Dean&Francis
ISSN 2959-6157

A Decision-Making Framework for Task Allocation in Single-Pilot Operations: Synergistic Empowerment Mechanisms of Reinforcement Learning and Deep Learning

Qingyun Hou[1] **and Yanjun Zhou**[2,*]

[1]College of China-New Zealand Collaboration, Dalian Ocean University, Dalian, 116000, China
[2]Department of Civil Engineering, Beijing Forestry University, Beijing, 100080, China
*Corresponding author: Zhouyanan.322@outlook.com

Abstract:

The task allocation decision-making architecture for Single Pilot Operation (SPO) represents a critical development direction in aviation to address soaring operational costs and the global pilot shortage. This paper systematically reviews the synergistic enabling mechanisms of Reinforcement Learning (RL) and Deep Learning (DL) within this architecture. DL serves as the perceptual foundation, processing visual information via CNNs and optimizing human-machine interaction through Transformers to achieve efficient multimodal data comprehension and situational awareness. RL functions as the decision core, leveraging methods such as multi-agent Proximal Policy Optimization (PPO) and Deep Q-Network (DQN) to model complex task allocation problems as Markov Decision Processes, enabling dynamic resource scheduling and multi-constraint optimization. Through deep integration in a "perception-decision-optimization" closed-loop, this dual approach significantly enhances the SPO system's responsiveness, robustness, and safety in high-real-time, high-uncertainty environments. This collaborative mechanism provides critical theoretical foundations and technical pathways for developing next-generation intelligent aviation systems compliant with airworthiness standards and enabling efficient human-machine collaboration.

Keywords: Single-pilot operation; reinforcement learning; deep learning; collaborative empowerment; task allocation decision-making

ISSN 2959-6157

1. Introduction

The civil aviation transport sector currently faces soaring operational costs, compounded by the International Air Transport Association's projection of a global pilot shortage reaching 340,000 by 2035. This dual pressure makes the Single Pilot Operation (SPO) model an inevitable choice for next-generation passenger aircraft[1]. SPO achieves this by reducing the traditional two-person cockpit to a single pilot, supplemented by intelligent systems and remote ground support. This reconfigures the three-tier collaborative architecture of 'pilot-airborne AI agent-ground station', making it a strategic choice for reducing labor costs and enhancing decision-making efficiency. However, existing SPO systems suffer from three major shortcomings: insufficient safety redundancy, uncontrolled cognitive load, and delayed dynamic decision-making [2-5]. The core contradiction lies in the difficulty of traditional rule engines to handle the multi-constraint optimization problem of flight task allocation. Conversely, RL and DL, through data-driven decision-making, offer new pathways for real-time responses in dynamic environments [6].

Current research in aviation intelligent decision-making exhibits a trend towards technological convergence, with RL serving as the decision core and DL forming the perceptual foundation. Early RL algorithm studies employed traditional Q-learning to address discrete action space problems, yet struggled with high-dimensional state spaces. Subsequent developments progressed to multi-agent near-term policy optimization and deep Q-network variants, which significantly enhanced dynamic obstacle avoidance decision efficiency through value function approximation, thereby resolving multi-aircraft coordinated collision avoidance challenges [7]. Deep learning technologies enhance state perception through feature extraction, employing convolutional neural networks to capture pilot eye-tracking and gesture patterns, using temporal models to predict cognitive load trends, and leveraging graph neural networks to optimize air-ground coordination topologies [8]. However, most RL models exhibit poor adaptability to dynamic environments, suffer from strategy mismatches across flight phases, and demonstrate insufficient generalization capabilities [9].

This paper aims to systematically review the 'perception-decision-verification' technology chain within the SPO mission decision-making domain. It summarizes the synergistic enabling mechanisms of reinforcement learning and deep learning identified in the literature, while employing content analysis to deconstruct technical adaptation bottlenecks. Future research should establish an airworthiness certification-oriented intelligent decision-making paradigm, propelling human-machine collaboration beyond technical feasibility towards operational reliability.

