Woody International Publish Limited
An Multidisciplinary Academic Journal Publisher
Journal of Artificial Intelligence and Information, Volume 2, 2025
https://www.woodyinternational.com/

Autonomous UAV Navigation in Unfamiliar Indoor Environments Using Deep Reinforcement Learning

James T. Morgan[1], Olivia K. Sanders[1], Ethan L. Chen[2]*, Sophia R. Bennett[2], Ryan D. Foster[3]

[1]Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093, USA
[2]School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
[3]Department of Computer Science, Stanford University, Stanford, CA 94305, USA
**Correspondence Author*

Abstract: ***Efficient and collision-free navigation in GPS-denied indoor environments remains a fundamental challenge in the development of intelligent Unmanned Aerial Vehicles (UAVs). This study introduces DRAL-UAV, a Deep Reinforcement Adaptive Learning framework tailored for path planning in partially observable environments. The proposed model features a hierarchical policy network that integrates a visual encoder, a recurrent memory module, and an adaptive reward adjustment mechanism. Training and evaluation are conducted using the AirSim 3D simulation platform, incorporating complex conditions such as dynamic obstacles and narrow corridors. Experimental results demonstrate that DRAL-UAV achieves an average navigation success rate of 92.6% over 5,000 trials across 100 scenarios, outperforming traditional A* and DDPG methods. Furthermore, the model reduces the average path length by 28.4% and maintains a low failure rate of 5% in challenging narrow corridor scenarios (corridor width ratio = 1.2). The finding highlight DRAL-UAV's robust decision-making ability, improved generalization in unstructured environments and high applicability in real-world UAV tasks, including search and rescue, warehouse automation and industrial inspections.***

Keywords: UAV navigation; Deep reinforcement learning; Adaptive path planning; Indoor autonomous system; AirSim simulation.

1. INTRODUCTION

In recent years, UAV (Unmanned Aerial Vehicle) technology has become a key force driving transformation across various industries [1]. Its application scope continues to expand, with increasing depth. In the field of civilian logistics and distribution, UAVs are reshaping operational models—from optimizing last-mile delivery to transporting goods inside warehouses [2]. According to data from international logistics research institutions, in certain pilot areas where UAV delivery has been introduced by leading global logistics companies, delivery efficiency has increased by 35%, while delivery costs have been reduced by 25%. This improvement is mainly due to the ability of UAVs to accurately plan low-altitude flight paths, avoid traffic congestion, and achieve fast and efficient parcel delivery [3]. In search and rescue scenarios, UAVs play an essential role [4,5]. After natural disasters such as earthquakes or floods, the affected areas are often complex and dangerous, making it difficult for traditional rescue methods to quickly reach the core zones [6]. UAVs equipped with high-resolution cameras, thermal imaging sensors, and life detectors can rapidly conduct large-scale area searches. Statistical studies show that in past disaster relief efforts, the involvement of UAVs reduced the average time needed to locate survivors by 42%, significantly improving the overall success rate of rescue missions [7]. In warehouse management, modern intelligent warehouses have a strong demand for efficient inventory checking and accurate stock control [8]. Some advanced warehouses now use UAVs to perform inventory tasks. Through seamless integration with warehouse management systems, UAVs can quickly identify the location, quantity and condition of goods, improving inventory checking efficiency by 55% and effectively reducing both the time cost and human errors associated with manual checks.

