i,m}^{(t)}) \tag{30}$$

The predicted action and value of the UAV at time t are Equations (31), (32):

$$h_{i,\pi}^{(t)}, a_i^{(t)} = \pi(O_i^{(t)}, h_{i,\pi}^{(t-1)}, \theta_i) \tag{31}$$

$$h_{i,V}^{(t)}, q_i^{(t)} = V_i(S_i^{(t)}, h_{i,V}^{(t-1)}, \phi_i) \tag{32}$$

where $h_{i,\pi}^{(t-1)}$ represents the state of the hidden layer t at a time on the actor, $h_{i,V}^{(t-1)}$ represents the state of the hidden layer at a time on the critic, $O_i^{(t)}$ represents the observation embedding value of the UAV i to other UAVs, $S_i^{(t)}$ represents the state embedding value of each UAV, θ_i is the parameter defining π, and ϕ_i is the parameter defining V_i. In this way, the critic network of UAV i evaluates the state of other UAVs.As shown in Figure 3.

Here the hidden layer is treated as part of each UAV, each input holds its own hidden layer, and each output maps to a new hidden layer for the next time step. Using n to represent the number of UAVs, L to represent the number of fully connected layers, the actor network input and output relationship expression are Equations (33), (34):

$$R_\pi^1 = \{h_{1,\pi}^{(t-1)}, h_{2,\pi}^{(t-1)}, ..., h_{n,\pi}^{(t-1)}, O_i^{(t)}\} \tag{33}$$

$$\{h_{1,\pi}^{(t)}, h_{2,\pi}^{(t)}, ..., h_{n,\pi}^{(t)}, O_L^{(t)}\} = R_\pi^L \tag{34}$$

The input-output relationship expression of critic network are Equations (35), (36):

$$R_V^1 = \{h_{1,V}^{(t-1)}, h_{2,V}^{(t-1)}, ..., h_{n,V}^{(t-1)}, S_1^{(t)}\} \tag{35}$$

$$\{h_{1,V}^{(t)}, h_{2,V}^{(t)}, ..., h_{n,V}^{(t)}, S_L^{(t)}\} = R_V^L \tag{36}$$

Because of the centralized training distributed execution mode, it is necessary to calculate the global value function, and the global function can be calculated by the individual value function. The global action value function can be calculated using all individual action value functions. The formula is as follows (Equation 37):

$$V_\pi(s) = f(q_i^t, ..., q_n^t) \tag{37}$$

where f denotes the credit allocation function for each UAV. In this study, the calculation method in Value-Decomposition Networks (VDN) is used, and the summation function is used to solve the credit allocation problem of UAVs. The calculation formula is shown in Equation (38):

$$f(q_i^t, ..., q_n^t) = \sum_{i=1}^{n}(q_i^t) \tag{38}$$

4.3.3. Transformer internal attention calculation

The Transformer in the traditional sequence modeling task is input by position coding, and the hidden representation of the coding is automatically regressed and decoded. This article uses a mechanism with a lower triangular matrix to calculate attention (Equations 39–42):

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}} + M)V \tag{39}$$

$$R_i^1 = \{h_i^{(t-1)}, O_i^{(t)}\} \tag{40}$$

$$Q_i^k, K_i^k, V_i^k = L_{Q,K,V}(R_i^k) \tag{41}$$

$$R_i^{(k+1)} = Attention(Q_i^k, K_i^k, V_i^k) \tag{42}$$

where M is the mask matrix, which ensures that the input at time step t can only be associated with the input from $< 1, 2, ..., t-1 >$. Where K, Q, V represent keys, queries, values respectively. d_k represents a scaling influence factor equal to the key dimension.

$$V_i(S_i^{(t)}, h_{i,V}^{(t-1)}, \phi_i) = P(R_i^L, u_i^{(t)}) \tag{43}$$

where L represents a linear function used to calculate Q, K, V. The output R_i^L of the last Transformer layer is mapped to the space of V_i, using a linear function P to implement the prediction, as shown in Equation (43):

4.3.4. Algorithm description

The Trans-UTPA algorithm is based on the MAPPO algorithm and is also composed of two independent networks, namely, the policy network $\pi_\theta(a_i|o_i)$ with the parameter θ and the evaluation network $V_\phi(s)$ with the parameter ϕ. First collect trajectory $\tau = \{\tau_1, \tau_2, ..., \tau_n\}$ where UAV i ' s trajectory $\tau_i = \{s_i^{(t)}, o_i^{(t)}, a_i^{(t)}\}$, $t \in (1, T)$, T is the training duration.

Initialize the hidden layer state of actor and critic, input the observation value into the actor network to obtain the optional action, and input the state value into the critic network to obtain the reward value. Perform action a_t, observe r_t,s_{t+1},o_{t+1}, then write it into τ_i, and calculate the dominance function $\hat{A}$ through GAE. The calculation formula is shown in Equation (44):

$$\hat{A}(s, a_i) = \sum_t \gamma^t r(s, a_i) - V_\phi(s) \tag{44}$$

The calculation formula of importance weight is as follows (Equation 45):

$$\omega = \pi_\theta(o_i, a_i)/\pi_{\theta old}(O_i, a_i) \tag{45}$$

Update the actor network parameter θ_i of UAV i (Equation 46):

$$\theta_i = argmaxE_t[min(\omega\hat{A}_t, clip(\omega, 1-\varepsilon, 1+\varepsilon)\hat{A}_t] \tag{46}$$