2. Multidimensional Constraints in SPO Task Allocation

2.1 Human Constraints: The Dynamic Equilibrium of Cognitive Load and Trust Mechanisms

In SPO mode, pilots must simultaneously undertake operational, monitoring, and decision-making functions, with their cognitive load (CL) becoming the primary constraint. Quantifying CL requires integrating physiological indicators such as eye movement frequency with behavioural data including operational latency and error rates. Its threshold is calibrated by the NASA-TLX scale: peak load must not exceed 70% [9]. For instance, during emergency obstacle avoidance tasks, the pilot's visual attention allocation must satisfy the following requirements [10]:

$$\sum_{i=1}^{N} a_i \cdot T_i \leqslant T_{max} (a_i \in [0.2, 0.8]) \tag{1}$$

Where T_i denotes the gaze duration for the i-th instrument, a_i represents the weighting coefficient, and T_{max} signifies the total task duration threshold.

2.2 Real-time Constraints: Delay-Sensitive Decision Chains

The allocation of SPO tasks is fundamentally a multi-stage Markov decision process (MDP), a mathematical model for sequential decision-making used to study optimisation problems solvable via dynamic programming. Its response latency must satisfy the hard real-time requirements of the avionics system [8,9].

The state $s(t_k)$ is defined within the state space as a task-resource pair, representing the concatenation of the partition state $s_n^{t_k}$ within the RIMA system processing module at task arrival and the task state s_k. The resource requirements for the kth task, comprising computational and memory demands, are denoted as u_{k1} and u_{k2} respectively. Similarly, the computational resource utilisation and memory resource utilisation of partition n at this point are represented by $u_{k1}^{t_k}$ and $u_{k2}^{t_k}$. The execution time of the kth task is denoted as v_k [9].

$$s(t) = [s_1^{t_k}, \ldots, s_N^{t_k}, s_k] = [u_{11}^{t_k}, u_{12}^{t_k} \ldots, u_{N1}^{t_k}, u_{N2}^{t_k}, u_{k1}, u_{k2}, v_k] \tag{2}$$

Within the action space, the process of selecting actions is fundamentally a resource allocation procedure for reconfiguring avionics systems oriented towards SPO mode

switching. This involves loading task k into partition n. Assuming a total of N possible actions are available for selection, the action space may be defined as [9]:

$$a(t) = n \in \{1, 2, \ldots, N\} \tag{3}$$

2.3 Resource Constraints: Dynamic Adaptation of Heterogeneous Computing Resources

The hydropower system employs an Integrated Modular Architecture (IMA), which must satisfy real-time and reliability requirements within constrained physical resources. Within the dynamic partition management of computational resources, to prevent partition isolation caused by overload, the IMA mandates that the CPU utilisation of N partitions must satisfy:

$$\sum_{n=1}^{N} U_n^{CPU} \leqslant 80\% \tag{4}$$

Exceeding limits will trigger partition isolation. Ground-air communication bandwidth must ensure a transmission rate of ≥50Mbps. Topology optimization using graph neural networks (GNN) reduces cross-node communication latency by 37% [9].

2.4 Safety Constraints: An Airworthiness Certification-Driven Verification System

The SPO system must comply with the DO-178C Level A certification standard—DO-178C being the core standard for airworthiness certification of aviation software, with Level A applicable to flight safety-critical systems.

2.5 The Aforementioned Four-Dimensional Constraints Exhibit a Strong Coupling Relationship

Resource allocation impacts real-time performance, while human factors impose constraints on safety redundancy. For instance, CPU utilisation exceeding an 80% threshold will cause decision delays exceeding 200 milliseconds, thereby triggering excessive cognitive load for pilots. Consequently, we require hierarchical decoupling optimisation—implementing Proximal Policy Optimisation (PPO) for resource allocation, as shown in Figure 1.

