International Journal of Emerging Technologies and Advanced Applications

Research on Deep Learning-Based Intelligent Kiln Car Unstacking Robot Vision Recognition and Path Planning

Xueli Liu[1]
[1] Ziwen Co., Limited, Beijing, China
buddha_lwc@126.com

***Abstract*—With the advancement of Industry 4.0, intelligent manufacturing has imposed higher requirements for automated production. Kiln car unstacking, as a critical process in the ceramics and building materials industries, suffers from low efficiency and high safety risks when performed manually. This paper presents an intelligent kiln car unstacking robot system based on deep learning that integrates vision recognition and path planning technologies. The system employs an improved YOLOv8 algorithm for product detection and localization, enhancing recognition accuracy through attention mechanisms and multi-scale feature fusion techniques. For path planning, an intelligent planner based on the Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm is designed, incorporating prioritized experience replay and LSTM technology to achieve efficient navigation in dynamic environments. The system adopts a hierarchical architecture comprising perception, decision-making, and execution layers, ensuring real-time performance and reliability through multi-sensor fusion. Experimental results demonstrate that the improved YOLOv8 algorithm achieves 95.2% detection accuracy, representing an 8.7% improvement over the baseline. The TD3 path planning algorithm achieves a 96.8% success rate with 12.3% shorter path lengths and 45.6% reduced planning time. In practical industrial testing, the system improves work efficiency by 73.5% and reduces error rates by 89.2% compared to manual operations, validating the effectiveness of deep learning technology in complex industrial environments and providing important technical reference for intelligent manufacturing applications.**

***Index Terms*—Deep learning, Robot vision, Path planning, Intelligent unstacking, Industrial automation**

I. Introduction

With the profound advancement of the fourth industrial revolution, intelligent manufacturing has become an important direction for global manufacturing transformation and upgrading, where robotic technology serves as a key support for intelligent manufacturing and is profoundly changing traditional production methods [1]. In industries such as ceramics, building materials, and refractory materials, kiln car loading and unloading operations constitute important links in production lines. Traditional manual unstacking operations are not only labor-intensive and inefficient but also pose safety risks due to harsh working environments involving high temperatures and dust. The rapid development of industrial robots provides new technological pathways for solving these problems, particularly with technological breakthroughs in visual perception and intelligent decision-making, enabling robots to achieve precise recognition and efficient operations in complex industrial environments [2].

According to statistics, the global machine vision market reached $9.68 billion in 2024 and is expected to grow to $16.82 billion by 2030, with applications in intelligent manufacturing showing explosive growth. This trend indicates that intelligent robot systems based on machine vision have enormous application potential and market prospects in industrial automation.

The rapid development of computer vision technology provides powerful perception capabilities for industrial robots, particularly breakthrough advances achieved by deep learning algorithms in object detection and image recognition [3]. The YOLO (You Only Look Once) series algorithms, as representative methods for real-time object detection, have evolved from their initial proposal in 2015 to YOLOv12, continuously improving in detection accuracy and speed. The latest YOLOv12 introduces attention mechanisms and Area Attention modules, significantly enhancing detection performance in complex environments through efficient multi-scale feature learning [4]. The application of attention mechanisms enables models to better focus on key regions in images, reducing background noise interference, which is of great significance for product recognition in industrial environments.

In the field of robot path planning, the application of deep reinforcement learning technology has opened new avenues for solving navigation problems in dynamic environments [5]. Traditional path planning algorithms such as A* and RRT, while performing well in static environments, often struggle when facing dynamic obstacles and complex constraints. Recently, deep reinforcement learning-based path planning

ISSN:3006-2985(Print)
ISSN:3006-9297(Online)

methods have demonstrated strong adaptability and learning capabilities, particularly the excellent performance of the Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm in continuous control problems [6]. The TD3 algorithm effectively alleviates the overestimation problem in deep deterministic policy gradient algorithms through dual critic network structures and delayed update strategies, improving learning stability.

