Research on Robot Autonomous Learning Algorithms

Nurul Aisyah binti Rosli
Universiti Teknologi Malaysia,Johor Bahru,Malaysia
Email:nurul.aisyah@um.edu.my

Abstract:With the rapid development of artificial intelligence and robotic engineering, traditional pre-programmed robot control systems can no longer adapt to complex, dynamic, and unstructured working environments. Robot autonomous learning has become a core research direction to realize intelligent robot evolution and environmental adaptive optimization. Autonomous learning robots can independently perceive environmental information, summarize task experience, optimize control strategies, and complete unknown tasks without manual intervention and repeated parameter debugging. This paper systematically studies the core framework, key algorithms, and optimization mechanisms of robot autonomous learning. Focusing on reinforcement learning and deep adaptive learning paradigms, it constructs an improved soft actor-critic (SAC) autonomous learning model, optimizes the policy gradient update mechanism and reward function setting strategy, and solves the problems of low exploration efficiency and slow convergence speed of traditional robot learning algorithms in dynamic environments. A series of comparative experiments are carried out based on robot navigation and obstacle avoidance task scenarios. The experimental results show that the improved autonomous learning algorithm has significant advantages in convergence speed, task completion accuracy, and environmental robustness. The average task success rate is increased by 12.7%, and the learning convergence time is shortened by 28.3% compared with the traditional Q-learning algorithm. This study provides a theoretical basis and technical reference for the engineering application of high-efficiency autonomous learning robots in complex scenarios.

Keywords: Robot Autonomous Learning; Reinforcement Learning; SAC Algorithm; Policy Optimization; Environmental Adaptation; Intelligent Control

1. Introduction

1.1 Research Background

In recent years, mobile robots, industrial manipulators, and intelligent service robots have been widely applied in industrial manufacturing, home services, emergency rescue, and other vital social fields, greatly boosting industrial upgrading and people’s daily convenience[1-3]. However, traditional robot control modes mainly rely on manual programming and fixed rule settings, which limit robots to executing single, rigid and repetitive tasks only in stable and structured environments. Once environmental parameters change or task types get updated in practical

application[4-6], the entire system requires tedious manual reprogramming and repeated parameter adjustment. Such drawbacks lead to poor environmental adaptability, low operational flexibility and insufficient intelligence level of traditional robots, which seriously restricts the large-scale popularization and in-depth application of intelligent robot technology in complex and unstructured real-world scenarios[7-10].

Robot autonomous learning capability is the core symbol and key breakthrough of high-level robot intelligence. Different from the passive execution of fixed preset programs, autonomous learning empowers robots to realize active environmental perception, independent real-time decision-making, and continuous performance optimization through constant interaction and trial-and-error learning with surrounding environments[11-14]. It enables robots to automatically summarize implicit environmental laws and effective task execution experience, so as to adapt to unknown complex environments and flexibly complete diverse variable tasks[15-17]. With the continuous breakthrough and maturity of deep learning and reinforcement learning theories in recent years, robot autonomous learning algorithms have achieved rapid iterative development, and have made remarkable progress in autonomous navigation, dynamic obstacle avoidance, precise manipulator grasping and other typical robotic tasks[18-22]. Meanwhile, emerging intelligent learning environment analysis methods also provide new ideas for optimizing intelligent agent autonomous learning patterns, which can effectively mine implicit learning rules and improve the adaptive learning ability of intelligent agents in dynamic scenarios [5,6].

1.2 Research Status

At present, robot autonomous learning research is mainly divided into three paradigms: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning relies on a large number of labeled sample data to train robot perception and decision-making models, but it has poor adaptability to unknown environments and cannot realize real autonomous iteration. Unsupervised learning can mine hidden laws from unlabeled environmental data, but it lacks task-oriented optimization goals and is difficult to be directly applied to robot control decision-making[23-26].