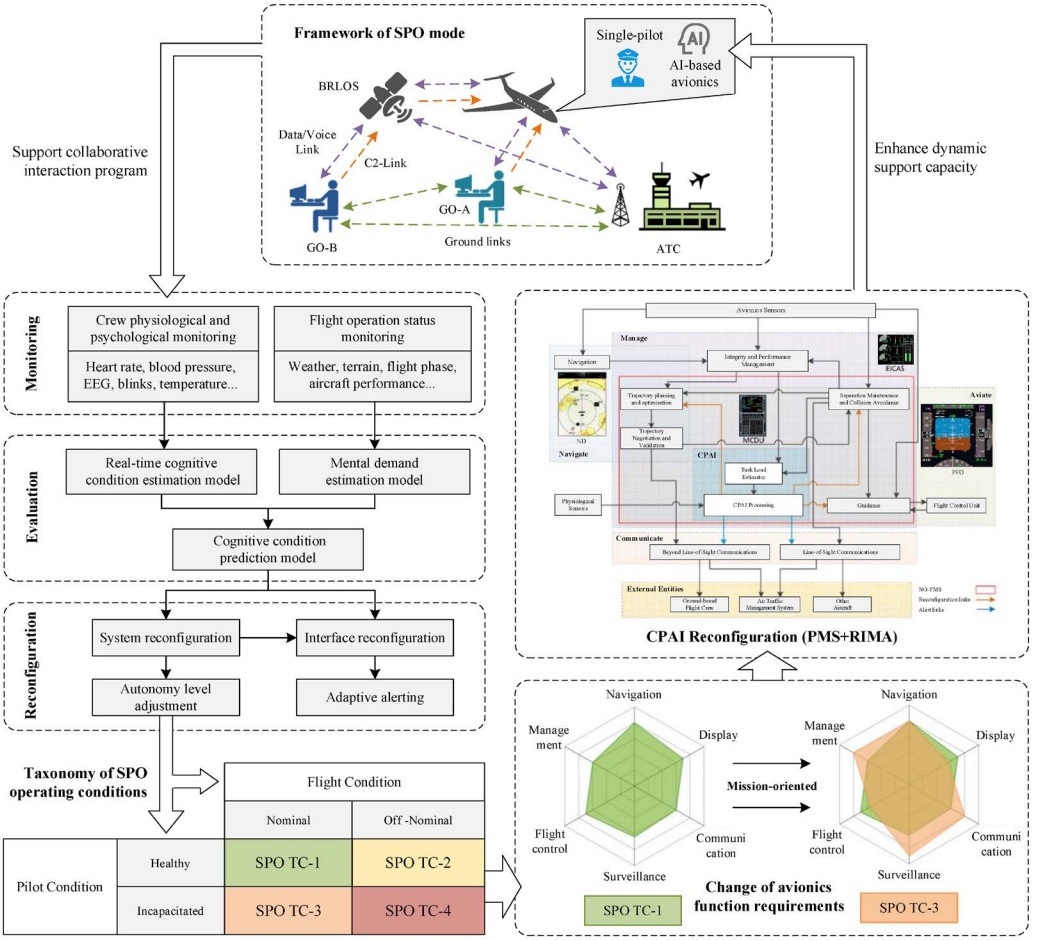

Fig. 1 Resource Allocation Framework for Avionics Systems in SPO Mode [9]

ISSN 2959-6157

3. Application of Reinforcement Learning in SPO Task Allocation Decision-Making

3.1 Collaborative Framework for Reinforcement Learning Algorithm Models

SPO reduces aviation operational costs by decreasing cockpit crew numbers, though it necessitates addressing sharply increased workloads and the allocation of unexpected tasks. RL demonstrates significant advantages in SPO task allocation by learning optimal strategies through interaction with the environment. Compared to model-based control and optimisation methods, RL provides a data-driven, learning-based framework for formulating and solving sequential decision problems. Figure 2 illustrates two classical RL algorithm categories: value-based and policy-based approaches [11].