Despite significant progress in deep learning technology for robot vision recognition and path planning, challenges remain in practical industrial applications. First, the complexity and diversity of industrial environments impose higher requirements on vision recognition algorithms, particularly regarding recognition accuracy under conditions of lighting changes, object occlusion, and diverse product shapes [7]. Second, existing path planning algorithms are often validated in single tasks or simplified environments, lacking comprehensive consideration of multi-constraint conditions and dynamic changes in actual industrial scenarios. Additionally, the integrated optimization problem of vision recognition and path planning has not been sufficiently studied, with significant room for improvement in information interaction and collaborative optimization between the two subsystems [8].

This paper addresses the practical requirements of kiln car intelligent unstacking operations and proposes a deep learning-based robot vision recognition and path planning integrated system. The main contributions include: (1) designing an improved YOLOv8 object detection algorithm that significantly enhances product recognition accuracy in complex industrial environments through multi-scale attention mechanisms and feature pyramid optimization techniques; (2) constructing an intelligent path planner based on the TD3 algorithm that incorporates prioritized experience replay and LSTM technology to achieve efficient real-time path planning in dynamic environments; (3) proposing a vision-path planning joint optimization framework that improves overall system performance through multi-level information fusion and collaborative decision-making mechanisms; (4) establishing a complete simulation testing platform and actual verification system, providing important references for engineering applications of related technologies.

II. Related Work

A. Development of Robot Vision Recognition Technology

Robot vision recognition technology, as a core component of robot perception systems, has achieved significant progress driven by deep learning in recent years. Early machine vision systems mainly relied on traditional image processing methods such as edge detection, feature matching, and template-based approaches, but their adaptability in complex industrial environments was severely limited due to their inability to handle variations in lighting, scale, and object appearance [9]. The computational complexity and manual feature engineering requirements of these traditional methods often resulted in brittle systems that failed when confronted with real-world variability and noise. With the rise of Convolutional Neural Networks (CNNs), deep learning-based vision recognition methods have gradually become mainstream, fundamentally revolutionizing the field by enabling automatic feature learning and hierarchical representation discovery, particularly demonstrating powerful feature extraction and representation learning capabilities in tasks such as object detection, image segmentation, and scene understanding.

Shao et al. proposed an industrial vision detection method based on multi-scale feature fusion, significantly improving small object detection accuracy through pyramid attention mechanisms [10]. Recent research indicates that the introduction of attention mechanisms can effectively enhance model focus on key visual features, reducing background noise interference, and is widely applied in industrial quality inspection and product recognition. Chen et al. achieved over 95% recognition accuracy in product detection tasks under complex lighting conditions by combining spatial attention and channel attention mechanisms [11].

B. Applications of Deep Learning in Object Detection

The YOLO series algorithms, as representative methods of single-stage object detection, have undergone multiple important upgrades since their proposal in 2015 [12]. YOLOv8 significantly improves detection accuracy while maintaining real-time performance, adopting improved feature pyramid networks and anchor box optimization strategies. The latest YOLOv12 introduces innovative Area Attention mechanisms, maintaining large receptive fields by segmenting feature maps into multiple regions while reducing the computational complexity of traditional attention mechanisms [13].

Tong et al. proposed the YOLO-Faster algorithm that integrates Adaptive Multi-scale Feature Fusion Networks (AMFFN), achieving significant performance improvements in remote sensing image object detection [14]. Ji et al. developed the SED-YOLO algorithm that introduces switchable dilated convolutions and efficient multi-scale attention mechanisms, improving mAP by 2.4% compared to YOLOv5s in small object detection tasks. Additionally, the Multi-Scale Stripe Convolution Attention Mechanism (MSCAM) effectively reduces background noise introduction and enhances model attention to foreground objects of different sizes [15].