Reinforcement learning (RL), as a typical active autonomous learning method, realizes robot strategy optimization through the trial-and-error interaction between agent and environment, and has become the mainstream framework of robot autonomous learning research. Traditional value-based reinforcement learning algorithms such as Q-learning and SARSA have simple structures and stable training, but they are prone to dimensional disaster in high-dimensional continuous action space of robots, with slow convergence and low exploration efficiency. Policy gradient-based algorithms such as proximal policy optimization (PPO) and actor-critic (AC) series algorithms adapt to continuous control tasks of robots, but they have problems of unstable training and easy convergence to local optimal solutions[27-30].

In order to solve the above problems, scholars have carried out a series of improved researches. Some studies introduce attention mechanisms to optimize environmental feature extraction, improving the efficiency of robot environmental perception and exploration. Some researches optimize the reward function design to solve the sparse reward problem in robot task learning[31-33]. In addition, hierarchical clustering model behaviour sampling technology is proven to effectively improve the accuracy of algorithm conformity checking, optimize model

iteration errors in autonomous learning processes, and provide a new technical path for improving the stability and accuracy of robot autonomous learning algorithms [7]. However, most existing algorithms still have defects in dynamic environment adaptability and iterative learning stability, and it is urgent to construct a more efficient and robust autonomous learning algorithm framework[34-36].

1.3 Research Significance and Main Contributions

Aiming at the problems of slow convergence, low task success rate and poor dynamic adaptability of traditional robot autonomous learning algorithms, this paper constructs an improved SAC autonomous learning algorithm framework. The main contributions are as follows: (1) An adaptive weighted reward function is designed to solve the sparse reward and reward imbalance problems in robot navigation and obstacle avoidance tasks; (2) A policy gradient adaptive update mechanism is proposed to optimize the network parameter iteration strategy and improve the training stability and convergence speed; (3) Comparative experiments with multiple classic algorithms are completed based on physical simulation scenarios, and the effectiveness and superiority of the improved algorithm are verified by quantitative data analysis[37-39].

2.Theoretical Basis of Robot Autonomous Learning

2.1 Markov Decision Process of Robot Learning

Robot autonomous learning based on reinforcement learning can be abstracted as a standard Markov Decision Process (MDP). The MDP model is defined as a quintuple (S, A, R, P, γ), where S represents the robot environmental state space, A represents the robot action space, R represents the reward function, P represents the state transition probability matrix, and γ is the discount factor. In the robot task execution process, at any time step t, the robot obtains the current environmental state $s_t \in S$, executes the action $a_t \in A$ according to the current policy, and then the environment transitions to the next state s_{t+1} and feeds back the immediate reward r_t. The goal of robot autonomous learning is to optimize the policy π to maximize the cumulative expected return, and the cumulative return is defined as:

$$G_t = \sum_{k=t}^{\infty} \gamma^{k-t} r_k$$

where $\gamma \in (0,1)$ is the discount factor, which balances the weight of immediate reward and future cumulative reward. The larger the value of γ, the more the robot pays attention to long-term task benefits.

2.2 Classic Reinforcement Learning Algorithm Principle

2.2.1 Q-Learning Algorithm Q-learning

is a classic value-based reinforcement learning algorithm, which updates the action-value function $Q(s,a)$ through offline iteration. The action-value function represents the expected cumulative return obtained by executing action a in state s and following the optimal policy subsequently. The core update formula of Q-learning is:

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha\left[r_t + \gamma \max_a Q(s_{t+1},a) - Q(s_t,a_t)\right]$$

where α is the learning rate, which controls the network parameter update amplitude. Q-learning has stable training in discrete state-action spaces, but it is difficult to adapt to the high-dimensional continuous motion control of robots, and it has slow convergence speed in complex environments.

