

Research Article

Multi-UAV Collaborative Path Planning Method Based on Attention Mechanism

Tingzhong Wang¹, Binbin Zhang¹, Mengyan Zhang², and Sen Zhang²

¹School of Information Engineering, Luoyang Normal University, Luoyang 471934, China

²College of Information Engineering, Henan University of Science and Technology, Luoyang 471023, China

Correspondence should be addressed to Tingzhong Wang; wangtingzhong@lynu.edu.cn

Received 19 May 2021; Revised 17 July 2021; Accepted 30 August 2021; Published 14 September 2021

Academic Editor: Dongsheng Guo

Copyright © 2021 Tingzhong Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at the problem that traditional heuristic algorithm is difficult to extract the empirical model in time from large sample terrain data, a multi-UAV collaborative path planning method based on attention reinforcement learning is proposed. The method draws on a combined consideration of influencing factors, such as survival probability, path length, and load balancing and endurance constraints, and works as a support system for multimachine collaborative optimizing. The attention neural network is used to generate the cooperative reconnaissance strategy of the UAV, and a large amount of simulation data is tested to optimize the attention network using the REINFORCE algorithm. Experimental results show that the proposed method is effective in solving the multi-UAV path planning issue with high real-time requirements, and the solving time is less than the traditional algorithms.

1. Introduction

Earthquake usually causes serious damages to roads and buildings in the local area. The relief staffs can grasp the local disaster situation at the first time and then take effective rescue measures which play a vital role in the post-emergency rescue and road repair after earthquake.

In recent years, with the rapid development of UAV technology, it is possible to quickly acquire and evaluate the disaster situation [1, 2]. When the disaster occurs, the first time to carry out on-site investigation by using UAVs has become a universally accepted emergency measure.

Due to its simple operation and sturdy maneuverability, UAV has, moreover, been widely utilized in several fields such as military, agriculture, and environmental protection [3–8] and has attracted the attention of countries around the world. Within the multi-UAV collaborative reconnaissance path planning framework, their distributed characteristic notably improves the working ability and robustness of the system. And, multi-UAV collaborative path planning strategy, as the basic of collaborative exploration, determines

the utilization degree of high quality for current sources by aircraft. In the real environment, restricted by various environmental factors, the collaborative path planning method of multi-UAV can be regarded as an NP combinatorial optimization problem with a number of constraints [9]. At present, the main classical models for the multi-UAV cooperative reconnaissance problem include multitraveling salesman model (MTSP), mixed linear integer programming model (MILP), and vehicle scheduling and path planning model VRP [9]. These models are solved by heuristic algorithms such as genetic algorithm [10], simulated annealing algorithm, and evolutionary algorithm [11]. However, the traditional model cannot completely describe the constraints of multi-UAV cooperative reconnaissance missions. Under complex environmental conditions and in the face of constantly changing terrain data, the above-mentioned heuristic algorithm needs to be reoptimized and solved, so it has poor adaptability and is unable to give the corresponding solution quickly [12]. For the model with variable environmental information, Yuan [13] proposed a genetic algorithm based on the bipartite successive revision

method for solving the problem of which the distance matrix is dynamic over time and Shi [14] investigated the algorithm for the VRP problem with variable demand. In fact, all these models simply introduced variables in the original specific problem, which only were optimized and improved for the traditional heuristic algorithm. They have not found a reasonable internal law of path planning and could not avoid the problem of needing to reoptimize iterations of the traditional heuristic algorithm, and the real-time performance was not high.

With the rapid development of deep learning, neural networks are gradually being used to deal with combinatorial optimization problems. Vinyals et al. proposed pointer network (PN) [15] based on end-to-end model and used supervised learning to train networks and solve TSP problems. However, the effectiveness of supervised learning depends on the training sample data, which limits PN networks to find better solutions to problems. Nazari et al. used reinforcement learning to train the end-to-end network, which greatly improved the effectiveness compared with the PN network on the TSP problem [16].

In view of the shortcomings of traditional heuristic algorithm, this paper proposes a reinforcement learning collaborative algorithm based on the attention model. It uses neural networks to represent the collaborative reconnaissance strategy of UAVs, introduces a large amount of simulation data and uses the REINFORCE learning algorithm [17] to pretrain the model, and explores the inherent characteristics between the site model and optimal path. The neural network is used to represent the cooperative reconnaissance strategy of the UAV, and a large amount of simulation data is introduced to pretrain the model using the REINFORCE algorithm [17] for finding out the internal law between data (site model) and optimized path. Experimental results show that the algorithm based on the attention model can give an optimized solution more quickly when solving similar problems with the same location distribution, number of nodes, and constraints for UAVs' collaborative reconnaissance issues. No reoptimization iterations are required during the computation, and its solution time and strain capacity are significantly better than traditional heuristic algorithms.