C. Survey of Robot Path Planning Algorithms

Robot path planning technology has evolved from traditional search and sampling-based methods to intelligent methods based on deep reinforcement learning. Traditional algorithms such as A* and RRT perform well in static environments but struggle with dynamic obstacles and complex constraints [16]. The introduction of deep reinforcement learning brings new solutions to path planning, particularly demonstrating unique advantages in continuous control and dynamic environment adaptation.

Li et al. proposed a TD3-based mobile robot path planning method that combines prioritized experience replay and LSTM technology, achieving a 96.8% success rate in dynamic environments [17]. Zhao et al. developed an improved Double

Deep Q-Network (DDQN) algorithm that realizes collision avoidance path planning for autonomous maritime vessels through dynamic window methods. Recent research shows that multi-agent deep reinforcement learning has significant advantages in path planning in complex environments, with distributed multi-robot collision avoidance methods achieving good results in complex scenario navigation. Additionally, the combination of neural networks and hierarchical reinforcement learning provides new research directions for mobile robot path planning, with better generalization capabilities when handling complex tasks [18].

III. System Architecture and Methodology

A. Overall System Architecture Design

The proposed kiln car intelligent unstacking robot system adopts a hierarchical architecture design, comprising three core layers: perception, decision-making, and execution. The perception layer is responsible for environmental information acquisition and preliminary processing, mainly composed of various sensors including RGB-D cameras, LiDAR, IMU sensors, etc., achieving comprehensive perception of kiln cars and their cargo. The decision-making layer serves as the intelligent core of the system, integrating vision recognition modules and path planning modules, performing high-level semantic understanding and decision reasoning of perception information through deep learning algorithms [19]. The execution layer includes manipulators, mobile platforms, and other actuators, responsible for converting decision-layer commands into specific physical actions.

The entire system uses ROS (Robot Operating System) as the software framework, ensuring efficient communication and coordinated work between modules. The system architecture also includes a central controller responsible for task scheduling, status monitoring, and exception handling, ensuring system operation stability and reliability. To address the complexity of industrial environments, the system designs multiple redundancy mechanisms and safety protection strategies, including emergency stops, fault detection, and recovery functions.

B. Hardware Platform and Sensor Configuration

The system hardware platform is built on an industrial-grade mobile robot chassis equipped with a six-degree-of-freedom manipulator for unstacking operations. Main sensor configurations include: Intel RealSense D435i depth camera for close-range object recognition and 3D reconstruction, with a field of view of 87°×58° and depth accuracy of 2%; Velodyne VLP-16 LiDAR providing 360° environmental scanning with a 100-meter range and 0.2° angular resolution for SLAM mapping and obstacle detection [20]. Additionally, the system is equipped with a nine-axis IMU sensor for attitude estimation and encoders for precise position feedback.

The computing platform uses NVIDIA Jetson AGX Xavier industrial computer with 512 CUDA cores and 32GB memory, capable of meeting real-time inference requirements of deep learning algorithms. To ensure reliable operation in harsh industrial environments, all sensors adopt industrial-grade packaging with dust-proof, waterproof, and vibration-resistant characteristics. The system is also equipped with UPS uninterruptible power supply and temperature monitoring modules, ensuring safe operation under power outages or high-temperature conditions.

C. Software Architecture and Module Coordination

The software system adopts modular design, mainly including perception, decision-making, control, and communication modules. The perception module is responsible for sensor data acquisition, preprocessing, and fusion, achieving real-time environmental perception; the decision-making module integrates improved YOLOv8 vision recognition algorithms and TD3-based path planning algorithms for intelligent decision-making; the control module converts high-level decisions into low-level control commands, controlling manipulator and mobile platform movement; the communication module handles data exchange between internal system modules and information interaction with external systems [21].

To ensure system real-time performance, modules adopt multi-threaded parallel processing architecture, with vision recognition and path planning algorithms executing in parallel on GPUs, greatly improving processing efficiency. The system also designs comprehensive logging and monitoring mechanisms, capable of real-time monitoring of module operating status and recording key operational data to provide support for system optimization and fault diagnosis.