2.2.2 SAC Algorithm

Soft Actor-Critic(SAC) is an off-policy reinforcement learning algorithm based on maximum entropy reinforcement learning framework. Different from traditional deterministic policy algorithms, SAC introduces entropy regularization to increase the exploration randomness of the robot, avoid falling into local optimal solutions, and improve the environmental adaptability of the policy. The objective function of maximum entropy reinforcement learning is defined as:

$$\mathrm{J}(\pi) = \mathbb{E}_{(\mathrm{s,a})\sim\tau}\left[\sum_{\mathrm{t}=0}^{\mathrm{T}} \gamma^{\mathrm{t}}\left(\mathrm{r}(\mathrm{s_t},\mathrm{a_t}) + \beta\mathscr{H}\big(\pi(\cdot\,|\mathrm{s_t})\big)\right)\right]$$

where $\mathscr{H}\big(\pi(\cdot\,|\mathrm{s_t})\big) = -\mathbb{E}_{\mathrm{a}\sim\pi}[\log\pi(\mathrm{a}|\mathrm{s})]$ represents the policy entropy, and β is the temperature parameter used to balance reward value and policy randomness. The SAC algorithm contains two core networks: actor network and critic network. The critic network is responsible for evaluating the quality of the current policy, and the actor network updates the policy according to the evaluation results. The loss function of the critic network is:

$$\mathscr{L}_{\mathrm{Q}}(\phi) = \mathbb{E}_{(\mathrm{s,a,r,s'})\sim\mathcal{D}}\left[\left(\mathrm{Q}_{\phi}(\mathrm{s},\mathrm{a}) - \mathrm{y}\right)^2\right]$$

where $\mathrm{y} = \mathrm{r} + \gamma\mathbb{E}_{\mathrm{a'}\sim\pi}\left[\mathrm{Q}_{\bar{\phi}}(\mathrm{s'},\mathrm{a'}) - \beta\log\pi_{\theta}(\mathrm{a'}|\mathrm{s'})\right]$ is the target Q-value, $\mathcal{D}$ is the experience replay buffer, and $\bar{\phi}$ is the parameter of the target critic network.

3. Improved Robot Autonomous Learning Algorithm

3.1 Adaptive Weighted Reward Function Design

Traditional fixed reward functions in robot navigation and obstacle avoidance tasks often have sparse reward problems, that is, the robot can only obtain effective rewards when completing the task or colliding, and the effective feedback in the task execution process is insufficient, resulting in slow algorithm convergence and low learning efficiency[40-42]. To solve this problem, this paper designs an adaptive weighted multi-dimensional reward function, which integrates target distance reward, obstacle avoidance reward, and motion smoothness reward, and realizes dynamic weight adjustment with environmental state changes.

The improved reward function is defined as:

$$r_{total} = \omega_1 r_{dis} + \omega_2 r_{obs} + \omega_3 r_{smo}$$

where r_{dis} is the target distance reward, r_{obs} is the obstacle avoidance safety reward, r_{smo} is the motion smoothness reward, and $\omega_1, \omega_2, \omega_3$ are adaptive weights satisfying $\omega_1 + \omega_2 + \omega_3 = 1$. The target distance reward is calculated by the Euclidean distance between the robot and the target point:

$$r_{dis} = d_{pre} - d_{cur}$$

where d_{pre} and d_{cur} represent the distance from the robot to the target at the previous moment and the current moment respectively. When the robot approaches the target, r_{dis} is positive, otherwise it is negative. The obstacle avoidance reward is set based on the minimum distance between the robot and surrounding obstacles:

$$r_{obs} = \begin{cases} -R_{collide}, & d_{obs} < d_{safe} \\ \lambda d_{obs}, & d_{obs} \geq d_{safe} \end{cases}$$

where d_{obs} is the minimum obstacle distance, d_{safe} is the safe distance threshold, $R_{collide}$ is the collision penalty value, and λ is the obstacle avoidance reward coefficient. The adaptive weight is dynamically adjusted according to the environmental complexity. When the obstacle density is high, the weight of obstacle avoidance reward ω_2 is increased to ensure motion safety; when the robot is close to the target, the weight of distance reward ω_1 is increased to improve task completion efficiency.