2. Multi-UAV Collaborative Reconnaissance Model

For multi-UAV collaborative reconnaissance mission, the threat and related location information of the target regions have been roughly assessed before the investigation, and multiple UAVs need to conduct detailed reconnaissance of multiple target regions in the affected area.

It is assumed that N_V UAVs are required to detect the target regions with a number of N_T in detail. Now, the endurance range of UAVs and the threat level of targets are comprehensively considered to set up model of the reconnaissance mission.

In this paper, the path planning issue of multi-UAV collaborative reconnaissance mission is described as follows:

Base: it is the location from which the UAVs depart and the location to which they return after completing the reconnaissance mission, with coordinates (x_0, y_0) .

UAV: assume that there are N_V UAVs, which are represented by set $V = \{v_1, \dots, v_{N_V}\}$, and the maximum endurance range of UAVs is L .

Detection target: there are N_T targets to be detected, which are represented by set $T_0 = \{t_1, \dots, t_{N_T}\}$ so that $T = \{0\} \cup T_0$, where 0 represents base and elements in T are called nodes, which represent local targets to be detected. The coordinates of the target $i \in T_0$ is (x_i, y_i) , and the survival probability of the UAV when detecting the target i is $P_i \in [0.7, 0.9]$.

Path set: $\pi^v = \{(i, j) | i, j \in T\}$ represents the detection path of UAV, where v is according to the basic principle of safe flight after considering various terrain conditions.

Decision variables of multi-UAV collaborative reconnaissance mission are defined as

$$X_{i,j}^v = \begin{cases} 1, & \text{UAV } v \text{ from } i \text{ fly to } j, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Then, the multi-UAV collaborative reconnaissance model is established:

$$\min L(\pi) = a * f_1 + b * f_2 + c * f_3, \quad 0 \leq a, b, c \leq a + b + c = 1, \quad (2)$$

s.t.

$$\sum_{i=0}^{N_T} \sum_{v=1}^{N_V} X_{i,j}^v = 1 \quad \forall j = 1, \dots, N_T, \quad (3)$$

$$\sum_{j=0}^{N_T} \sum_{v=1}^{N_V} X_{i,j}^v = 1 \quad \forall i = 1, \dots, N_T, \quad (4)$$

$$\forall i \in R^v, j \in R^v, i \neq j, s_{i,j} \cdot X_{i,j}^v \leq L, \quad \forall v = 1 \dots N_V, \quad (5)$$

$$\min f_1 = \sum_{i=1}^{N_T} \sum_{j=1}^{N_T} \sum_{v=1}^{N_V} X_{i,j}^v \cdot \leq s_{i,j}, \quad (6)$$

$$\prod_{i=a^v}^{b^v} P_i, \quad (7)$$

$$\sum_{n_i^v}^{b^v} \prod_{i=a^v}^{b^v} P_i, \quad a^v \leq n_i^v \leq b^v, \quad (8)$$

$$\max f_2 = \sum_{n_i^v}^{b^v} \prod_{i=a^v}^{b^v} P_i, \quad (9)$$

$$\min f_3 = \text{var}(\pi^v). \quad (10)$$

Taking into account the UAV's endurance range, survival probability, and execution capability, this paper adopts a relatively simple linear combination method, which uses different proportionality coefficients to combine the individual objective functions in the process of UAVs performing reconnaissance missions. Equation (2) represents the model function for multi-UAV cooperative detection, in which f_1 , f_2 , and f_3 are the three optimized objective functions in the UAV reconnaissance model, and a , b , and c , respectively, correspond to the proportional coefficients of loss function of the three objective functions.

Formula (3) represents one visit to the node per UAV, and formula (4) represents departure from the node once. There are many restrictive factors when flying long time multiple UAVs at high altitude. For example, if the cruising range is too long, the mission may fail directly due to the limited power of the on-board battery in the first place. It may also cause communication failure due to interference from various external uncertainties, resulting in the inability to transmit valid data to the ground station. Therefore, in order to reduce the flight risk, the shorter the total flight distance of the UAV, the better it is recommended. Equation (5) indicates that the range constraint of each UAV cannot exceed its maximum range, i.e., $X_{i,j}^v \leq L$. The total flight mileage of all UAVs is described by formula (6), which is one of the optimization objectives of the reconnaissance mission.