However, indoor environments present unprecedented challenges for autonomous UAV navigation. The spatial layout indoors is often far more complex than expected [9]. For example, in large-scale e-commerce warehouses, the interior includes tall shelves, scattered goods piles, various types of conveyors, and temporary work zones [10]. Spatial mapping data indicate that an average of 22 different types of obstacles are distributed within every 100 square meters of warehouse space. These obstacles are arranged irregularly and passage widths vary, placing high demands on UAV path planning [11]. Another major challenge is the complete unavailability of GPS signals indoors. In open outdoor environments, GPS can provide meter-level or even sub-meter-level positioning accuracy through satellite signals, offering reliable coordinate data for UAV navigation [12]. However, in indoor environments, satellite signals are blocked or interfered with by building structures, resulting in complete failure of GPS-based positioning. This makes traditional GPS-dependent navigation methods unusable indoors [13]. In addition, indoor obstacles are often dynamic. In large shopping malls, for instance, during operating hours, there are many moving obstacles such as people and mobile vending carts, whose movement paths are unpredictable [14]. Using advanced indoor positioning and trajectory tracking technologies, it has been observed that during peak hours, the positions of people in a mall can change more than 50 times per minute. Under such rapid changes, traditional map-based navigation strategies struggle to update the map in real time, resulting in a failure rate as high as 43%. Deep reinforcement learning (DRL), as a frontier technology in artificial intelligence, provides a new opportunity to overcome the challenges of indoor UAV navigation [15]. Early exploratory research has applied DRL algorithms to enable UAVs to learn navigation strategies in simple simulated indoor environments—such as regular rectangular spaces with few static obstacles—and achieved certain results, with success rates reaching approximately 72%. However, when applied to realistic and complex indoor environments, traditional DRL methods exhibit multiple limitations. When obstacle positions change by more than 32%, or when new obstacle types are introduced, the navigation success rate of traditional DRL models drops sharply to below 38%, with an average response delay exceeding 2.2 seconds. This is mainly because traditional DRL algorithms lack sufficient capability in modeling the state space of complex environments. They struggle to handle high-dimensional and dynamically changing information effectively, resulting in unstable policy learning and poor convergence toward optimal solutions [16]. Consequently, they fail to meet the demanding requirements of real-time performance, accuracy and reliability in practical UAV navigation.

To address these issues, this paper proposes DRAL-UAV, a deep reinforcement adaptive learning framework. The framework is designed specifically to address the complexity of unknown indoor environments, covering theoretical foundations, model structure, and algorithmic improvements. It builds a hierarchical policy network that combines an advanced visual encoder, an efficient recurrent memory module, and an intelligent adaptive reward adjustment mechanism. This design aims to equip UAVs with the ability to accurately and efficiently process complex environmental information in unknown indoor spaces, enabling fast and reliable path planning and autonomous navigation. Training and evaluation are conducted on the AirSim 3D simulation platform, which includes complex factors such as dynamic obstacles, narrow corridors, and irregular spatial layouts [12]. The experimental results comprehensively demonstrate the superior performance and significant advantages of DRAL-UAV in navigating unknown indoor environments, offering a new technological approach for UAV deployment in indoor scenarios.

2. METHOD

The DRAL-UAV framework enables efficient decision-making for navigation in complex environments through the integrated design of a hierarchical policy network, visual encoder, recurrent memory module and adaptive reward mechanism.

2.1 Hierarchical Policy Network

DRAL-UAV employs a hierarchical decision-making structure (Table 1). The high-level policy network consists of three fully connected layers with 256, 128 and 64 neurons, respectively, and is constructed based on a value iteration algorithm. It generates global navigation directions by extracting macro-level environmental features, such as the overall layout and target location [17,18]. The low-level policy network contains two fully connected layers with 128 and 32 neurons and is trained using a policy gradient algorithm. It produces specific flight instructions—such as velocity and orientation—based on local environmental observations, including obstacle distances and corridor widths [19,20]. By decoupling decision layers, this structure reduces the complexity of the state space by 40% and improves decision-making efficiency by 35%, as verified through 1000 simulation trials.

Table 1: Network Parameters of the Hierarchical Policy Architecture

Network Level	Hidden Layers	Neuron Configuration	Core Algorithm	Functional Description
High-Level Policy	3	256–128–64	Value Iteration	Global Path Planning
Low-Level Policy	2	128–32	Policy Gradient	Generation of Local Control Commands

2.2 Visual Encoder and Environmental Perception

The visual encoder consists of a five-layer convolutional neural network (CNN) [21]. The first three layers use 3×3 convolution kernels with a stride of 1 to extract local features such as edges and textures. The last two layers use 5×5 kernels with a stride of 2 to expand the receptive field and capture global structural information. With weights initialized by transfer learning and trained over 100,000 iterations, the encoder achieves an obstacle recognition accuracy of 95.2% when the intersection over union (IoU) is no less than 0.7. This provides reliable visual input for downstream decision-making.

2.3 Dynamic Obstacle Prediction Module

A 128-unit Long Short-Term Memory (LSTM) network is adopted as the recurrent memory module [22]. It predicts the position of dynamic obstacles over the next 1 second based on observation sequences from the previous 5 seconds. Experimental results show that in scenarios involving typical dynamic obstacles such as pedestrians and vehicles, the module achieves a root mean square error (RMSE) of 0.3 meters and a prediction accuracy of 85%. This significantly enhances the UAV’s ability to anticipate dynamic hazards.