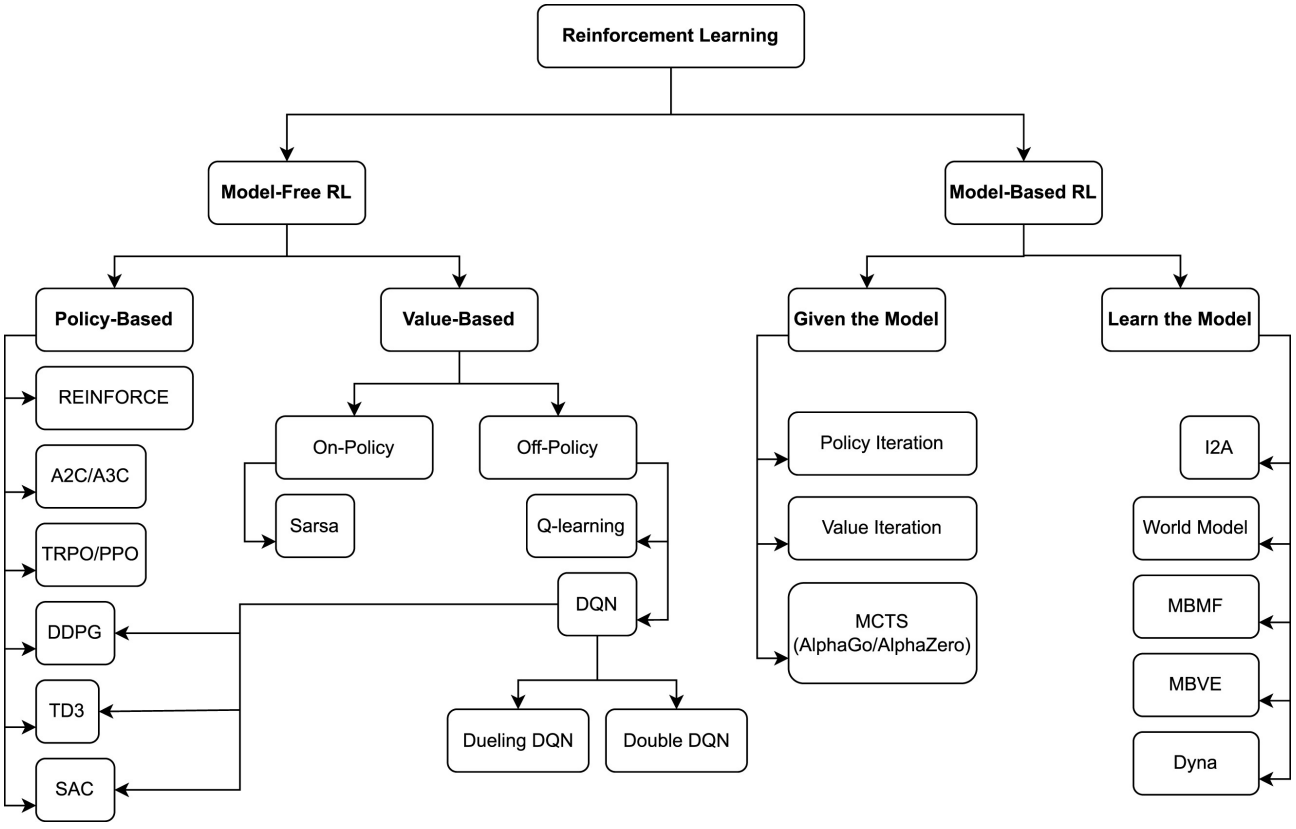

Fig. 2: Categories of Reinforcement Learning Models and Algorithms [11]

3.2 Multi-Constraint Decision Models

A coalition task allocation model is constructed that accounts for payload resource requirements, agent resource spatial constraints, and execution windows. By designing a Q-value network function approximator via Deep Q-Network (DQN), optimal coalition execution paths are generated to achieve adaptive scheduling optimisation for multi-agent systems [12].

Q-learning is one of the fundamental value-based RL algorithms introduced by Watkins[13] in the late 1980s. The Q-value for each combination of state and action in the environment can be defined as given by Equation (5). It represents the expected value of the cumulative reward over time step t for action (a) when following policy π:

$$Q_{\pi}(s,a) = \mathbb{E}_{\pi}[\Sigma_{i=0}^{\infty}\gamma^{i} r_{t+i+1} \mid s,a] \quad (5)$$

Where i denotes the number of steps forward in time step t. After updating the Q-values, the algorithm attempts to determine the value of taking a specific action in a particular state. The Q-table comprises all stored Q-values for every state-action pair in the discrete space. The policy $\pi(s) = argmaxQ(s,a)$ yields the highest total return. The agent selects an action to explore the environment and observe the outcome. Q-values may be updated using temporal difference techniques [11]:

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha[r_t + \gamma \max_{a_i®} Q(s_{t+1},a_{t+1}) - Q(s_t,a_t)] \quad (6)$$

Where $Q = r_t + \gamma \max_{a_i®} Q(s_{t+1},a_{t+1})$ is treated as the temporal difference target, with a denoting the learning rate. The

off-policy method learns the optimal policy value independently of the agent's actions.

3.3 Algorithm Performance Validation

The SPO model's distributed multi-agent coalition task allocation problem is transformed into a multi-stage sequential decision problem based on a Markov Decision Process (MDP). The fundamental elements of the DQN method are explicitly defined, including the state S, action A, state transition probability P, policy π, reward function R, and the state-action function Q(S, A; φ) [12]. Moreover, each interaction between the agent and the environment yields a reward signal. Reinforcement learning aims to maximise the cumulative total reward over a complete decision process, i.e., the aggregate reward within a single round. This total reward and the optimisation objective are termed the expected return. The simplest method for calculating the reward is shown in Equation (8) [7]:

$$G_t^{simple} = r_{t+1} + r_{t+2} + \ldots + r_T = T_{t+1} + G_{t-1}^{simple} = \sum_{k=0}^{T-t+1} r_{t+k+1} \quad (7)$$

Here, G_t^{simple} denotes the expected return at time t, and T represents the time at which the round concludes.