Since each node to be detected has an inconsistent threat level to the UAV, the UAV has to pick the safest flight path while detecting multiple nodes. The defined UAV survival coverage function is shown in equation (7), where a^v is the start node of reconnaissance path for the UAV v , b^v is the end node, and P_i is the survival probability of the UAV v when detecting node i . Assuming that UAV passes through the nodes 1, 2, and 3, the survival probabilities of them are 0.9, 0.8, and 0.7, respectively. Then, survival coverage function of the UAV at node 1 is 0.9, node 2 is $0.9 * 0.8 = 0.72$, and node 3 is $0.9 * 0.8 * 0.7 = 0.50$. The survival coverage function of entire path of UAVs for reconnaissance missions is the cumulative sum of the node's survival coverage functions, as shown in equation (8), where $a^v \leq n_i^v \leq b^v$. Formula (9) indicates that the survival coverage function f_2 is optimized to take the maximum value.

In the process of performing actual tasks, in order to avoid excessive load of a single UAV, the load of each UAV is adjusted by introducing the variance of nodes, as shown in formula (10).

3. Solving Algorithm for Model

3.1. Attention Model. The attention model [18] is a framework proposed by Google for dealing with problems related to sequence models and has achieved peer-leading results in neural network translation applications. While most traditional neural network machine translation utilizes RNN or CNN as the basis of encoder-decoder framework [9], the attention model is based only on its transformer structure, which discards the inherent stereotypes. The attention model can work highly in parallel manner, so it significantly improves the training

speed of the model as well improves its translation performance.

The optimal combination of the multi-UAV collaborative reconnaissance model also belongs to one of the sequential models. In this paper, an end-to-end model based on attention is used to train a random strategy $p(\pi | s)$, and the path planning $\pi = (\pi_1, \dots, \pi_{N_T})$, $\pi_t \in T$, is given from the constraints of the current terrain, where s as an instance of the current planning:

$$p_\theta(\pi | s) = \prod_{t=1}^n p_\theta(\pi_t | s, \pi_{1:t-1}). \quad (11)$$

In the end-to-end model, the encoder processes the embeddings' encoding of the input coordinate nodes, and the decoder accepts the input from the encoder. The decoding process is constrained by a mask mechanism to give the rational path planning π .

3.2. Encoder. In the problem of multi-UAV collaborative detection, the relevant information of each node is not sequentially correlated. The encoder of the attention model makes use of the transformer structure [18], due to the fact the transform model encoder will examine the sequence correlation of the data by default. In order to keep away from the interference of the order of the input nodes on the path planning, the position embedding layer is not used in the encoder structure adopted in this paper, as shown in Figure 1.

The coordinates of the nodes are mapped into $d_h = 128$ dimensional embeddings by using a linear mapping layer with parameters W^x and b^x , and aiming at distinguishing the base and the node, different parameters are used:

$$h_i^{(0)} = \begin{cases} W^x(x_i, y_i) + b^x, & i = 0, \\ W_0^x(x_0, y_0) + b_0^x, & i = 1, \dots, n. \end{cases} \quad (12)$$

The embedding of the input nodes is updated through N -layer attention layers, each of which consists of two sublayers: multilead attention (MHA) layer and the nodewise fully connected feedforward (FF) layer [19]. Each sublayer is connected by a skip-connection structure, which solves the prominent problem that the learning efficiency deteriorates as the number of network layers deepens. Then, the batch normalization (BN) [20] is regularized, and the calculation formulas are shown in equations (13) and (14). $h_i^{(l)}$ is defined as the node embedding, $l \in \{1, \dots, N\}$, generated by the first layer, in which BN and MHA with superscript represent that parameters are not shared between layers. The MHA layer uses $M=8$ heads, each of which has a dimension of $d_h/M = 16$. The FF layer uses a hidden layer with a dimension of 512 and uses the ReLU activation function. In course of the experiment, the results show that the effect of using BN regularization is better than the transformer encoder, so this paper uses BN to regularize the network output:

$$\hat{h}_i = \text{BN}^l(h_i^{(l-1)} + \text{MHA}(h_1^{(l-1)}, \dots, h_n^{(l-1)})), \quad (13)$$

$$H_i^{(l)} = \text{BN}^l(\hat{h}_i + \text{FF}^l(\hat{h}_i)). \quad (14)$$

Graph embedding, $\bar{h}^{(N)} = (1/n) \sum_{i=1}^n h_i^{(N)}$, is obtained by calculating the average value of the network output layer embedding, and then, the output layer embedding and graph embedding are passed to the decoder.