Update the parameter ϕ of the critic network of the UAV (Equation 47):

$$\phi = argmin\frac{1}{2}[\sum_t(\gamma^t r(s, a_i) - V_\phi(s)]^2 \tag{47}$$

Based on the above content, the specific UAV swarm cooperative path planning algorithm based on Trans-UTPA is shown in Algorithm 1.

```
BEGIN
    Initialize UAV parameters, power consumption parameters, action space dimension $action\_dim$ and observation space dimension $obs\_dim$;
    Initialize training parameters Parameter $\theta$ of actor network $\pi(a|o)$, $\phi$ parameter of critic network $V(s)$, learning rate $\alpha$, discount factor $\gamma$, number of training rounds;
    Initializes the hidden-layer state $h_\pi$ of the actor, the hidden-layer state $h_V$ of the critic;
    for $i = 1, 2, 3, ..., episodes$ do:
        Initialize data buffer $D = \{\}$
        for $j = 1$ to batch_size $do$:
            Set empty list $l = []$
            The observation value is input into the actor network to get the optional action $a_t$, and the state value is input into the critic network to get the value v.
            Execute action $a_t$ to get reward value $r_t$, observation value $o_{t+1}$ for the next moment, global state value $s_{t+1}$;
            Write the resulting value $a_t, r_t, o_{t+1}, s_{t+1}$ to the list $l$
            According to $l$, calculating Dominance Function $\hat{A}$ by Formula $A(s, a_i) = \sum_t (\gamma^t) r(s, a_i) - V_\phi(s)$;
            Write $l$ into buffer $D$;
        end for;
        Random sampling mini-batch $H$ from $D$;
        Calculate the loss function $policy\_loss$ of the actor from the data in H and the loss function $value\_loss$ of the critic;
        Update parameter $\theta_i$of actor network of UAV $i$, update parameter $\phi$ of critic network
    Select random mini-batch as $b$ from $B$
    end for
END
```

Algorithm 1. Cooperative path planning algorithm for UAV swarm based on Trans-UTPA (TUTPA).

5. Experiment

5.1. PSO-based clustering

The data set used in this paper is 7,366 scenic spot location information obtained from Mafengwo Tourism Online, which is represented in the coordinate map. As shown in Figure 4, the abscissa represents the longitude value, and the ordinate is the dimension value.

Each location information in the dataset includes attributes: longitude, dimension, and label (region attribution). The location information of the attractions is shown in Table 2.

In the experiment, we use PSO algorithm to cluster the obtained location information. Here $K = 5$, after 10,000 iterations, 1,666 attractions are clustered into 5 clusters. The results are shown in Figure 5. The circles of different colors are used to represent the attractions belonging to different clusters, and the large black circle represents the cluster center.

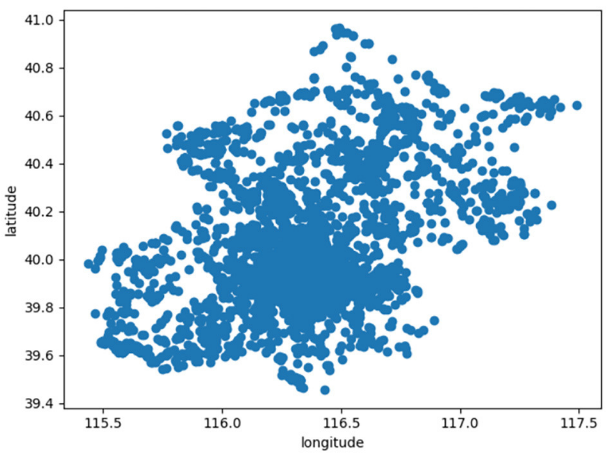

FIGURE 4
It shows the location distribution of scenic spots in Liaoning.

TABLE 2 Sample location information for attractions.

Longitude	Latitude	Region ownership
116.47545	40.96251398	Beijing
116.4757993	40.96299956	Beijing
116.4862425	40.96156428	Beijing
123.46856	41.808196	Shenyang
123.407483	41.799246	Shenyang
123.362155	41.786272	Shenyang

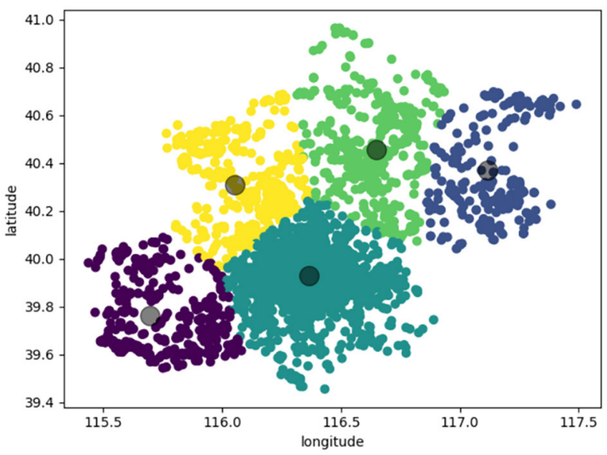

FIGURE 5
PSO algorithm attractions clustering distribution map.

5.2. target-POI recommendation based on PSO

5.2.1. Acquisition of experimental data

In the emergency rescue scenario, it is necessary to comprehensively consider the distance between the rescue crowd and

TABLE 3 Sample visit statistics based on attraction information.

ID	Name	Location	Visits	The latest statistical time
1	Shenyang expo park	41.863424, 123.645846	2,999	May 2022
2	Guanmenshan national forest park	41.121535, 124.188346	1,355	March 2022
3	Benxi water cave	41.304582, 124.08402	1,256	June 2022
4	Shenyang imperial palace	41.803551, 123.462138	1,000	October 2022

the UAV, the crowd density and the rescue time in the rescue area, and reasonably dispatch the UAV to complete the emergency rescue task. In this experiment, the scenic spot data set in the Mafengwo tourism network is used. The target-POI position is represented by the location information of the scenic spot, and the number of visits represents the crowd density of the position. The attributes of each attraction include: the number of visits to the attraction, the location of the attraction (latitude and longitude), the name of the attraction, and the ID of the attraction. Access statistics based on attractions information are shown in Table 3.

5.2.2. Evaluation criteria

In order to evaluate the performance of the algorithm for target-POI recommendation, we use Normalized Discounted Cumulative Gain (NDCG) and Hit Ratio (HR) to evaluate the performance of the algorithm in top-K recommendation.