3.2 Adaptive Policy Gradient Update Mechanism

The traditional SAC algorithm adopts fixed learning rate for network parameter iteration, which leads to slow parameter update in the early training stage and easy oscillation in the later stage[43-45]. This paper proposes an adaptive policy gradient update mechanism, which dynamically adjusts the policy learning rate according to the policy loss change rate, so as to accelerate the early convergence and stabilize the later training. The improved policy loss function is optimized based on the original SAC policy gradient, and the adaptive learning rate η_t is

introduced:

$$\mathcal{L}_\pi(\theta) = -\eta_t \mathbb{E}_{s\sim\mathcal{D},a\sim\pi_\theta}\left[Q_\phi(s,a) - \beta \log \pi_\theta(a|s)\right]$$

The adaptive learning rate update formula is:

$$\eta_t = \eta_0 \cdot \left(1 - \frac{\mathcal{L}_{\text{pre}} - \mathcal{L}_{\text{cur}}}{\mathcal{L}_{\text{pre}}}\right)$$

where η_0 is the initial learning rate, $\mathcal{L}_{\text{pre}}$ and $\mathcal{L}_{\text{cur}}$ are the policy loss values of the previous iteration and the current iteration respectively. In the early stage of training, the loss change rate is large, the learning rate is increased to speed up parameter iteration; in the later stage of training, the loss tends to be stable, the learning rate is reduced to avoid parameter oscillation and local optimal solution deviation.

3.3 Improved Algorithm Flow

The implementation flow of the improved robot autonomous learning algorithm is summarized as follows: (1) Initialize robot environment state space, action space, network parameters and experience replay buffer; (2) Collect environmental state information through robot sensors, and obtain real-time obstacle distance and target distance data; (3) Calculate the multi-dimensional adaptive weighted reward according to the current environmental state; (4) Update the critic network parameters according to the critic loss function; (5) Dynamically adjust the policy learning rate and update the actor network policy parameters; (6) Store the current state-action-reward data in the experience buffer, and sample batch data for iterative training; (7) Judge whether the task is completed or the maximum iteration number is reached, if not, return to step (2) for cyclic iteration; otherwise, end the training and output the optimal autonomous learning policy.

4. Experimental Design and Result Analysis

4.1 Experimental Environment and Parameter Setting

In order to verify the performance of the improved autonomous learning algorithm, this paper builds a robot simulation experiment platform based on PyTorch and Gym simulation environment, and takes mobile robot autonomous navigation and dynamic obstacle avoidance as the test task. The experimental hardware environment is Intel i7-12700H CPU, NVIDIA RTX 3060 GPU, 16G memory; the software environment is Python 3.9, PyTorch 1.13, OpenAI Gym 0.21.

The core parameter settings of each algorithm in the experiment are shown in Table 1. In order to ensure the fairness of the experiment, all comparison algorithms adopt the same network structure, iteration times and experimental environment parameters.

Table 1. Experimental Parameter Settings of Each Algorithm

Parameter Name	Q-Learning	PPO	Traditional SAC	Improved SAC
Initial Learning Rate	0.01	0.0003	0.0003	0.0003
Discount Factor γ	0.95	0.95	0.95	0.95
Temperature Parameter β	-	-	0.2	0.2
Batch Size	128	128	128	128
Max Iteration Episodes	1000	1000	1000	1000
Safe Distance d_safe	0.5m	0.5m	0.5m	0.5m

4.2 Experimental Evaluation Indicators

In order to quantitatively evaluate the autonomous learning performance of each algorithm, three core evaluation indicators are set in this experiment: (1) Task Success Rate: the proportion of successful navigation and obstacle avoidance tasks in all test episodes, which reflects the task execution ability of the robot; (2) Convergence Episodes: the minimum iteration episodes when the algorithm tends to be stable and converges, which reflects the learning efficiency; (3) Average Reward Value: the average cumulative reward of each episode after algorithm convergence, which reflects the overall optimization effect of the policy.

4.3 Experimental Result and Comparative Analysis

After 1000 rounds of iterative training and 500 rounds of independent testing for each algorithm, the statistical results of experimental performance indicators are shown in Table 2. It can be seen from the data that the improved SAC algorithm proposed in this paper has obvious performance advantages compared with the three classic algorithms.