3.3. Decoder. The decoder generates path planning π by receiving embedding, graph embedding [21], and context embedding of output layer's node of the encoder. The decoding process is carried out step by step, as shown in Figure 2. At each time point t , the decoder generates context embedding based on the embedding of the input node and π_t at time $t-1$. Define context embedding as follows:

$$h_{(c)}^{(N)} = \begin{cases} [\bar{h}^{(N)}, h_{\pi_{t-1}}^{(N)}, \hat{D}_{i,t}, p_i], & t > 1, \\ [\bar{h}^{(N)}, h_{\pi_0}^{(N)}, \hat{D}_{0,t}, p_0], & t = 1, \end{cases} \quad (15)$$

where $\hat{D} = L - \delta_{\pi_t}$ is the remaining cruising range of UAV at the current moment and δ_{π_t} represents the distance during this mission at π_t time. The symbol $[., ., .]$ depicts a horizontal concatenation operator, which means to join three vectors horizontally into one vector.

It then uses an MHA layer (unlike the MHA of the encoder), where $q_{(c)}$ is calculated by $h_{(c)}^{(N)}$ and k_i and v_i are calculated from $h_i^{(N)}$. For computational efficiency skip connection, batch normalization and feedforward are not used:

$$\begin{aligned} q_{(c)} &= W^Q h_{(c)}^{(N)}, \\ k_i &= W^K h_i^{(N)}, \\ v_i &= W^V h_i^{(N)}, \\ u_{(c)i} &= \begin{cases} \frac{q_{(c)}^T k_i}{\sqrt{d_k}}, & \text{if } j \neq \pi_t \forall t' < t, \\ -\infty, & \text{otherwise.} \end{cases} \end{aligned} \quad (16)$$

The similarity $u_{(c)i}$ can be calculated by the above formula, where $d_k = d_h/M$; at the same time, the nodes that do not meet the constraint conditions are shielded, and the compatibility is set to $-\infty$. To derive the strategy in equation (11), a single attention layer is used at the last layer of the network:

$$u_{(c)j} = \begin{cases} C \cdot \tanh\left(\frac{q_{(c)}^T k_j}{\sqrt{d_k}}\right), & \text{if } j \neq \pi_t \forall t' < t, \\ -\infty, & \text{otherwise.} \end{cases} \quad (17)$$

Finally, the access probability of each node is obtained, in which the visited nodes and the nodes that do not meet the

constraints have been masked when calculating the similarity and probability:

$$p_i = p_\theta(\pi_t = i | s, \pi_{1:t-1}) = \frac{e^{u_{(c)i}}}{\sum_j e^{u_{(c)j}}}. \quad (18)$$

In the process of decoding, the current best-fit node is sampled probabilistically by equation (18).

3.4. Pretraining Model. REINFORCE algorithm adopted in this paper is a strategy gradient algorithm, which means that the strategy is parameterized. The strategy function can be expressed as

$$\pi_\theta = P(a | s, \theta) \approx \pi(a | s). \quad (19)$$

The strategy function determines the probability of taking any action in a given state and a certain parameter setting; thus, it is substantially a probability density function. In order to find the optimal strategy, the gradient formula (20) is obtained by differentiating the objective function, wherein $\nabla_\theta \log \pi_\theta(s, a)$ is the score function and $Q_\pi(s, a)$ is the action state variable value:

$$\nabla_\theta J(\theta) = E_{\pi_\theta}[\nabla_\theta \log \pi_\theta(s, a) Q_\pi(s, a)]. \quad (20)$$

After constructing the model, the loss function equation (21) is defined in order for it to be trained. And, the objective function of reinforcement learning is modified to make $L(\pi)$ achieve the purpose of optimizing planning:

$$\mathcal{L}(\theta | s) = E_{p_\theta(\pi | s)}[(L(\pi) - b(s)) \nabla \log p_\theta(\pi | s)]. \quad (21)$$

In REINFORCE algorithm, the baseline $b(s)$ can reduce variance and improve training speed of the model. In this paper, the sliding average model $M \leftarrow \beta M + (1 - \beta)L(\pi)$ is regarded as the baseline for the REINFORCE algorithm.

As shown in Figure 3, after defining the model and the optimization goal, it is trained by using a large amount of simulation data so that the model can dig out the connection between the data and the optimization path. If the indicators are effective, the model can be extracted, and its parameters can be fixed, then the real data of the terrain can be input to solve the UAV reconnaissance problem in real time. Traditional heuristic algorithms, such as genetic algorithms, ant colony algorithms, and particle swarm algorithms, all have to go through a process of optimization, iteration, and evolution when processing new data [12, 13, 22], which is equivalent to re-executing training on a single set of data with long iteration times. However, terrain data are rapidly changing, and such a process is obviously not suitable for applications requiring high real-time performance. The trained neural network model finds rules and features from the simulated data and directly gives the optimized path from the data. This means that it skips the process of reiteration and directly uses known experience to solve similar problems, and its adaptability is significantly stronger than traditional heuristic models. Meanwhile, the model can also introduce an online training mechanism (as shown by the dashed box) and use the real data of the terrain to train the model, which continuously improve the effect and realize the self-evolution of the model.