The calculation formula of HR is shown in Equation (48). $hits(i)$ records the predicted score of the sample, whether it is in the recommended first K, is 1, otherwise it is 0, and N represents the total number of samples in the test set.

$$HR = \frac{\sum_{i=0}^{N}(hits(i))}{N} \tag{48}$$

NDCG is a measure of ranking, which tends to evaluate the order of recommendation. The calculation formula is shown in Equation (49).

$$NDCG_K = \frac{DCG_K}{IDCG_K} \tag{49}$$

$$DCG_K = \sum_{i=0}^{K}(\frac{y_i}{\log_2^{i+1}}) \tag{50}$$

The DCG obtained by the Equation (50) represents the cumulative gain of the loss, in which the ranking order factor will be considered. Starting from the first item in the obtained ranking, each item is multiplied by the decreasing coefficient, so that the top item gain is higher, and the following items will have a loss.

5.2.3. Analysis of experimental results

Outputs the heat ranking of attractions in the area based on the area 's location data and number of visits to the attraction. Figure 6 shows the changes of HR and NDCG in each iteration of the model when taking top-10 recommendation, and the total number of iterations of the experiment is 30. It can be seen from Figure 6A that the growth rate of the model is obvious in the first 10 iterations. When iterating to 15 times, the model begins to converge and HR is 0.7347. When iterating to the 25 th time, the fluctuation range of HR is small and stable in the range of [0.7586, 0.7607]. As can be seen from Figure 6B, NDCG also changes greatly at the initial stage of iteration. When iterating to 15 times, the NDCG is 0.4379. When iterating to 25 times, the NDCG fluctuates between [0.4422, 0.4434].

Figure 7 shows the numerical comparison of HR and NDCG at different top-K recommendations. It can be seen from Figure 7A that HR increases greatly before is <3, and then the HR gap between top-K gradually narrows with the increase of K. HR fluctuates around 0.75. Similarly, it can be seen from Figure 7B that the NDCG increases greatly before is <3, and then the NDCG gap between top-K gradually narrows with the increase of. When, NDCG fluctuates around 0.44.

From the results in Figures 6, 7, it can be seen that the HR and NDCG values are higher when using the algorithm for top-K recommendation. When K is 10, HR can reach 0.68 and NDCG can reach 0.45. It shows that the algorithm can be used to predict the popularity of scenic spots and get the target-POI ranking that UAVs need to prioritize.

5.3. Experiment of UAV swarm cooperative path planning algorithm based on Trans-UTPA

5.3.1. Experimental environment

The experimental environment of this algorithm includes a training environment and a test environment. The test environment is a given map_ size, the number of UAVs, the number of TaskPOIs, the starting position of the UAV, and the generation density of obstacles to generate obstacles. Randomly distributed test scenarios based on global environmental data. The purple dots are clustered to obtain the target-POI area marked by a red circle. The square red area is the target area for all UAVs to finally network. The following is the environmental scene map of the unreleased UAV after clustering. Figure 8 shows the training environment under different specifications, where Figures 8A, B are the environment maps of 32*32 and 64*64 under fixed scenarios, respectively. Figure 8C shows a 64*64 environment map with an obstacle generation density of 0.1. Figure 8D shows a 64 * 64 environment map with an obstacle generation density of 0.3. Obstacles in the environment are represented by gray squares, and green and blue circles represent UAVs. In order to ensure that the scenes used by various algorithms are consistent, the scene files used by the USCTP algorithm (Li J. et al., 2022) are saved to generate the corresponding fixed scenes, and then compared with the Trans-UTPA algorithm.

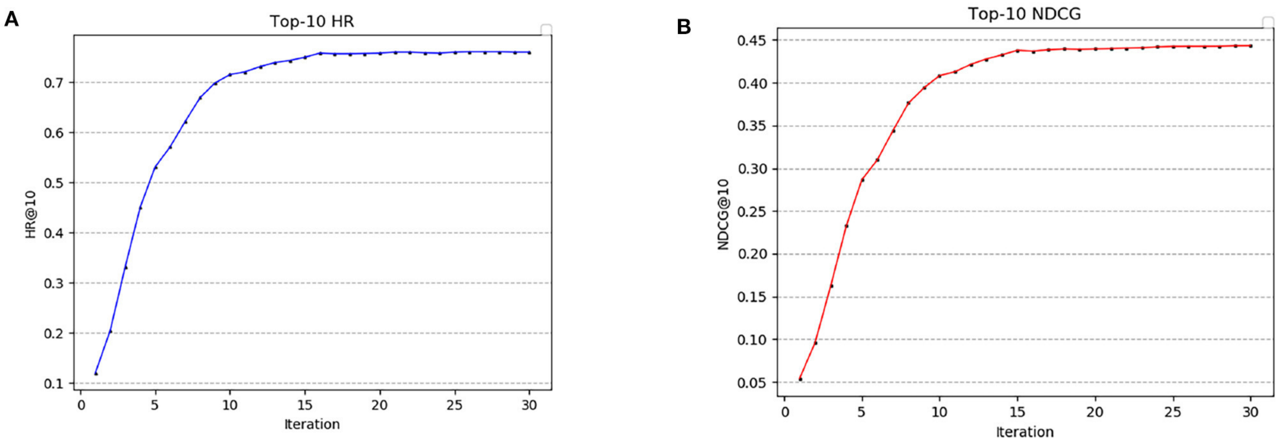

FIGURE 6
Hit rate and normalized cumulative gain of top-10. **(A)** Top-10 HR. **(B)** Top-10 NDCG.