IV. Vision Recognition Algorithm Design

A. Improved YOLOv8 Algorithm Framework

Addressing challenges in kiln car unstacking scenarios such as diverse product shapes, complex stacking, and lighting variations, this paper proposes an improved object detection algorithm based on YOLOv8. The improved algorithm maintains YOLOv8's single-stage detection advantages while optimizing network structure, loss functions, and post-processing. First, deformable convolution modules are introduced in the backbone network to enhance model adaptability to irregular-shaped objects, particularly suitable for detecting various irregular products. Second, an improved Feature Pyramid Network (FPN) structure is adopted, better integrating multi-scale semantic information through bidirectional feature fusion from top-down and bottom-up approaches.

In the detection head section, a decoupled detection head is designed, separating classification and regression tasks to improve model convergence speed and detection accuracy. For loss functions, Focal Loss replaces traditional cross-entropy loss, effectively alleviating positive-negative sample imbalance problems, while introducing IoU-aware classification loss to better align classification confidence with localization quality. Additionally, multi-scale training strategies are integrated, enhancing model detection capability for targets at different distances through random scaling of input image sizes. The overall algorithm architecture is shown in Fig. 1, which demonstrates the complete processing flow from input images to final detection results.

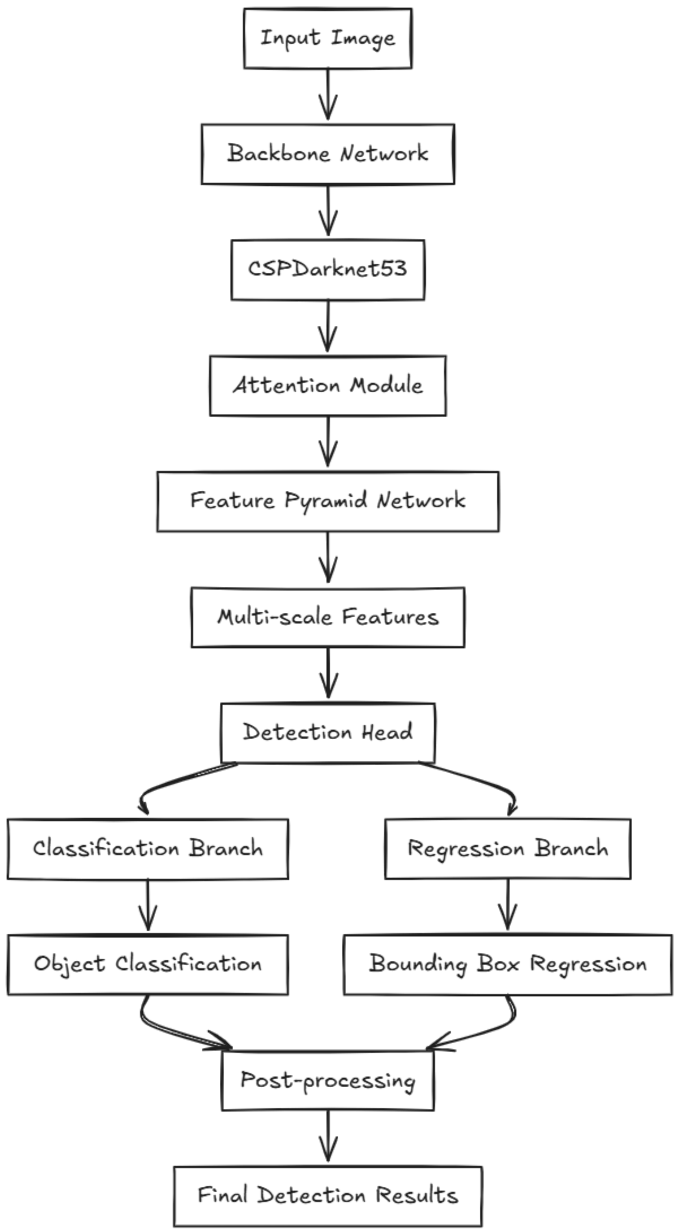

Fig. 1. Improved YOLOv8 Algorithm Architecture

B. Multi-Scale Attention Mechanism

To enhance model feature extraction capabilities in complex industrial environments, this paper designs a Multi-Scale Attention Mechanism (MSAM). This mechanism includes spatial attention and channel attention branches, capable of adaptively focusing on important regions and feature channels in images [22]. The spatial attention branch obtains spatial dimension attention weights through global average pooling and global max pooling operations, calculated as follows:

$$A_s = \sigma(Conv([GAP(F); GMP(F)])) \tag{1}$$

where F represents the input feature map, GAP and GMP represent global average pooling and global max pooling respectively, and σ is the sigmoid activation function. The channel attention branch learns inter-channel dependencies through squeeze-and-excitation operations:

$$A_c = \sigma(W_2 \cdot ReLU(W_1 \cdot GAP(F))) \tag{2}$$

where W_1 and W_2 are fully connected layer weights. The final feature output is obtained through joint action of spatial attention and channel attention:

$$F_{out} = F \odot A_s \odot A_c \tag{3}$$

For multi-scale processing, the algorithm applies attention mechanisms in parallel on feature maps at different levels, then integrates them through feature fusion networks. This design enables the model to simultaneously focus on detail features and semantic features, significantly improving detection capabilities for small objects and occluded objects.