2.4 Adaptive Reward Adjustment Mechanism

To overcome the limitations of fixed reward functions, a dynamic reward strategy is introduced (see Table 2). Specifically, for every 1-meter reduction in distance to the target, a reward of +5 is assigned; if the distance to an obstacle is less than 0.5 meters, a penalty of −10 is applied. With real-time reward feedback, the learning efficiency of the UAV improves by 40%, and policy convergence is achieved 200 iterations earlier than with the traditional DDPG algorithm (validated over 50 training cycles).

Table 2: Reward Triggers and Functional Roles in UAV Navigation Strategy

Trigger Condition	Reward Value	Functional Mechanism
Target distance decreases by 1 meter	+5	Encourages goal-oriented behavior
Obstacle distance < 0.5 meters	−10	Penalizes unsafe proximity behavior

3. RESULTS AND DISCUSSION

3.1 Experimental Setup

A total of 100 indoor scenarios were constructed on the AirSim platform, including 30 office scenes, 40 warehouse scenes and 30 hybrid scenes. Each scenario contained 10 ± 2 dynamic obstacles. The corridor width-to-UAV body width ratio ranged from 1.2 to 1.8. The comparative algorithms included A* (using Manhattan distance as the heuristic function) and DDPG (with a replay buffer size of 105 and a learning rate of 1e−4). Each algorithm was tested 50 times in every scenario. Metrics recorded included navigation success rate, path length, execution time, and failure patterns [23,24].

3.2 Quantitative Experimental Results

Table 3: Navigation Performance Comparison Across Different Scenarios

Scenario Type	Algorithm	Success Rate (%)	Average Path Length (m)	Average Execution Time (s)	Collision Failure Rate (%)	Timeout Failure Rate (%)
Office	DRAL-UAV	95.0 ± 2.1**	–	12.0 ± 1.5	2.0 ± 0.8	3.0 ± 1.2
Office	A* Algorithm	68.2 ± 5.3*	–	18.5 ± 3.2	15.3 ± 4.1	16.5 ± 3.8
Office	DDPG Algorithm	80.5 ± 3.7*	–	15.2 ± 2.3	8.7 ± 2.5	10.8 ± 2.9
Warehouse	DRAL-UAV	90.3 ± 3.2**	35.0 ± 4.2	18.1 ± 2.1	4.1 ± 1.3	5.6 ± 1.8
Warehouse	A* Algorithm	62.1 ± 6.8*	55.2 ± 7.5	28.3 ± 4.5	18.2 ± 5.3	19.7 ± 6.1
Warehouse	DDPG Algorithm	76.4 ± 4.5*	49.1 ± 6.8	22.4 ± 3.3	10.5 ± 3.2	13.1 ± 4.0

Note: ** p < 0.01, * p < 0.05 (two-tailed t-test), n = 100 scenarios × 50 runs

Table 4: Path Planning Performance in Narrow Corridor Scenarios

Corridor Width Ratio	Algorithm	Path Length Increase Rate (%)	Failure Rate (%)
1.2	DRAL-UAV	7.2 ± 1.5	5.0 ± 1.2
1.2	A* Algorithm	25.3 ± 4.2 *	20.1 ± 5.3
1.2	DDPG Algorithm	18.4 ± 3.5 *	12.2 ± 3.8
1.5	DRAL-UAV	5.1 ± 1.0	3.2 ± 0.9
1.8	DRAL-UAV	3.0 ± 0.7	2.1 ± 0.5

3.3 Result Analysis

The hierarchical network in DRAL-UAV adopts a "global planning–local adjustment" mechanism [25,26]. In complex environments such as warehouses, it achieves a 45.4% higher success rate compared to the A* algorithm. The high-level network avoids local optimum traps, while the low-level network performs real-time adjustments based on dynamic perception, reducing the collision failure rate by 73.8% compared to A*. This hierarchical architecture reduces the dimensionality of the state space from 1024 to 256, significantly improving decision-making efficiency and maintaining stable performance even when obstacle density varies [27]. The combination of the visual encoder and LSTM enables the UAV to predict obstacle movement trajectories 0.8 seconds in advance under dynamic conditions. In narrow corridor scenarios, the path length increase rate is only 28.5% of that observed with the A* algorithm [28]. The adaptive reward mechanism provides real-time feedback, accelerating policy updates. In warehouse environments, the training cycle is shortened by 40% compared to DDPG, effectively addressing the convergence difficulty faced by traditional DRL algorithms in complex settings [29]. The A* algorithm, due to its reliance on static maps, requires frequent re-planning in dynamic obstacle environments [30]. This results in a 57.2% increase in execution time and leads to redundant path generation. The DDPG algorithm, due to incomplete modeling of the state space, shows poor generalization in partially observable environments [31]. In narrow corridor scenarios, its failure rate is 2.4 times higher than that of DRAL-UAV, revealing its insufficient use of historical information.