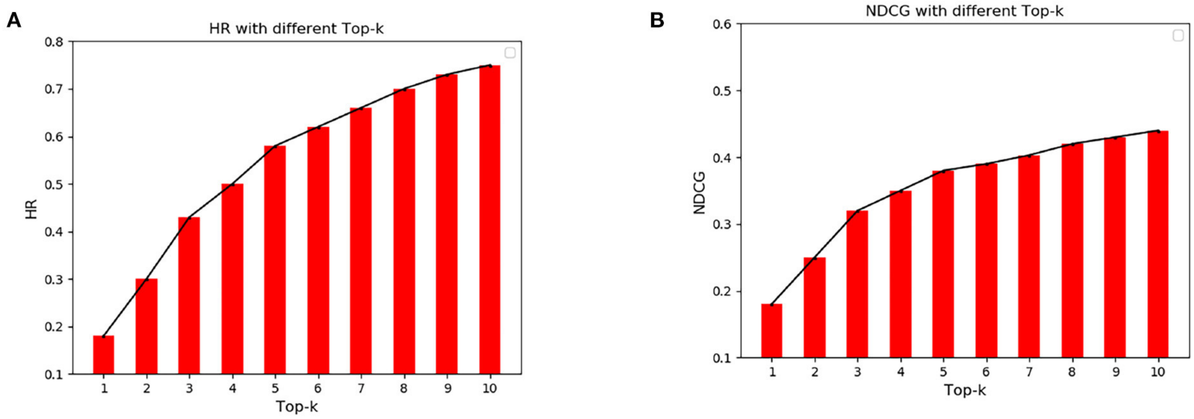

FIGURE 7
Cumulative Gain of Hit Rate and Normalized Discount at different top-K. **(A)** Hit rate under different top-k. **(B)** Cumulative increase of normalized loss for different top-k.

5.3.2. Evaluation criteria

We mainly consider the success rate of the UAV reaching the target position, the number of collisions during the movement and the average reward of the algorithm.

(1) Success rate: The probability of each UAV reaching the target position accurately is counted, that is, whether they have completed the task of reaching the target position within a given time or number of times.
(2) The number of collisions: the action conflict between dynamic agents and the conflict between agents and static obstacles, that is, during the movement of the UAV, the number of collisions between the UAV and other UAVs in an episode or between the UAV and obstacles in the environment is counted.
(3) Average reward: each UAV relies on obstacle avoidance during flight, performs access tasks, and controls power consumption to obtain corresponding rewards and penalties. After training several times, the average reward of the UAV reaches a convergent state, indicating that the UAV has trained a rough trajectory and will not be blindly explored again.

5.3.3. Analysis of experimental results

The model compares the success rate of reaching the target position, the number of collisions and the average reward of the two algorithms in the test environment 32 * 32 map with different obstacle densities. Figure 9 shows the comparison of the success rate of the UAV swarm based on the two algorithms to reach the target position under different obstacle densities. Figure 9A is the success rate of the UAV reaching the target position when the density is 0.1, and Figure 9B is the success rate of the UAV reaching the target position when the density is 0.3. It can be seen from the figure that under the Trans-UTPA algorithm, the success rate of the UAV swarm to reach the target position has always been higher than that of the USCTP algorithm. With the increase of the number of iterations, the more than part reaches about 8%.

Figure 10 shows the number of collisions of the two algorithms in the flight process of the UAV under different obstacle densities. Figure 10A is the number of collisions of the UAV in the case of a density of 0.1. It can be seen from the graph that the number of collisions of the Trans-UTPA algorithm has been lower than the USCTP algorithm. Even in the 300–400 rounds, the peak is still lower