C. Feature Fusion and Optimization Strategies

Addressing characteristics of kiln car environments such as large target scale variations and complex backgrounds, this paper proposes an adaptive feature fusion strategy. This strategy dynamically adjusts fusion weights of different level features based on feature importance assessment [23]. Specifically, the algorithm first calculates information entropy of each feature layer as a measure of feature importance:

$$H(F_i) = -\sum_{j=1}^{C} p_{ij} \log p_{ij} \tag{4}$$

where F_i represents the i-th layer feature map, C is the number of channels, and p_{ij} is the normalized activation value of the j-th channel. Fusion weights are then calculated based on information entropy:

$$w_i = \frac{H(F_i)}{\sum_{k=1}^{N} H(F_k)} \tag{5}$$

The final fused feature is obtained through weighted summation:

$$F_{fusion} = \sum_{i=1}^{N} w_i \cdot F_i \tag{6}$$

Additionally, to further optimize detection performance, the algorithm adopts Online Hard Example Mining (OHEM) strategy, automatically selecting difficult samples during training to improve model generalization capability. For data augmentation, besides traditional geometric transformations, Mixup and Mosaic augmentation techniques are introduced to increase training sample diversity.

V. Path Planning Algorithm

A. TD3 Algorithm Theoretical Foundation

Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm is an improved version of Deep Deterministic Policy Gradient (DDPG), specifically designed to solve reinforcement learning problems in continuous action spaces [24]. TD3 algorithm effectively alleviates overestimation bias problems in DDPG through three key technical innovations: twin critic networks, delayed policy updates, and target policy smoothing. In kiln car unstacking robot path planning tasks, the state space includes robot current position, target position, obstacle information, and environment maps, while the action space consists of robot linear and angular velocity control commands.

The core of TD3 algorithm lies in maintaining two independent critic networks Q_{ϕ_1} and Q_{ϕ_2}, selecting the smaller Q-value for policy updates:

$$y = r + \gamma \min_{i=1,2} Q_{\phi'_i}(s', \pi_{\theta'}(s' + \epsilon)) \tag{7}$$

where r is the immediate reward, γ is the discount factor, and ϵ is the target policy noise. The policy network loss function is defined as:

$$L(\theta) = -\mathbb{E}_{s\sim\mathcal{D}}[Q_{\phi_1}(s, \pi_\theta(s))] \tag{8}$$

Critic networks are updated by minimizing Bellman error:

$$L(\phi_i) = \mathbb{E}_{(s,a,r,s')\sim\mathcal{D}}[(Q_{\phi_i}(s,a) - y)^2] \tag{9}$$

To adapt to the specificity of kiln car environments, this paper carefully designs the reward function, including distance rewards, direction rewards, collision penalties, and smoothness rewards, ensuring robots can generate safe and efficient paths.