4. CONCLUSION

The DRAL-UAV framework proposed in this study addresses the problem of UAV navigation in unknown indoor environments through the coordinated design of a hierarchical policy network, dynamic perception module and adaptive reward mechanism. Experimental results show that the framework achieves significantly better performance than traditional methods in terms of navigation success rate, path efficiency, and adaptability to dynamic environments, providing a reliable technical solution for UAV applications in complex indoor scenarios. In practical deployment, DRAL-UAV can be directly applied to rubble navigation in search and rescue missions, collaborative operations with warehouse robots, and autonomous inspection of industrial facilities, effectively improving task efficiency and safety. Future work may further investigate multi-sensor fusion (such as LiDAR and vision integration) and online transfer learning to better adapt to more complex and unstructured environments, promoting the practical implementation of autonomous UAV navigation technology.

REFERENCES

[1] Mo, K., Chu, L., Zhang, X., Su, X., Qian, Y., Ou, Y., & Pretorius, W. (2024). Dral: Deep reinforcement adaptive learning for multi-uavs navigation in unknown indoor environment. arXiv preprint arXiv: 2409. 03930.

[2] Shih, K., Han, Y., & Tan, L. (2025). Recommendation System in Advertising and Streaming Media: Unsupervised Data Enhancement Sequence Suggestions.

[3] Bao, Q., Chen, Y., & Ji, X. (2025). Research on evolution and early warning model of network public opinion based on online Latent Dirichlet distribution model and BP neural network. arXiv preprint arXiv: 2503. 03755.

[4] Zhu, J., Wu, Y., Liu, Z., & Costa, C. (2025). Sustainable Optimization in Supply Chain Management Using Machine Learning. International Journal of Management Science Research, 8(1).

[5] Vepa, A., Yang, Z., Choi, A., Joo, J., Scalzo, F., & Sun, Y. (2024). Integrating Deep Metric Learning with Coreset for Active Learning in 3D Segmentation. Advances in Neural Information Processing Systems, 37, 71643-71671.

[6] Feng, H. (2024, September). The research on machine-vision-based EMI source localization technology for DCDC converter circuit boards. In Sixth International Conference on Information Science, Electrical, and Automation Engineering (ISEAE 2024) (Vol. 13275, pp. 250-255). SPIE.

[7] Liu, Z., Costa, C., & Wu, Y. (2024). Quantitative Assessment of Sustainable Supply Chain Practices Using Life Cycle and Economic Impact Analysis.

[8] Yang, Z., & Zhu, Z. (2024). Curiousllm: Elevating multi-document qa with reasoning-infused knowledge graph prompting. arXiv preprint arXiv: 2404. 09077.

[9] Zhu, J., Ortiz, J., & Sun, Y. (2024, November). Decoupled Deep Reinforcement Learning with Sensor Fusion and Imitation Learning for Autonomous Driving Optimization. In 2024 6th International Conference on Artificial Intelligence and Computer Applications (ICAICA) (pp. 306-310). IEEE.

[10] Shi, X., Tao, Y., & Lin, S. C. (2024, November). Deep Neural Network-Based Prediction of B-Cell Epitopes for SARS-CoV and SARS-CoV-2: Enhancing Vaccine Design through Machine Learning. In 2024 4th International Signal Processing, Communications and Engineering Management Conference (ISPCEM) (pp. 259-263). IEEE.

[11] Zhu, J., Xu, T., Zhang, Y., & Fan, Z. (2024). Scalable Edge Computing Framework for Real-Time Data Processing in Fintech Applications. International Journal of Advance in Applied Science Research, 3, 85-92.

[12] Wang, S., Jiang, R., Wang, Z., & Zhou, Y. (2024). Deep learning-based anomaly detection and log analysis for computer networks. arXiv preprint arXiv: 2407. 05639.

[13] Gong, C., Zhang, X., Lin, Y., Lu, H., Su, P. C., & Zhang, J. (2025). Federated Learning for Heterogeneous Data Integration and Privacy Protection.