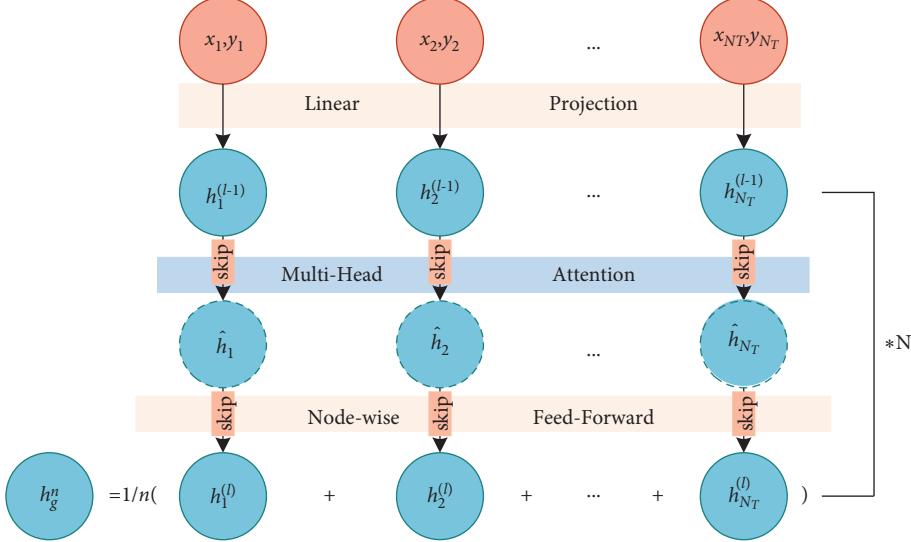


FIGURE 1: Schemes follow the same formatting.

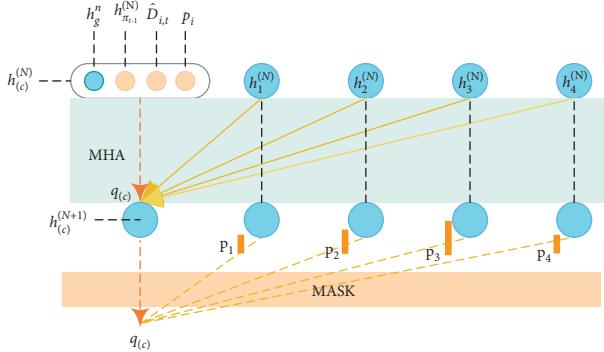


FIGURE 2: Decoder network structure.

4. Simulation Test

M of the moving average model is initialized to $L(\pi)$, and the attenuation rate $\beta = 0.8$. Using Adam optimizer [23], initialize the learning rate $\eta = 10^{-4}$, the number of encoder attention layers $N = 3$, and the number of nodes as 20, and the network parameters obey a uniform distribution at $(-1/\sqrt{d}, 1/\sqrt{d})$. In this paper, five UAVs are simulated to scout 20 nodes in a fixed area, and the survival probability and total range of the UAVs are optimized under the condition that the UAV's endurance range is satisfied. At the same time, the simulation scene meets the following conditions:

- (1) Each UAV reaches only one target
- (2) Each target is covered by a UAV
- (3) During the flight, the UAV should keep a certain safe distance from the obstacles
- (4) During the flight, UAV should keep a certain safe distance from other UAVs.

In the UAV collaborative reinforcement learning algorithm based on the attention model, the node and base

coordinate data of the UAV obey the uniform distribution on $[0, 1]$, and the survival probability of the UAV when detecting the target i is randomly generated from $P_i \in [0.7, 0.9]$. The objective function coefficients in the loss function are $[0.3, -0.5, 0.2]$. In Table 1, the nodes, base coordinates, and survival data of the detection area are listed.

The research studies are implemented based on the deep learning framework PyTorch. The configuration and environment of the computer in the experiment are as follows. The CPU is AMD ryzen5 1600x and the GPU is GTX1080, with 16 GB of memory. After 100 rounds of network training, the test data sets are used to compare the model proposed in this paper with other traditional ways, and the results are shown in Table 2. Among them, AM is the model of this article, CVPR is the traditional model, and TSC-GA is a genetic algorithm based on the two-side successive revision method proposed in [3], and the time-consuming data of TSC-GA comes from the original text. The attention prefix in the table indicates that attention neural network is used to build the model, and the performance of different node numbers based on CPU and GPU are compared, respectively. The AM model used in this paper was tested 10 times with 1000 sets of data, and the traditional CVPR100 model was tested 100 times with 100 sets of test data, which all are a total 10000 sets. The adaptive large neighborhood search algorithm proposed in [2] for the SDVRP problem with variable demand points has an optimized iteration time magnitude of around 500 s, which is not listed in Table 1.