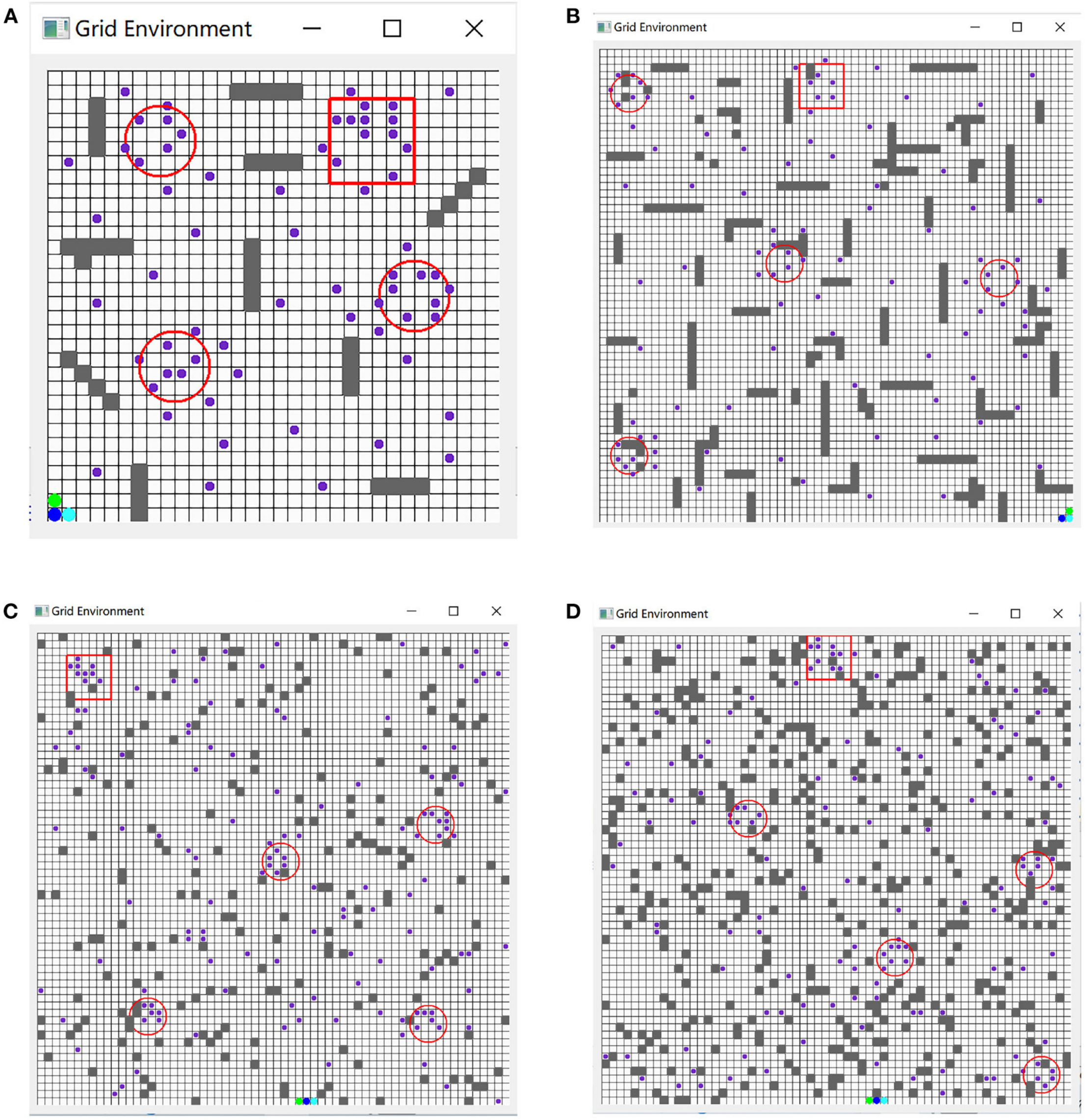

FIGURE 8
Scene maps of different specifications. **(A)** 32*32 map of fixed scene. **(B)** 64*64 map of fixed scene. **(C)** 64*64 map with a random scene with obstacle density = 0.1. **(D)** 64*64 map with a random scene with obstacle density = 0.3.

than the USCTP algorithm, and the convergence speed is faster than the USCTP algorithm. Figure 10B shows the number of collisions of UAVs at a density of 0.3. Compared with the case of obstacle density of 0.1, the number of collisions of UAVs under the two algorithms is significantly increased, but it can be seen from the figure that the number of collisions of UAVs based on Trans-UTPA algorithm is the lowest.

Figure 11 shows the average rewards obtained by the UAV swarms of the two algorithms during flight under different obstacle densities. Figure 11A is the average reward value of the two algorithms under the density of 0.1. It can be seen from the figure that the average reward value based on the Trans-UTPA algorithm has always been higher than the USCTP algorithm. Figure 11B shows the average reward value of the two algorithms when the density is 0.3. It can be seen from the figure that the two algorithms are not much different. In the first 1,000 rounds, the USCTP algorithm is higher. As the number of rounds increases, Trans-UTPA is gradually higher than the USCTP algorithm.

5.3.4. Energy consumption experimental results

In the energy consumption part, we use simulation experiments to verify the Trans-UTPA algorithm. The relevant parameters in the experiment are shown in Table 4.

The relevant parameters in the Trans-UTPA experiment are shown in Table 5.

In the simulation environment, we trained 10,000 times to explore the feasibility of using Trans-UTPA algorithm energy consumption model for UAV group.

Figure 12 shows the comparison of POI visits between Trans-UTPA and A *. The UAV uses Trans-UTPA (Trans-UTPA is abbreviated as TUTPA) algorithm and A * algorithm to complete

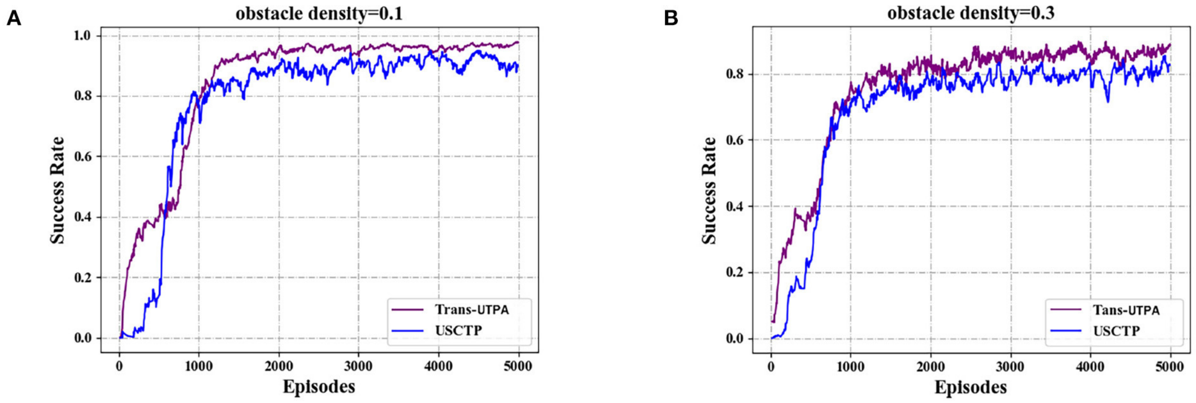

FIGURE 9
Comparison of success rates of UAV swarms reaching target locations. **(A)** Success rate with obstacle density = 0.1. **(B)** Success rate with obstacle density = 0.3.

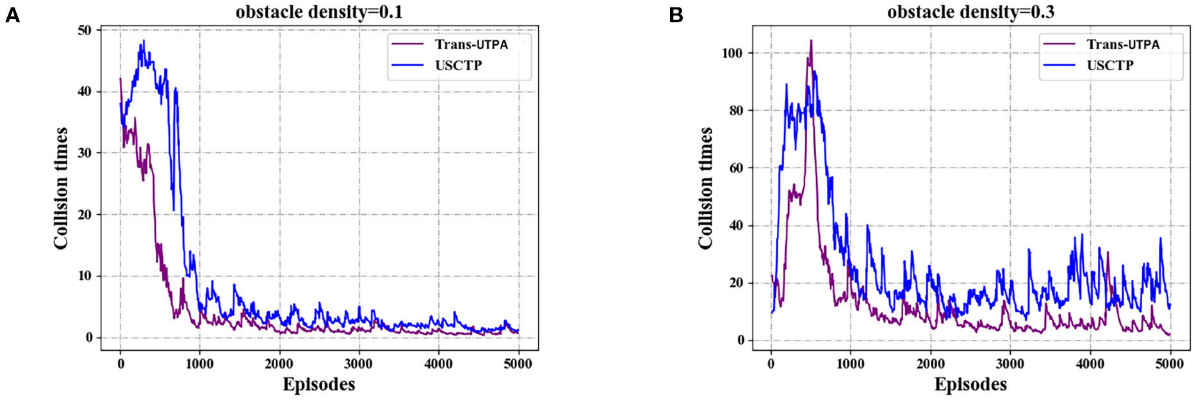

FIGURE 10
Comparison experiment of collision times of UAVs. **(A)** Collision times with obstacle density = 0.1. **(B)** Collision times with obstacle density = 0.3.