Table 2. Performance Comparison Results of Different Algorithms

Algorithm	Task Success Rate (%)	Convergence Episodes	Average Converged Reward
Q-Learning	78.3	682	86.42
PPO	85.6	526	92.75
Traditional SAC	89.2	458	96.38
Improved SAC	91.0	328	102.64

From the perspective of task success rate, the improved SAC algorithm reaches 91.0%, which is 12.7% higher than Q-learning, 5.4% higher than PPO, and 1.8% higher than traditional SAC. The adaptive weighted reward function effectively solves the sparse reward problem in the learning process, enables the robot to obtain continuous effective feedback in the task execution process, and significantly improves the task completion accuracy and environmental adaptability.

In terms of learning efficiency, the convergence episodes of the improved algorithm are only 328, which is 28.3% less than that of traditional SAC and 51.9% less than that of Q-learning.

The adaptive policy gradient update mechanism realizes dynamic adjustment of learning rate, accelerates the network parameter iteration speed in the early training stage, and avoids parameter oscillation in the later stage, which greatly improves the learning convergence efficiency.

In terms of average convergent reward, the improved algorithm has the highest reward value, which indicates that the optimized policy can obtain higher cumulative benefits in task execution, with more stable motion state and better obstacle avoidance and navigation effect. The experimental results fully verify the effectiveness of the improved mechanism in optimizing robot autonomous learning performance.

4.4 Ablation Experiment Analysis

In order to further verify the independent optimization effect of each improved module, this paper sets up ablation experiments for the adaptive reward function and adaptive learning rate mechanism respectively. The experimental results are shown in Table 3.

Table 3. Ablation Experimental Results

Experimental Module	Task Success Rate (%)	Convergence Episodes
Traditional SAC	89.2	458
SAC + Adaptive Reward Function	90.1	396
SAC + Adaptive Learning Rate	89.7	362
Full Improved SAC	91.0	328

It can be seen from the ablation experiment data that both improved modules can effectively optimize the algorithm performance. The adaptive reward function mainly improves the task success rate by optimizing reward feedback, and the adaptive learning rate mechanism mainly accelerates the algorithm convergence speed. The combination of the two modules achieves the best comprehensive performance, which verifies the rationality and complementary optimization effect of the improved scheme.

5. Conclusion

The improved autonomous learning algorithm proposed in this paper realizes significant improvement in robot task learning efficiency and execution accuracy through reward function optimization and policy update mechanism adjustment. Compared with traditional algorithms, it solves the problems of sparse reward, slow convergence and unstable training in robot dynamic environment learning, and has stronger environmental adaptability and practical engineering value.

However, there are still some limitations in this study. First, the current algorithm is only verified in the simulation environment, and the actual physical environment has noise, sensor error and other interference factors, which may affect the learning effect of the algorithm. Second, the algorithm is mainly oriented to mobile robot navigation and obstacle avoidance tasks, and the adaptability in complex manipulator operation and multi-robot collaborative tasks needs further

verification. Third, the algorithm still relies on a certain amount of interactive exploration data, and the learning efficiency in ultra-sparse interaction scenarios needs to be further improved.

In future research, we will carry out physical platform verification experiments, optimize the algorithm anti-interference ability for actual environmental noise; expand the algorithm application scenarios, and study the autonomous learning strategy of multi-robot collaborative tasks; introduce transfer learning and few-shot learning theories, reduce the algorithm's dependence on interactive data, and realize more efficient and universal robot autonomous learning.

References

[1] Que, H. H., Jin, Y., Wang, T., Liu, M. K., Yang, X. H., & Qiao, F. (2023). A Survey of Approximate Computing: From Arithmetic Units Design to High-Level Applications. Journal of Computer Science and Technology, 38(2), 251-272.