It can be seen from Table 2 that whether the algorithm based on the attention model is run by CPU or accelerated by CUDA, the time of a single group of simulated reconnaissance operations is much shorter than that of the traditional model based on evolutionary algorithm (TSC-GA). Various situations can be handled fast and real time in the actual terrain, which is much better than the adaptive large neighboring search algorithm also. The model based on traditional evolutionary algorithm needs to be iteratively optimized [13], when encountering different tasks with the

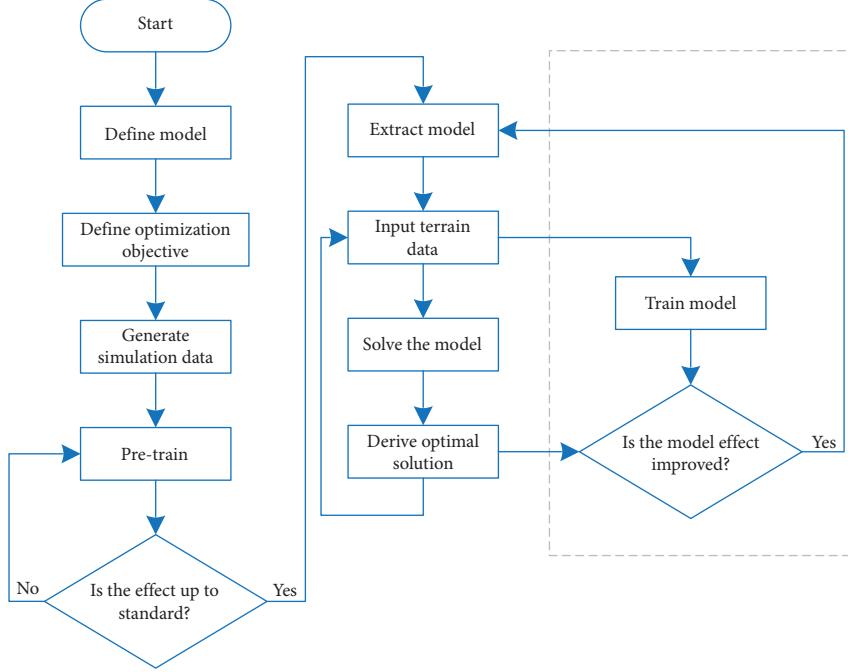


FIGURE 3: Program flowchart.

TABLE 1: Training data schematic.

Serial no.	0	1	2	3	4
x_i	0.128	0.9498	0.1891	0.2267	0.2715
y_i	0.4532	0.4065	0.5248	0.0061	0.7013
P_i	1	0.8	0.9	0.9	0.9
Serial no.	5	6	7	8	9
x_i	0.2527	0.6383	0.8405	0.5754	0.0855
y_i	0.5374	0.3824	0.9356	0.3803	0.3179
P_i	0.8	0.7	0.8	0.7	0.7
Serial no.	10	11	12	13	14
x_i	0.6551	0.8748	0.9738	0.5008	0.9691
y_i	0.9123	0.8385	0.7627	0.839	0.592
P_i	0.9	0.8	0.7	0.7	0.7
Serial no.	15	16	17	18	19
x_i	0.5378	0.2525	0.6076	0.1459	0.5593
y_i	0.5028	0.8478	0.848	0.1864	0.0544
P_i	0.7	0.8	0.7	0.7	0.7

same number of nodes, while the trained collaborative model based on attention mechanism can quickly deal with similar problems without retraining. Moreover, the model based on the attention mechanism can simulate in parallel in batches and can simultaneously process 1000 (100) sets of data at a single time. The throughput performance of problem processing is far superior to the traditional heuristic algorithm.

In terms of optimal path and survival probability on 1000 sets of test data, as can be seen in Table 3, the AM model UAV ensures that the path distance is within a reasonable acceptable range, and the survival probability has a great improvement compared to the traditional model.

The path planning of the AM model is shown in Figure 4, in which the legend lists the survival probability and path

length of each UAV path planning, respectively. By comparing the objective function equation (2), it can be found that the trained model has well grasped the characteristics among the data and optimized the cruising range, load, and survival probability of UAV. The diagram shows that the model plans the path for the UAV in a balanced way and maintains a high survival probability, and none of the UAV flight paths exceed the range.