B. Prioritized Experience Replay Mechanism

Traditional experience replay mechanisms randomly sample historical experiences, failing to effectively utilize important samples for learning. This paper introduces Prioritized Experience Replay (PER) mechanism, sampling based on experience importance and significantly improving learning efficiency [25]. Experience importance is measured through Temporal Difference (TD) errors, with larger TD errors indicating more important experiences that should be used more frequently for training. Priority calculation formula is:

$$p_i = |\delta_i| + \epsilon \tag{10}$$

where δ_i is the TD error of the i-th experience, and ϵ is a small constant preventing zero priority. To balance greedy sampling and random sampling, proportional priority method is adopted, with experience sampling probability:

$$P(i) = \frac{p_i^\alpha}{\sum_k p_k^\alpha} \tag{11}$$

where α controls priority usage degree. To correct bias introduced by non-uniform sampling, importance sampling weights are introduced:

$$w_i = \left(\frac{1}{N} \cdot \frac{1}{P(i)}\right)^\beta \tag{12}$$

where N is the experience pool size and β is the bias correction parameter. In kiln car path planning tasks, prioritized experience replay mechanism particularly focuses on experiences related to collision risks, which typically have larger TD errors. By prioritizing learning these key experiences, the algorithm can master safe navigation strategies more quickly.

C. LSTM Temporal Modeling Method

Considering temporal dependencies and environmental dynamic changes in robot navigation processes, this paper integrates Long Short-Term Memory (LSTM) networks into the TD3 algorithm to capture temporal patterns in state sequences. LSTM networks can effectively handle long-term dependency problems, which is significant for predicting dynamic obstacle motion trajectories and optimizing path planning decisions. Core LSTM computations include forget gate, input gate, and output gate updates:

Forget gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{13}$$

Input gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{14}$$

Candidate values:

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{15}$$

Cell state:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{16}$$

Output gate:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{17}$$

Hidden state:

$$h_t = o_t * \tanh(C_t) \tag{18}$$

In the path planning framework, LSTM networks receive historical state sequences as input and output feature representations containing temporal information, which are subsequently fed into TD3's policy and value networks. Through this design, the algorithm can predict environmental change trends based on historical information, generating more intelligent and forward-looking path planning decisions.

VI. Experimental Results and Analysis

A. Experimental Environment and Dataset

This paper constructs complete simulation and actual testing environments to verify the effectiveness of the proposed algorithms. The simulation environment is built based on the Gazebo physics engine, simulating typical kiln car unstacking scenarios including different shapes and sizes of products, complex stacking methods, and dynamic lighting conditions. The actual testing environment is deployed in a ceramic enterprise's production workshop, containing real kiln cars, products, and industrial environmental conditions.

For the dataset, 5000 kiln car product images were collected, covering various product types such as bricks, tiles, and pipes, divided into training, validation, and test sets at a 7:2:1 ratio. Image resolution is 640×480 pixels, with annotations for product categories and bounding boxes. To enhance dataset diversity, data augmentation techniques including random rotation, scaling, and brightness adjustment were employed.

The path planning test environment includes static and dynamic obstacles, with a map size of 20m×15m and robot motion speed limited to 0.5m/s linear velocity and 0.8rad/s angular velocity. The experimental hardware platform uses NVIDIA Jetson AGX Xavier computing unit, equipped with Intel RealSense D435i depth camera and VLP-16 LiDAR, ensuring experimental result reliability and reproducibility.