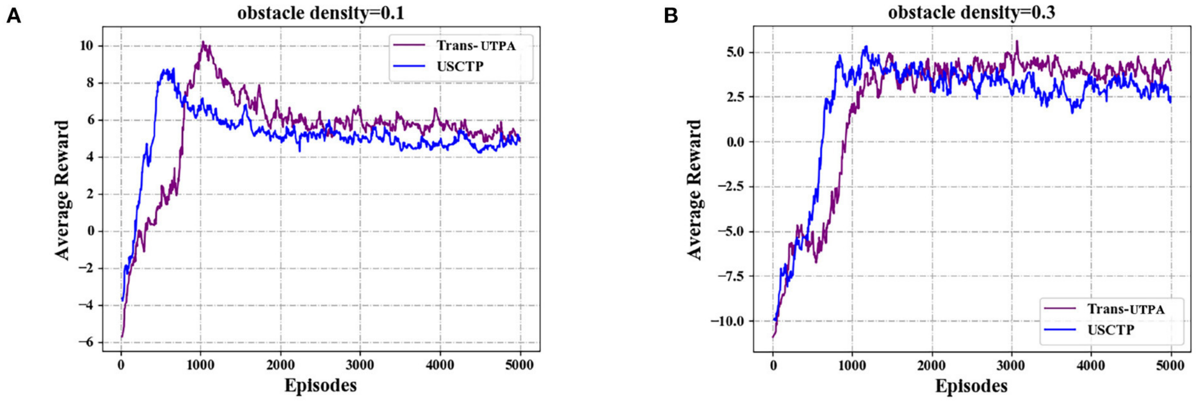

FIGURE 11
Comparison experiment of the average rewards. **(A)** Average rewards with obstacle density = 0.1. **(B)** Average rewards with obstacle density = 0.3.

the mission in the same flight area. From the figure, we can find that the TUTPA algorithm can access more target-POI at the same time, because the TUSCTP algorithm can mobilize the UAV to access more data acquisition areas during the process of going to the network communication area.When accessing the same number of target-POIs, the TUTPA algorithm takes less time than the A *

TABLE 4 Statistics of UAV simulation parameters.

Parameter	Value
E	$5.3 * 10^4 J$
d_{safe}	$400m$
c_1, c_2	0.0661, 15.97
σ	0.2
Bit	$1Mbit$
V_{MAX}	$10m/s$
u	3
α	2
E_m^B	$1J/(m^\alpha.Mbit)$
$P_{collect}$	$0.37J/s$
$E_{hover}^i(t)$	$10J/min$
Q	$10J/r$
d_{com}	$20m$
E_{up}^i	$2.5Jeachtime$
E_{down}^i	$2.5Jeachtime$

TABLE 5 The relevant parameters in the Trans-UTPA experiment.

Description	Value
Learning rate of the actor network	0.001
Learning rate of the critic network	0.001
Max clipped value loss	0.2
Entropy regularization coefficient	0.5
Value function update coefficient	0.5
Batch size	512

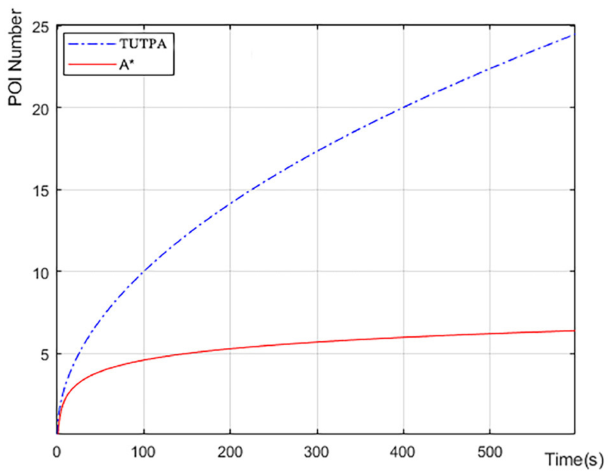

FIGURE 12
Average number of POI visits for multiple experiments.

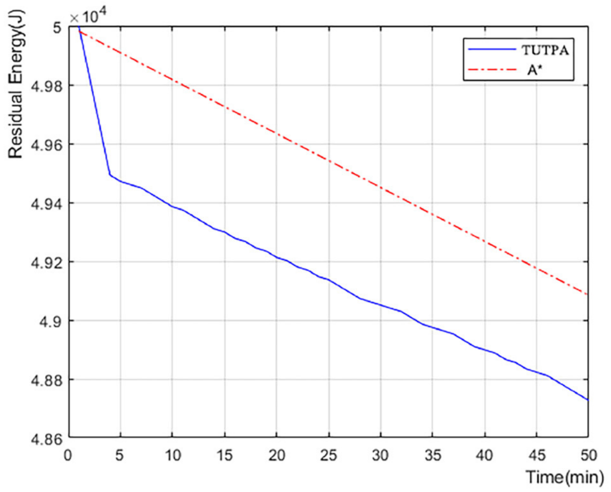

FIGURE 13
The remaining power of a single UAV after performing data acquisition on target-POI.

algorithm, which better shows that the TUTPA algorithm can cover more POIs and work efficiently.