The proposed path planning based on the attention-UAV model could be used in different cases, mainly in military and civil. In military aspect, UAV could be used as reconnaissance aircraft and target aircraft. In the civil aspect, UAV technology could be widely used in aviation shooting, plant protection, news reporting, microselfie, express transportation, agriculture, wildlife observation, disaster

TABLE 2: Solving time.

Name	Data batch * times	Number of nodes	Time consuming (s)	Single set of data time consuming
Attention-AM (cpu)	1000 * 10	20	34.45	3.45 ms
Attention-AM (cuda)	1000 * 10	20	5.45	0.55 ms
Attention-CVRP100 (cpu)	100 * 100	100	266	26.6 ms
Attention-CVRP100 (cuda)	100 * 100	100	66.18	6.62 ms
TSC-GA	Eil101 * 1	101	2	2 s

TABLE 3: Model performance.

Name	Average distance	Average survival coverage
Attention-AM	7.766	12.916
Attention-CVRP20	6.575	10.219

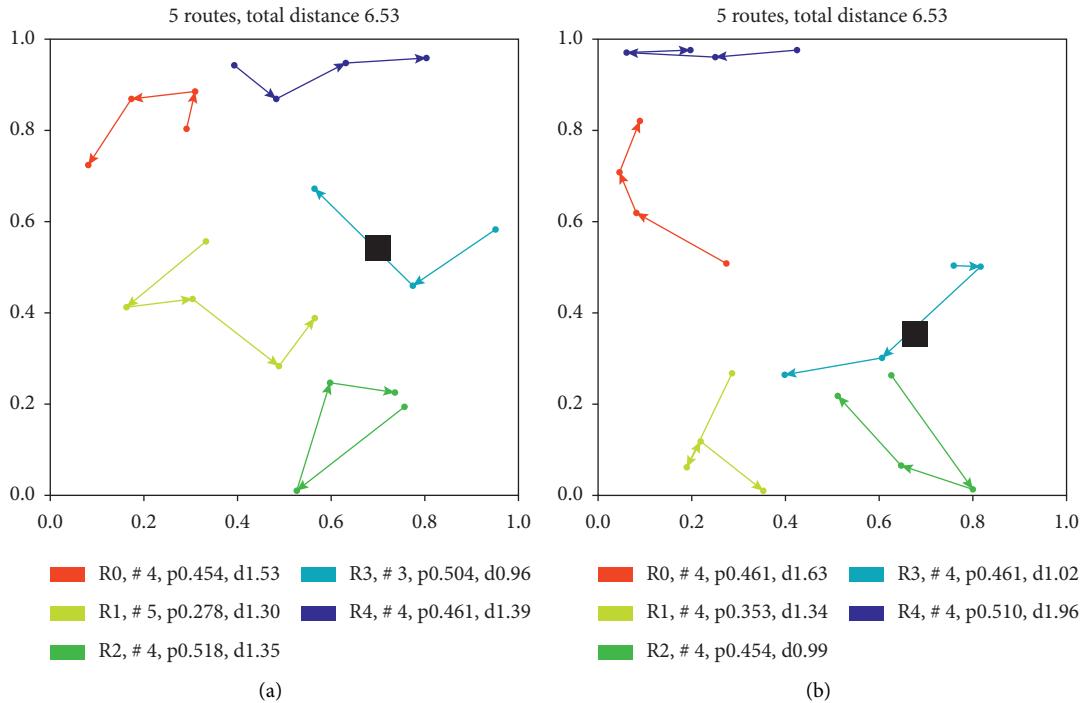


FIGURE 4: Path planning based on the attention-UAV model.

rescue, film and television shooting, infectious disease monitoring, disaster relief, power inspection, romantic manufacturing, and other fields.

5. Conclusions

Through data analysis and experimental results, it can be seen that the path planning algorithm for UAVs' cooperative reconnaissance based on the attention model has a great advantage over the traditional heuristic algorithm in terms of real time, which can cope with complex terrain changes and can be applied to earthquake relief. In addition, the algorithm based on the attention model can well excavate the internal relationship between terrain data and optimized path and make reasonable path planning for multi-UAV collaborative reconnaissance mission.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

T.W. developed the methodology; T.W. and S.Z. helped with software; T.W., B.Z., and M.Z. validated the study; B.Z. carried out formal analysis; T.W. investigated the study; T.W. helped with the resources; B.Z. curated the data; T.W. wrote and prepared the original draft; T.W. and B.Z. reviewed and edited the study; M.Z. visualized the study;

T.W. supervised the study; T.W. administrated the project; T.W. helped with funding acquisition. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This work was supported in part by the Science and Technology Key Project of Henan Province, under Grant no. 212102210520, and the Science and Technology Key Project in the High-Tech Fields of Henan Province, under Grant no. 152102210123.