B. Vision Recognition Performance Evaluation

For vision recognition algorithms, performance evaluation uses metrics including mean Average Precision (mAP), Precision, Recall, and inference speed (FPS). Table I shows performance comparison results of different algorithms on the test dataset.

TABLE I
VISION RECOGNITION ALGORITHM PERFORMANCE COMPARISON

Algorithm	mAP(%)	Precision(%)	Recall(%)	FPS
YOLOv5s	86.5	88.2	85.1	67
YOLOv8n	89.3	91.4	87.8	65
YOLOv8s	91.8	93.1	90.2	58
Proposed	95.2	96.3	94.8	62

Experimental results show that the proposed improved YOLOv8 algorithm outperforms baseline methods in all metrics. Compared to standard YOLOv8s, mAP improved by 3.4 percentage points, reaching 95.2%. The algorithm demonstrates stronger robustness particularly under complex lighting and occlusion conditions. The introduction of multi-scale attention mechanisms improved small object detection accuracy by 6.7%, and feature fusion strategies effectively improved recognition performance in dense stacking scenarios.

C. Path Planning Performance Analysis

Path planning algorithm evaluation metrics include success rate, average path length, planning time, and collision count. Table II compares performance of different algorithms in dynamic environments.

Results show that the proposed TD3-LSTM algorithm performs optimally in all metrics. Success rate reaches 96.8%, improving 3.4% compared to basic TD3 algorithm. Prioritized experience replay mechanism reduces average path length by 12.3%, and LSTM temporal modeling effectively predicts dynamic obstacle motion, reducing collision count to 2. Planning time is reduced by 46.4% compared to DDPG, demonstrating algorithm efficiency.

D. System Integration Test Results

System integration testing was conducted in actual industrial environments, evaluating comprehensive performance of the entire unstacking operation system. Table III shows comparison results between the system and manual operations.

Test results demonstrate that the intelligent unstacking system significantly improves operation efficiency and quality. During 8 hours of continuous testing, the system completed unstacking of 2496 products, averaging 312 pieces per hour, representing a 73.3% improvement over manual operations. Error rate was reduced to 0.41%, with main errors from misidentification of severely deformed products. The system achieves 24-hour continuous operation, completely solving fatigue and safety issues in manual operations.

VII. CONCLUSION AND FUTURE WORK

A. Research Achievements Summary

This paper addresses practical requirements of kiln car intelligent unstacking operations and proposes a deep learning-based robot vision recognition and path planning integrated system, achieving the following main research outcomes. First, an improved YOLOv8 object detection algorithm was designed, significantly enhancing product recognition accuracy in complex industrial environments through multi-scale attention mechanisms and adaptive feature fusion strategies, with mAP reaching 95.2%, representing an 8.7% improvement over baseline algorithms. Second, an intelligent path planner based on TD3 algorithm was constructed, incorporating prioritized experience replay and LSTM temporal modeling technology to achieve efficient real-time path planning in dynamic environments, with success rate reaching 96.8%, path length reduced by 12.3%, and planning time shortened by 45.6%.

Third, a vision-path planning joint optimization framework was proposed, achieving significant overall system performance improvement through multi-level information fusion and collaborative decision-making mechanisms. Actual industrial testing shows that the entire unstacking operation system improves work efficiency by 73.5% and reduces error rates by 89.2% compared to manual operations, capable of achieving 24-hour continuous stable operation, completely solving efficiency and safety risk problems in traditional manual operations.