Figure 13 we recorded the remaining power of a single UAV starting to perform data acquisition tasks on target-POI. According to the data in Figure 12, the number of target-POI that A * algorithm can access is 6 in 600 s (10 min), and the number of target-POI that TUTPA algorithm can access is 25, which is about 4 times that of A * algorithm. However, in this process, TUTPA consumes less energy than A * in a target-POI area on average. It is concluded that the energy consumption of TUTPA algorithm has certain advantages and can achieve energy saving.

6. Conclusion

With the coming of 6G network, UAVs will provide more and more help in emergency disaster relief. We consider the distribution of the rescue area, the type of mission, and the flight characteristics of the UAV. Firstly, according to the distribution of the crowd, the PSO algorithm is used to cluster the target-POI of the task area, and the neural collaborative filtering algorithm is used to prioritize the target-POI. Then we design a Trans-UTPA algorithm and introduce Transformer mechanism to sequence modeling. The UAV completes flight movements (horizontal flight, ascent and descent) and emergency missions (data acquisition and networked communications) in a three-dimensional space, sharing information about the global UAV. The multi-UAVs cooperative flight constraints and multi-UAVs green energy consumption calculation model are designed. The multi-agent reinforcement learning is used to design the flight route according to the maximum global reward. The experimental results show that the Trans-UTPA algorithm makes the success rate of each UAV reaching the target position, the number of collisions and the average reward performance of the algorithm further improve than the USCTP algorithm. Among them, the average reward algorithm exceeds USCTP algorithm 13%, the number of collisions reduced by 60%. Compared with the heuristic algorithm, it can cover more target-POI, and has less energy consumption than the heuristic algorithm. There are still some defects in the algorithm, and there is a lack of three-dimensional actual environment simulation experiments. In the future, we will continue to study along this point.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Author contributions

JL, SC, XL, RY, and XW conceived the idea of the study and interpreted the results. JL, XL, and SC analyzed the data. SC wrote the paper. All authors discussed the results and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Key Research and Development Projects (2019YFB1802800), the Liaoning Province Science and Technology Fund Project (2020MS086), the Shenyang Science and Technology Plan Project (20206424), the Fundamental Research Funds for the Central Universities (N2116014), China University Industry-University-Research Innovation Fund (2021ITA10011), and the National Natural Science Foundation of China (61872073).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ahmed, S., Mohamed, A., Harras, K., Kholief, M., and Mesbah, S. (2016). "Energy efficient path planning techniques for uav-based systems with space discretization," in *IEEE Wireless Communications and Networking Conference (IEEE)*, 1–6. Available online at: https://ieeexplore.ieee.org/abstract/document/7565126

Fu, Z., Yu, J., Xie, G., Chen, Y., and Mao, Y. (2018). A heuristic evolutionary algorithm of uav path planning. *Wireless Commun. Mobile Comput.* 2018, 2851964. doi: 10.1155/2018/2851964

Jin, T., and Yang, X. (2021). Monotonicity theorem for the uncertain fractional differential equation and application to uncertain financial market. *Elsevier* 190, 203–221. doi: 10.1016/j.matcom.2021.05.018

Jin, T., Yang, X., Xia, H., and Ding, H. (2021). Reliability index and option pricing formulas of the first-hitting time model based on the uncertain fractional-order differential equation with caputo type. *World Sci.* 29, 2150012. doi: 10.1142/S0218348X21500122

Kordos, M., a., Blachnik, M., and Scherer, R. (2022). Fuzzy clustering decomposition of genetic algorithm-based instance selection for regression problems. *Inf. Sci.* 587, 23–40. doi: 10.1016/j.ins.2021.12.016

Kowalski, P. A., and Jeczmionek, E. (2022). Parallel complete gradient clustering algorithm and its properties. *Inf. Sci.* 600, 155–169. doi: 10.1016/j.ins.2022.03.087

Krishna, E. R., Devarakonda, N., Al-Shamri, M. Y. H., and Revathi, D. (2022). "A novel hybrid clustering analysis based on combination of k-means and pso algorithm," in *Data Intelligence and Cognitive Informatics* (Singapore: Springer), 139–150.

Li, J., Liu, X., Han, G., Cao, S., and Wang, X. (2022). Taskpoi priority based energy balanced multi-uavs cooperative trajectory planning algorithm in 6g networks. *IEEE Trans. Green Commun. Netw.* 2022, 3187097. doi: 10.1109/TGCN.2022.3187097

Li, K., Ni, W., Wang, X., Liu, R. P., Kanhere, S. S., and Jha, S. (2015). Energy-efficient cooperative relaying for unmanned aerial vehicles. *IEEE Trans. Mobile Comput.* 15, 1377–1386. doi: 10.1109/TMC.2015.2467381

Li, W., Wang, L., Zou, A., Cai, J., He, H., and Tan, T. (2022). Path planning for uav based on improved prm. *Energies* 15, 7267. doi: 10.3390/en15197267

Lipare, A., Edla, D. R., and Dharavath, R. (2021). Fuzzy rule generation using modified pso for clustering in wireless sensor networks. *IEEE Trans. Green Commun. Network.* 5, 846–857. doi: 10.1109/TGCN.2021.3060324

Liu, W., Yang, Y., and Hao, J. (2022). "Design and research of a new energy-saving uav for forest fire detection," in *2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI)* (Changchun: IEEE), 1303–1316.