References

- [1] S. M. Adams and C. J. Friedland, "A survey of unmanned aerial vehicle (UAV) usage for imagery collection in disaster research and management," in *Proceedings of the 9th International Workshop on Remote Sensing for Disaster Response*, vol. 8, Stanford, CA USA, September 2011.
- [2] M. Erdelj and E. Natalizio, "UAV-assisted disaster management: applications and open issues," in *Proceedings of the 2016 International Conference on Computing, Networking and Communications (ICNC)*, pp. 1–5, IEEE, Kauai, HI, USA, February 2016.
- [3] B. Fan, Y. Li, R. Zhang, and Q. Fu, "Review on the technological development and application of UAV systems," *Chinese Journal of Electronics*, vol. 29, no. 2, pp. 199–207, 2020.
- [4] L. Gupta, R. Jain, and G. Vaszkun, "Survey of important issues in UAV communication networks," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1123–1152, 2015.
- [5] M. A. Ma'Sum, M. K. Arrofi, G. Jati et al., "Simulation of intelligent unmanned aerial vehicle (uav) for military surveillance," in *Proceedings of the 2013 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pp. 161–166, IEEE, Sanur Bali, Indonesia, September 2013.
- [6] U. R. Mogili and B. B. V. L. Deepak, "Review on application of drone systems in precision agriculture," *Procedia Computer Science*, vol. 133, pp. 502–509, 2018.
- [7] B. H. Y. Alsalam, K. Morton, D. Campbell, and F. Gonzalez, "Autonomous UAV with vision based on-board decision making for remote sensing and precision agriculture," in *Proceedings of the 2017 IEEE Aerospace Conference*, pp. 1–12, IEEE, Big Sky, MT, USA, March 2017.
- [8] Y. Wan, X. Hu, Y. Zhong, A. Ma, L. Wei, and L. Zhang, "Tailings reservoir disaster and environmental monitoring using the UAV-ground hyperspectral joint observation and processing: a case of study in Xinjiang, the belt and road," in *Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pp. 9713–9716, IEEE, Yokohama, Japan, July 2019.
- [9] J. Tian, *Research on Modeling and Optimization Technology of Multi-UAV Collaborative Detection Mission Planning*, National University of Defense Technology, Changsha, China, 2007.
- [10] J. Tian, L. Shen, and Y. Zheng, "Genetic algorithm based approach for multi-UAV cooperative reconnaissance mission planning problem," *Lecture Notes in Computer Science*, pp. 101–110, Springer, Berlin, Heidelberg, 2006.
- [11] K. Shi, X. Zhang, and S. Xia, "Multiple swarm fruit fly optimization algorithm based path planning method for multi-UAVs," *Applied Sciences*, vol. 10, no. 8, p. 2822, 2020.
- [12] L. Lin, *Research on Multi-UAV Mission Planning Based on Collaborative Mechanism*, Beijing University of Posts and Telecommunications, Beijing, China, 2013.
- [13] H. Yuan, *Research and Application of Traveling Salesman Problem*, Nanjing University of Posts and Telecommunications, Nanjing, China, 2017.
- [14] J. Shi, "Research on VRP model and algorithm of batch distribution with random demand point," *Decision and Control*, vol. 32, no. 2, pp. 213–222, 2017.
- [15] O. Vinyals, M. Fortunato, and N. Jaitly, "Pointer networks," 2015, <https://arxiv.org/abs/1506.03134>.
- [16] M. Nazari, A. Oroojlooy, L. V. Snyder, and M. Takáć, "Deep reinforcement learning for solving the vehicle routing problem," 2018, <https://arxiv.org/abs/1802.04240>.
- [17] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Machine Learning*, vol. 8, no. 3-4, pp. 229–256, 1992.
- [18] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," 2017, <https://arxiv.org/abs/1706.03762>.
- [19] M. T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," 2015, <https://arxiv.org/abs/1508.04025>.
- [20] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," 2015, <https://arxiv.org/abs/1502.03167>.
- [21] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," 2017, <https://arxiv.org/abs/1710.10903>.
- [22] J. L. Ryan, T. G. Bailey, J. T. Moore, and W. B. Carlton, "Reactive tabu search in unmanned aerial reconnaissance simulations," in *Proceedings of the 1998 Winter Simulation Conference*, vol. 1, pp. 873–879, IEEE, Washington, DC, USA, December 1998.
- [23] D. P. Kingman and J. Ba, "Adam: a method for stochastic optimization," in *Proceedings of the 3rd International Conference for Learning Representations*, San Diego, CA, USA, May 2015.