To validate the effectiveness of the proposed DQN-based SPO mode distributed multi-agent coalition task allocation method in coalition formation scenarios, a simulation case study was designed. Taking the dynamic allocation of limited resources among multi-agents within the SPO mode collaborative flight organization architecture as an example, the method testing comprises two parts. Firstly, the simulation selects optimal parameters for the method under complex constraints, testing their impact on simulation outcomes. Secondly, Q-Learning was introduced for comparative analysis to validate the convergence performance and computational efficiency of the designed coalition task allocation method. Finally, simulations demonstrated that DQN generates task allocation schemes with high real-time capability under complex constraints, achieving approximately 40% improvement in task allocation efficiency and reducing response time to 60% of the two-person mode [12], thereby providing an effective strategy for distributed multi-agent collaborative task allocation.

4. The Foundational Support Role Of Deep Learning in SPO Task Allocation

4.1 Intelligent Task Allocation and Scheduling

Deep learning plays a supporting role as the core decision-maker in SPO task allocation,It can perceive in real time,understand and make decisions for complex flight missions.It can continuously analyze a large amount of multimodal data from aircraft sensors, air traffic control instructions, meteorological data, etc, and then perform dynamic priority assessment.For example, input engine warnings, weather radar echoes, etc,It will process the model to calculate the key and urgency level of each task in real time, generating a dynamically optimal task prioritization order.

In addition, there is also resource optimal matching--To decide whether a task is completed most efficiently by a pilot, an AI agent, or a collaboration between both, we can adopt the model of a reward function.[13]

$$G(x_t, y_t) = E[r_t + \gamma max Q(x_{t+1}, y_{\text{¦}}^{-}; \varnothing^{-})] \quad (8)$$

Used to measure the immediate feedback an agent receives from the environment after executing an action[13]. Immediate rewards enable pilots to swiftly ascertain the effects of their current actions, allowing timely strategy adjustments to prevent short-term misjudgements from diminishing future rewards. This facilitates pilots' consideration of long-term benefits.

4.2 Multi-source Information Fusion and Understanding

Convolutional neural networks (CNNs) can be employed to process visual information. As a deep learning technique specifically designed for handling grid-like data such as images and videos, CNNs emulate the functioning of the human visual cortex to learn and recognise key features from images with exceptional efficiency. During phases such as landing and taxiing, CNNs can analyse camera video streams in real time, accurately identifying various signals, facilities, and other objects.

Even in complex environmental conditions, they maintain exceptionally high accuracy. Employing a three-layer convolutional neural network using the Super-Resolution Convolutional Neural Network (SRCNN) model enables not only automatic parameter learning but also endows the model with formidable resolution capabilities [14]. Should onboard cameras capture blurred imagery due to equipment limitations or meteorological factors, this model can perform super-resolution reconstruction. This allows pilots to discern critical information and make timely judgements.

4.3 Naturally efficient human-computer interaction

4.3.1 The blueprint for natural communication

Efficient task allocation requires smooth communication between pilots and the AI system, as well as accurate and

ISSN 2959-6157

precise voice data.With the continuous research and development of deep learning-related algorithms, they are becoming more suitable for natural language processing tasks.After the development of deep learning models, natural language processing (NLP) has rapidly advanced, Deep learning-based multi-layer neural networks can automatically learn languages from big data.

The breakthrough in the Transformer architecture represents a qualitative leap for the SPO model. Unlike the recurrent structure employed by RNNs, the Transformer employs a self-attention mechanism to design its encoder. This implies that it not only comprehends the intrinsic meaning of each individual lexical unit and its corresponding task, but also understands the inherent connections between each lexical unit and its task. Consequently, it is capable of making globally optimal decisions.