Ma, L., Cheng, S., and Shi, Y. (2021d). Enhancing learning efficiency of brain storm optimization via orthogonal learning design. *IEEE Trans. Syst. Man Cybern. Syst.* 51, 6723–6742. doi: 10.1109/TSMC.2020.2963943

Ma, L., Huang, M., Yang, S., Wang, R., and Wang, X. (2021a). An adaptive localized decision variable analysis approach to large-scale multiobjective and many-objective optimization. *IEEE Trans. Cybern.* 51, 6723–6742. doi: 10.1109/TCYB.2020.3041212

Ma, L., Li, N., Guo, Y., Wang, X., Yang, S., Huang, M., et al. (2021b). Learning to optimize: reference vector reinforcement learning adaption to constrained many-objective optimization of industrial copper burdening system. *IEEE Trans. Cybern.* 52, 12698–12711. doi: 10.1109/TCYB.2021.3086501

Ma, L., Wang, X., Wang, X., Wang, L., Shi, Y., and Huang, M. (2021c). Tcda: Truthful combinatorial double auctions for mobile edge computing in industrial internet of things. *IEEE Trans. Cybern.* 21, 4125–4138. doi: 10.1109/TMC.2021.3064314

Mughal, U. A., Ahmad, I., Pawase, C. J., and Chang, K. (2022). "Uavs path planning by particle swarm optimization based on visual-slam algorithm," in *Intelligent Unmanned Air Vehicles Communications for Public Safety Networks*, eds Z. Kaleem, I. Ahamad, and T. Q. Duong (Singapore: Springer). doi: 10.1007/978-981-19-1292-4_8

Pu, Q., Gan, J., Qiu, L., Duan, J., and Wang, H. (2022). An efficient hybrid approach based on pso, abc and k-means for cluster analysis. *Multimed. Tools Appl.* 81, 19321–19339. doi: 10.1007/s11042-021-11016-6

Qin, Z., Li, A., Dong, C., Dai, H., and Xu, A. (2018). "Fair-energy trajectory plan for reconnaissance mission based on uavs cooperation," in *2018 10th International Conference on Wireless Communications and Signal Processing (WCSP)* (Hangzhou: IEEE).

Sathyamoorthy, M., Kuppusamy, S., Dhanaraj, R. K., and Ravi, V. (2022). Improved k-means based q learning algorithm for optimal clustering and node balancing in wsn. *Wireless Pers. Commun.* 122, 2745–2766. doi: 10.1007/s11277-021-09028-4

Sharma, S., and Kaushik, B. (2021). A survey on nature-inspired algorithms and its applications in the internet of vehicles. *IEEE Access* 34, e4895. doi: 10.1002/dac.4895

Shinde, P., Kudalkar, G., Pawar, S., Tiwari, H., and Khairnar, H. (2022). "Accident hotspot detection by db-scan clustering," in *ICT Analysis and Applications. Lecture Notes in Networks and Systems, Vol. 314* (Singapore: Springer), 413–418.

Tang, S., Thomas, J., and Kumar, V. (2018). Hold or take optimal plan (hoop): a quadratic programming approach to multi-robot trajectory generatio. *Int. J. Rob. Res.* 37, 1062–1084. doi: 10.1177/0278364917741532

Tian, C., Jin, T., Yang, X., and Liu, Q. (2022). Reliability analysis of the uncertain heat conduction model. *Comput. Math. Appl.* 119, 131–140. doi: 10.1016/j.camwa.2022.05.033

Wang, C., Wang, W., Gao, Y., and Li, X. (2022). Parameters optimization of multipass milling process by an effective modified particle swarm optimization algorithm. *Discrete Dyn. Nat. Soc*. 2022, 8545739. doi: 10.1155/2022/8545739

Wang, F., Jiang, D., Qi, S., Qiao, C., and Shi, L. (2021). A dynamic resource scheduling scheme in edge computing satellite networks. *Mobile Netw. Appl*. 26, 597–608. doi: 10.1007/s11036-019-01421-5

Wang, Z., Yu, X., Feng, N., and Wang, Z. (2014). An improved collaborative movie recommendation system using computational intelligence. *J. Vis. Lang. Comput*. 34, 667–675. doi: 10.1016/j.jvlc.2014.09.011

Wangsheng, F., Chong, W., and Ruhua, Z. (2021). Application of simulated annealing particle swarm optimization in complex three-dimensional path planning. *IOP Publishing* 1873, 012077. doi: 10.1088/1742-6596/1873/1/012077

Wu, Q., Zeng, Y., and Zhang, R. (2018). Joint trajectory and communication design for multi-uav enabled wireless networks. *IEEE Trans. Wireless Commun*. 17, 2109–2121. doi: 10.1109/TWC.2017.2789293

Wu, Q., and Zhang, R. (2018). Common throughput maximization in uav-enabled ofdma systems with delay consideration. *IEEE Trans. Commun*. 66, 6614–6627. doi: 10.1109/TCOMM.2018.2865922

Yang, Z., Pan, C., Wang, K., and Shikh-Bahaei, M. (2019). Energy efficient resource allocation in uav-enabled mobile edge computing networks. *IEEE Trans. Wireless Commun*. 18, 4576–4589. doi: 10.1109/TWC.2019.2927313

Yao, P., Wang, H., and Su, Z. (2016). Cooperative path planning with applications to target tracking and obstacle avoidance for multi-uavs. *Aerospace Sci. Technol*. 54, 10–22. doi: 10.1016/j.ast.2016.04.002

Zhang, H., Yao, Q., Kwok, J. T., and Bai, X. (2022). Searching a high performance feature extractor for text recognition network. *IEEE Trans. Pattern Anal. Mach. Intell*. 2022, 1–15. doi: 10.1109/TPAMI.2022.3205748

Zhang, W., Wang, W., Zhai, H., and Li, Q. (2021). "A deep reinforcement learning method for mobile robot path planning in unknown environments," in *2021 China Automation Congress (CAC)* (Beijing: IEEE), 5898–5902.

Zhu, X., Wang, L., Li, Y., Song, S., Ma, S., Yang, F., et al. (2022). Path planning of multi-uavs based on deep q-network for energy-efficient data collection in uavs-assisted iot. *Vehicular Commun*. 2022, 100491. doi: 10.1016/j.vehcom.2022.100491