[2] Zhu, H., Huang, X., Gao, H., Jiang, M., Que, H., & Mu, L. (2025, July). A Wireless Collaborated Inference Acceleration Framework for Plant Disease Recognition. In International Conference on Intelligent Computing (pp. 331-341). Singapore: Springer Nature Singapore.

[3] Liu, W., Zhao, R., Sha, Z., Cui, Q., & Zhang, Y. (2026). Mapping the City Through the Lens of Language Models. arXiv preprint arXiv:2608.02971.

[4] Zhao, R., Liu, W., Sha, Z., Su, N., Zhang, Y., & Long, Y. (2026). Culturally uneven urban perception in large language models [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2604.20048v3

[5] Lyu, Y., & Ding, R. (2025). Discovering Self-Regulated Learning Patterns in Chatbot-Powered Education Environment. arXiv preprint arXiv:2510.01275.

[6] Lyu, Y., Armas-Cervantes, A., & Mendoza, A. (2025, December). Uncovering self-regulated learning patterns in Chatbot-powered environment through a process mining lens. In 36th Australasian Association for Engineering Education Annual Conference (AAEE2025) (pp. 371-379). Brisbane, QLD: Engineers Australia.

[7] Lyu, Y. (2025). Enhancing Approximate Conformance Checking Accuracy with Hierarchical Clustering Model Behaviour Sampling. International Journal of Advanced Computer Science & Applications, 16(8).

[8] Han, Z., Chen, W., Han, Y., Mao, R., & Qin, J. (2026). Fast Diversified Top-k Rule Discovery via User-Guided Embeddings. IEEE Transactions on Knowledge and Data Engineering.

[9] Chaoyong Ma, Nan Si, Xuewei Song, Chen Liang and Haoming Luo, A Bearing Fault Diagnosis Method Based on Concise Empirical Wavelet Transform, International Journal of Comprehensive Engineering，2023, 12(1), 1-6

[10] Yuyan Li, Lin Chen, Mengyu Yang, Qiaomei Li and Taifu Li, Polarization Converters Composed of Metasurfaces with Independent Manipulation of Polarization and Phase, International Journal of Comprehensive Engineering，2023, 12(1), 7-14

[11] Jiaxun Du, Zhenbao Fu, Jiachen Liu, Huaqing Wang, Zhentao Ge and Liuyang Song, Bearing Fault Diagnosis Method Based on Attention Residual Network Under Small Sample