4.3.2 Accurate voice data

The SEGAN speech enhancement model possesses a degree of self-adaptability, enabling it to be trained under varying channel noise conditions while producing favourable enhancement results. Owing to the constraints inherent in air traffic control system voice communications, pilots find it challenging to obtain clean speech data. Consequently, the collected speech data is categorised based on its frequency-domain signal-to-noise ratio F_{SNR} to determine whether it constitutes a clean speech signal [15, 16]:

$$F_{SNR} = 20 \times log_{10} \frac{S_{std}\left(\left|f_{STFT}\left(A_{signal}\right)\right)\right.}{\mu_{mean}\left(\left|f_{STFT}\left(A_{signal}\right)\right)\right. + 10^{-4}} \tag{9}$$

This facilitates the identification of noise interference within specific frequency bands, thereby preventing unexpected incidents during flight that might overwhelm pilots. Subsequently, by employing deep learning-based filtering techniques to suppress noise within the corresponding bands, communication stability can be maintained.

5. The Coupling Mechanism of Reinforcement Learning and Deep Learning in SPO Dynamic Allocation

In the task allocation decision-making of the SPO mode, DL and RL establish a comprehensive 'perception-decision-optimisation' closed-loop enabling mechanism through multi-level, multi-dimensional deep collaboration. This mechanism not only demonstrates the inherent complementarity of the two technologies but, more significantly, achieves holistic optimisation of task allocation problems in complex aviation environments through systematic architectural design.

DL technology underpins the core functions of environmental perception and state understanding within SPO systems. Visual processing modules built upon convolutional neural networks enable real-time analysis of multi-source heterogeneous data within the cockpit, encompassing visual information from instrument panels, pilot gestures, and external environmental video streams. Through multi-level feature extraction, these networks transform raw pixel-level inputs into high-level representations of environmental states, providing precise inputs for subsequent decision-making. Particularly under complex meteorological conditions, DL models demonstrate robust adaptability, effectively handling challenging scenarios such as low visibility and blurred imagery. Concurrently, the recurrent neural network architecture focuses on processing temporal data. Through long-short term memory mechanisms, it captures dynamic information such as pilot operational habits and mission execution trajectories, enabling accurate prediction of system state evolution trends. This temporal modelling capability is crucial for preventing cognitive overload, as the system can proactively identify load accumulation patterns, thereby providing a time window for active intervention.

However, DL possesses inherent limitations in dynamic decision-making, precisely where RL plays a pivotal role. Through reward-driven mechanisms, RL models task allocation problems as sequential decision processes, generating optimal task scheduling strategies based on the environmental state representations provided by DL. In multi-agent collaborative scenarios, value function approximation methods effectively handle high-dimensional state spaces, enabling efficient exploration of complex decision spaces via algorithms such as deep Q-learning. When the system detects unexpected events or resource conflicts, policy gradient methods can rapidly adjust task priorities to ensure critical operations receive timely responses. Notably, RL demonstrates unique advantages in handling uncertainty. Through a balanced exploration-exploitation mechanism, the system maintains robust decision-making even in partially observable environments.

In the practical deployment of the SPO system, this collaborative mechanism is concretely realised through the resource allocation framework. The avionics system resource allocation framework depicted in Figure 1 profoundly demonstrates the synergistic advantages of RL and DL. This framework employs deep learning to monitor resource utilisation across partitions in real time, predicting shifts in computational demands during task execution. Concurrently, the RL component dynamically adjusts task scheduling strategies based on these predictions, ensuring the system consistently operates at its optimal operating point. This collaborative design enables the

system to maximise resource utilisation efficiency while meeting stringent real-time requirements.

This deep collaborative mechanism between deep learning and reinforcement learning not only provides technical assurance for single-pilot operation modes but, more significantly, pioneers a new paradigm for intelligent aviation decision-making systems. Through systematic architectural innovation and algorithmic breakthroughs, it offers robust technical support for navigating increasingly complex aviation operational environments, thereby propelling aviation intelligence technology towards higher levels of advancement.

6. Conclusions

This paper systematically reviews the pivotal roles and synergistic mechanisms of RL and DL in SPO task decision-making, clarifying their application pathways and efficacy under multi-dimensional constraints.