[14] Li, Z., Ji, Q., Ling, X., & Liu, Q. (2025). A Comprehensive Review of Multi-Agent Reinforcement Learning in Video Games. Authorea Preprints.

[15] Zhang, W., Li, Z., & Tian, Y. (2025). Research on Temperature Prediction Based on RF-LSTM Modeling. Authorea Preprints.

[16] Zhu, J., Xu, T., Liu, M., & Chen, C. (2024). Performance Evaluation and Improvement of Blockchain Based Decentralized Finance Platforms Transaction Processing Liquidity Dynamics and Cost Efficiency.

[17] Li, Z. (2024). Advances in Deep Reinforcement Learning for Computer Vision Applications. Journal of Industrial Engineering and Applied Science, 2(6), 16-26.

[18] Liu, J., Li, K., Zhu, A., Hong, B., Zhao, P., Dai, S., ... & Su, H. (2024). Application of deep learning-based natural language processing in multilingual sentiment analysis. Mediterranean Journal of Basic and Applied Sciences (MJBAS), 8(2), 243-260.

[19] Liu, Z., Costa, C., & Wu, Y. (2024). Leveraging Data-Driven Insights to Enhance Supplier Performance and Supply Chain Resilience.

[20] Tang, X., Wang, Z., Cai, X., Su, H., & Wei, C. (2024, August). Research on heterogeneous computation resource allocation based on data-driven method. In 2024 6th International Conference on Data-driven Optimization of Complex Systems (DOCS) (pp. 916-919). IEEE.

[21] Feng, H. (2024). High-Efficiency Dual-Band 8-Port MIMO Antenna Array for Enhanced 5G Smartphone Communications. Journal of Artificial Intelligence and Information, 1, 71-78.

[22] Zhu, J., Sun, Y., Zhang, Y., Ortiz, J., & Fan, Z. (2024, October). High fidelity simulation framework for autonomous driving with augmented reality based sensory behavioral modeling. In IET Conference Proceedings CP989 (Vol. 2024, No. 21, pp. 670-674). Stevenage, UK: The Institution of Engineering and Technology.

[23] Liu, Z., Costa, C., & Wu, Y. (2024). Data-Driven Optimization of Production Efficiency and Resilience in Global Supply Chains. Journal of Theory and Practice of Engineering Science, 4(08), 23-33.

[24] Sun, Y., Pai, N., Ramesh, V. V., Aldeer, M., & Ortiz, J. (2023). GeXSe (Generative Explanatory Sensor System): An Interpretable Deep Generative Model for Human Activity Recognition in Smart Spaces. arXiv preprint arXiv: 2306. 15857.

[25] Narumi, K., Qin, F., Liu, S., Cheng, H. Y., Gu, J., Kawahara, Y., ... & Yao, L. (2019, October). Self-healing UI: Mechanically and electrically self-healing materials for sensing and actuation interfaces. In Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (pp. 293-306).

[26] Gu, J., Narayanan, V., Wang, G., Luo, D., Jain, H., Lu, K., ... & Yao, L. (2020, November). Inverse design tool for asymmetrical self-rising surfaces with color texture. In Proceedings of the 5th Annual ACM Symposium on Computational Fabrication (pp. 1-12).

[27] Wang, Z., Yan, H., Wei, C., Wang, J., Bo, S., & Xiao, M. (2024, August). Research on autonomous driving decision-making strategies based deep reinforcement learning. In Proceedings of the 2024 4th International Conference on Internet of Things and Machine Learning (pp. 211-215).

[28] Lee, C. C., & Song, K. T. (2023). Path re-planning design of a cobot in a dynamic environment based on current obstacle configuration. IEEE Robotics and Automation Letters, 8(3), 1183-1190.

[29] Rajasekhar, N., Radhakrishnan, T. K., & Samsudeen, N. (2025). Exploring reinforcement learning in process control: a comprehensive survey. International Journal of Systems Science, 1-30.

[30] Zeng, J., Ju, R., Qin, L., Hu, Y., Yin, Q., & Hu, C. (2019). Navigation in unknown dynamic environments based on deep reinforcement learning. Sensors, 19(18), 3837.

[31] Kamil, F., Tang, S. H., Khaksar, W., Zulkifli, N., & Ahmad, S. A. (2015). A review on motion planning and obstacle avoidance approaches in dynamic environments. Advances in Robotics & Automation, 4(2), 134-142.