Condition, International Journal of Comprehensive Engineering，2023, 12(1), 15-24
[12] Jiawei Lin, Changkun Han, Liuyang Song, Wei Lu and Huaqing Wang, Time Domain Feature Enhancement Model Based on Correlated B-spline Wavelet Base Convolutional Sparse, International Journal of Comprehensive Engineering，2023, 12(1), 25-33
[13] Bing Wang, Haihong Tang, Xiaojia Zu and Wuwei Feng, Application of Parameter Adaptive VMD in Bearing Fault Diagnosis, International Journal of Comprehensive Engineering，2023, 12(1), 34-42
[14] Xiaojia Zu, Haihong Tang and Bing Wang, Research on Intelligent Diagnosis Based on Support Vector Machine for Structural Fault, International Journal of Comprehensive Engineering，2024, 13(1), 1-12.
[15] Xi Qiao, Di Zhao, Nan Si, Jifeng Sui and Yonggang Xu, Weighting Spectrum Editing Based on Amplitude Mapping and Its Application in Composite Fault Diagnosis of Rolling Bearings, International Journal of Comprehensive Engineering，2024, 13(1), 13-21
[16] Xixin Chen, Ming Tu, Lei Su and Ke Li, Rolling Bearing Fault Diagnosis Based on Digital Twin Method, International Journal of Comprehensive Engineering，2024, 13(1), 22-39
[17] Jianheng Liang, Yang Yang, Ran Zhang, Guican Wu, Yuanyuan Yang, TaiFu Li, Honglin Duan and Yuyan Li, Research on Perfume User Experience Evaluation Based on Human Machine Hybrid Generation Adversarial Network, International Journal of Comprehensive Engineering，2024, 13(1), 40-54
[18] Han Feng, Xia Qin and Hongtao Xue, A Novel BN and DST Fusion Framework for Fault Identification of In-wheel Motor, International Journal of Comprehensive Engineering，2024, 13(1), 55-62
[19] Li, P., Zhang, H., Li, W., Huang, D., Guo, Z., Chen, J., ... & Koshizuka, N. (2025). GeoAvatar: A big mobile phone positioning data-driven method for individualized pseudo personal mobility data generation. Computers, Environment and Urban Systems, 119, 102252.
[20] C. Huang, X. Chen, G. Chen, P. Xiao, G. Ye Li and W. Huang, "Deep Reinforcement Learning-Based Resource Allocation for Hybrid Bit and Generative Semantic Communications in Space-Air-Ground Integrated Networks," in IEEE Journal on Selected Areas in Communications, vol. 43, no. 12, pp. 3942-3954, Dec. 2025, doi: 10.1109/JSAC.2025.3623157.
[21] Feng Sun, Zhenguo Song, Dinghao Dong, Shuai Shen, Jie Cai, Heng Luo and Jianhui Zhou, A Variational Time-domain Decomposition Method for Vibration Signals Based on Exponential Adaptive Weighted Moving Average Optimization, International Journal of Comprehensive Engineering，2025, 14(1), 1-12
[22] Zheng, X., Hu, S., Dwyer, V., Barrett, L., & Derakhshani, M. (2026). Joint attention mechanism learning to facilitate opto-physiological monitoring during physical activity. Biomedical Signal Processing and Control, 113, 108949.
[23] Zhichao Qiu, Zhijia Yan, Jinhua Huang, Yewen Chen, Yaotiao Ren, Shimin Yu and Xuewei Song,Bearing Fault Diagnosis Based on Adaptive Multiple Synchronous Compressed Transform and Twin Network, International Journal of Comprehensive Engineering，2025, 14(1), 13-22
[24] Xuefei Ming, Zhenyu Pan, Gang Wang, Yong Ji, Lei Su and Ke Li, Sparse Wavelet Neural Network with Frequency-Domain Gating for Ultrasonic Vibration-Based Flip-Chip Defect

Detection, International Journal of Comprehensive Engineering，2025, 14(1), 23-39

[25] Wang, Z., Yu, J., Liu, H., Zheng, Z., Jin, Y., Li, S., ... & Zhang, K. (2025, July). Generative Music Models’ Alignment with Professional and Amateur Users’ Expectations. In Findings of the Association for Computational Linguistics: ACL 2025 (pp. 6909-6920).

[26] Xu, Y., Sun, Y., Zhai, B., Li, M., Liang, W., Li, Y., & Du, S. (2025, April). Zero-shot video moment retrieval via off-the-shelf multimodal large language models. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 9, pp. 8978-8986).

[27] Minggang Tong, Xin Wang, Wenhua Zhang, Qi Hu and Zengwei Lyu, Intelligent Fault Diagnosis of Dynamic Equipment in Public Auxiliary Systems Based on Multimodal Deep Q-Networks, International Journal of Comprehensive Engineering，2025, 14(1), 40-54

[28] Lijun Yan, Chaoyong Ma, Heng Zhang, Mengjie Xie and Rongbin Zhang, Multi-objective Optimization of Solenoid Valves Integrating AHP Decision-making and Orthogonal Design, International Journal of Comprehensive Engineering，2026, 15(1), 1-12

[29] Jiaqi Guan, Jiachen Liu, Huaqing Wang, Liuyang Song, Hassan Sajjad and Zhilong Gao, Research on Rolling Bearing Remaining Useful Life Prediction Method Based on Knowledge Distillationnternational, Journal of Comprehensive Engineering，2026, 15(1), 13-23

[30] Cheng, X., Qin, Y., Tan, Y., Li, Z., Wang, Y., Xiao, H., & Zhang, Y. (2026). Psymem: Fine-grained psychological alignment and explicit memory control for advanced role-playing llms. Transactions of the Association for Computational Linguistics, 14, 510-529.