The SPO task allocation system faces four-dimensional strong coupling constraints involving human factors, real-time requirements, resources, and security. Traditional rule engines struggle to achieve dynamic optimisation. Reinforcement learning employs algorithms such as multi-agent near-term policy optimisation and deep Q-networks to model task allocation as a Markov decision process. In simulations, this approach significantly enhances task allocation efficiency, reduces response latency, and demonstrates exceptional sequential decision-making capabilities. Deep learning serves as the foundational support for perception and representation, employing models such as CNNs, SRCCNs, and Transformers to process multimodal data. This enhances situational awareness, natural interaction, and cognitive state recognition, providing high-quality environmental state representations for decision-making. RL and DL do not operate in isolation but form a deep collaborative mechanism through a 'perception-decision-optimisation' closed loop. DL supports RL's decision inputs through front-end perception and state representation, reducing interaction latency and state uncertainty. RL, in turn, dynamically generates task allocation strategies via reward-driven policy optimisation, while leveraging techniques such as policy extraction to influence DL model training and refinement. Together, they construct an adaptable, responsive intelligent decision system, providing a viable pathway for achieving efficient collaboration among 'pilot-airborne agent-ground station'. Nevertheless, integrating RL and DL within SPO systems remains fraught with challenges. Future research must first address the strategic mismatch and insufficient adaptability of a single system, algorithm, or model when confronted with various uncertainties, disturbances, or environmental changes in dynamic environments—particularly across flight phases and under extreme scenarios.

Most of the models and formulas referenced herein have now reached a relatively mature stage of development and can be applied to assess the feasibility of SPO modes. However, certain measures remain theoretical in nature, and their practical implementation will require continued effort. It is hoped that in the near future, the shortage of pilots can be alleviated through the adoption of single-pilot operation modes.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] IATA. Global Pilot Shortage Forecast 2035[R]. Montreal: International Air Transport Association, 2023.

[2] Wang M, Xiao G, Wang G Q. Single Pilot Operations Technology[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(4): 202-220.

[3] Zhang T R, Wei Z Q, Shi T Y. Design of Air-Ground Collaborative Operation Mode Based on SPO[J]. Journal of Civil Aviation University of China, 2020, 38(6): 12-17.

[4] Bilimoria K D, Johnson W W, Schutte P C. Conceptual Framework for Single Pilot Operations[C]. International Conference on Human-Computer Interaction in Aerospace, 2014: 1-8.

[5] Liu W, Yang J Z. Review of New Human-Computer Interaction Methods in Civil Aircraft Cockpits[J]. Chinese Journal of Ergonomics, 2024, 30(3): 81-86.

[6] Yang L. Research on Multi-Agent Aircraft Path Planning Based on Deep Reinforcement Learning[D]. Harbin Engineering University, 2024.

[7] Liu H F. Research on Aircraft Approach Decision-Making in Terminal Area Based on Deep Reinforcement Learning[D]. Sichuan University, 2024.

[8] Dong L, Liu J, Sun Z, et al. Resource Allocation Approach of Avionics System in SPO Mode Based on Proximal Policy Optimization[J]. Aerospace, 2024, 11(10): 812-812.

[9] Shan S Z, Zhang W W. Air Combat Intelligent Decision-Making Method Based on Self-Play and Deep Reinforcement Learning[J]. Acta Aeronautica et Astronautica Sinica, 2024, 45(4): 206-218.

[10] Ye K W, Bao H, Wei S D. Layout Optimization for Aircraft Cockpit Man-Machine Interface Based on Visual Attention Distribution[J]. Journal of Nanjing University of Aeronautics and Astronautics, 2018, 50(3): 416-421.

[11] Razzaghi P, Tabrizian A, Guo W, et al. A Survey on Reinforcement Learning in Aviation Applications[J]. Engineering Applications of Artificial Intelligence, 2024, 136: 108911.

[12] Dong L, Chen H B, Chen X, et al. Distributed Multi-Agent

ISSN 2959-6157

Coalition Task Allocation Strategy for Single Pilot Operations Based on DQN[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(13): 180-195.

[13] Cheng W Y. Research on Anomaly Identification of Network Traffic Data Based on Deep Reinforcement Learning[J]. Journal of Intelligent Internet of Things Technology, 2025, 57(4): 87-91.

[14] Li C, He S Q, Wei X, et al. Automatic Identification Method for Anomaly Traffic in Elastic Optical Networks Based on Isolation Forest Algorithm[J]. Laser Journal, 2024, 45(1): 179-183.

[15] Tian C W, Song M J, Zuo W M, et al. Application of Convolutional Neural Networks in Image Super-Resolution[J]. CAAI Transactions on Intelligent Systems, 2025, 20(3): 719-749.

[16] Wang Y Z, Li X, Mou R, et al. Air Traffic Control Speech Enhancement Algorithm Based on Improved SEGAN[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(12): 3930-3939.