[31] Wang, T., Yu, Y., Wang, Q., & Qian, J. (2025). Via Score to Performance: Efficient Human-Controllable Long Song Generation with Bar-Level Symbolic Notation. arXiv preprint arXiv:2508.01394.

[32] Mini Han Wang , " AI-Powered Innovations in Ophthalmic Diagnosis and Treatment ", Bentham Science Publishers (2025). https://doi.org/10.2174/9798988122871250101

[33] Tian Teng and Xiao Zhang, Design Principles and Feasibility Assessment of an Intelligent Spare-Parts Shelving System for Shipboard Storage Environments, Journal of Comprehensive Engineering，2026, 15(1), 24-30

[34] Su, J., Cai, C., Zhu, F., He, C., Xu, X., Guan, D., & Si, C. (2024, September). Momentum auxiliary network for supervised local learning. In European Conference on Computer Vision (pp. 276-292). Cham: Springer Nature Switzerland.

[35] Su, J., Zhu, F., Shi, H., Han, T., Qiu, Y., Luo, J., ... & Gao, J. (2025). MAN++: Scaling Momentum Auxiliary Network for Supervised Local Learning in Vision Tasks. arXiv preprint arXiv:2507.16279.

[36] Quan, Q., Gao, Y., & Wang, Q. (2025). The impact of teacher emotional support on students' engagement in AI-mediated English learning environments: The mediating role of resilience and self-efficacy. Acta Psychologica, 260, 105766.

[37] Xu, Z., Zhang, X., Li, R., Tang, Z., Huang, Q., & Zhang, J. (2024). Fakeshield: Explainable image forgery detection and localization via multi-modal large language models. arXiv preprint arXiv:2410.02761.

[38] Xu, Z., Zhang, X., Zhou, X., & Zhang, J. (2025). AvatarShield: Visual Reinforcement Learning for Human-Centric Video Forgery Detection. arXiv preprint arXiv:2505.15173.

[39] Huang, Q., Xu, Z., Zhang, X., & Zhang, J. (2025). UniShield: An Adaptive Multi-Agent Framework for Unified Forgery Image Detection and Localization. arXiv preprint

arXiv:2510.03161.

[40] Cui, J., Zhang, G., Chen, Z., & Yu, N. (2022). Multi-homed abnormal behavior detection algorithm based on fuzzy particle swarm cluster in user and entity behavior analytics. Scientific Reports, 12(1), 22349.

[41] Liu, T., He, J., Zhou, R., Song, J., Liu, H., & Liang, Y. (2025). Revolutionizing clinical decision making through deep learning and topic modeling for pathway optimization. Scientific Reports, 15(1), Article 28787. https://doi.org/10.1038/s41598-025-12679-z

[42] Wu, X., Xing, X., Yao, J., Qian, Q., & Song, J. (2025). Scnet: spectral convolutional networks for multivariate time series classification. Applied Intelligence, 55(6), Article 456. https://doi.org/10.1007/s10489-025-06352-1

[43] Ma, C., Song, J., Xu, Y., Fan, H., Wu, X., & Sun, T. (2024). Vehicle-Based Machine Vision Approaches in Intelligent Connected System. IEEE Transactions on Intelligent Transportation Systems, 25(3), 2827-2836. https://doi.org/10.1109/TITS.2023.3276325

[44] Song, J., Wu, X., Yao, J., Zhang, Q., Shang, C., Qian, Q., & Song, J. (2026). SPC: Self-supervised point cloud completion. Neural Networks, 194, Article 108107. Advance online publication. https://doi.org/10.1016/j.neunet.2025.108107

[45] Yu, C., Li, P., Wu, H., Wen, Y., Deng, B., & Xiong, H. (2024). USM: Unbiased Survey Modeling for Limiting Negative User Experiences in Recommendation Systems. arXiv preprint arXiv:2412.10674.