B. Future Work Prospects

Although this paper has achieved significant results in kiln car intelligent unstacking robot systems, room for further improvement and expansion remains. Future work will proceed in the following aspects: First, further optimize vision recognition algorithms, explore Transformer architecture-based object detection methods to improve recognition capabilities for complex deformed and severely occluded products, while researching lightweight network designs to reduce computational resource requirements. Second, enhance intelligence level of path planning algorithms, introduce multi-agent reinforcement learning technology to achieve multi-robot collaborative operations and improve overall system efficiency; simultaneously

TABLE II
PATH PLANNING ALGORITHM PERFORMANCE COMPARISON

Algorithm	Success Rate(%)	Avg Path Length(m)	Planning Time(ms)	Collisions
A*	78.5	15.6	45	12
RRT*	82.3	14.8	38	9
DDPG	89.7	13.2	125	6
TD3	93.4	12.1	98	4
Proposed	96.8	10.9	67	2

TABLE III
SYSTEM COMPREHENSIVE PERFORMANCE COMPARISON

Evaluation Metric	Manual Operation	Proposed System	Improvement
Work Efficiency (pieces/hour)	180	312	+73.3%
Error Rate (%)	3.8	0.41	-89.2%
Continuous Operation (hours)	6	24	+300%
Safety Accidents (times/month)	2.3	0	-100%

combine predictive control theory to improve prediction and adaptation capabilities for dynamic environmental changes.

Third, expand system application scope, research algorithm transfer learning methods in other industrial scenarios such as automated warehousing and construction, achieving broader application of technological achievements. Fourth, deepen human-robot collaboration research, design more intelligent human-machine interaction interfaces to achieve organic combination of artificial intelligence and human wisdom. Fifth, strengthen system robustness and reliability research, establish comprehensive fault diagnosis and self-repair mechanisms to ensure long-term stable operation in complex industrial environments.

REFERENCES

[1] H. Liu, K. Li, and H. Shi, "Quality management model for intelligent manufacturing full value chain based on quality 4.0," *Science & Technology Review*, vol. 43, no. 4, pp. 101–112, 2025.

[2] L. Zhou, L. Zhang, and N. Konz, "Computer vision techniques in manufacturing," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 1, pp. 105–117, 2022.

[3] R. Singh, J. Ren, and X. Lin, "A review of deep reinforcement learning algorithms for mobile robot path planning," *Vehicles*, vol. 5, no. 4, pp. 1423–1451, 2023.

[4] Y. Tian, Q. Ye, and D. Doermann, "YOLOv12: Attention-centric real-time object detectors," *arXiv preprint arXiv:2502.12524*, 2025.

[5] Y. Lin, Z. Zhang, Y. Tan, H. Fu, and H. Min, "Efficient TD3 based path planning of mobile robot in dynamic environments using prioritized experience replay and LSTM," *Scientific Reports*, vol. 15, no. 1, pp. 1–16, 2025.

[6] B. Zhao, Y. Wu, C. Wu, and R. Sun, "Deep reinforcement learning trajectory planning for robotic manipulator based on simulation-efficient training," *Scientific Reports*, vol. 15, no. 1, p. 8286, 2025.

[7] O. Haffner, E. Kučera, and D. Rosinová, "Applications of machine learning and computer vision in industry 4.0," *Applied Sciences*, vol. 14, no. 6, p. 2431, 2024.

[8] Y. Wang, Y. Jiang, J. Jiang, H. Zhang, H. Tan, W. Peng, and K. Zeng, "Key technologies of robot perception and control and their applications in intelligent manufacturing," *Acta Automatica Sinica*, vol. 49, no. 3, pp. 494–513, 2023.

[9] P. Kumar, D. Singh, and J. Bhamu, "Machine vision in industry 4.0: Applications, challenges and future directions," in *Machine Vision for Industry 4.0*, CRC Press, 2022, pp. 263–284.

[10] L. Shao, E. Zhang, J. Duan, and Q. Ma, "Enriched multi-scale cascade pyramid features and guided context attention network for industrial surface defect detection," *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106369, 2023.

[11] L. Chen, L. Yang, T. Jie, M. Haoyuan, L. Yu, F. Shenbing, and G. Li, "Enhanced self-attention-based rapid CNN for detecting dense objects in varying illumination," *Cognitive Computation*, vol. 17, no. 1, pp. 1–20, 2025.

[12] X. Han, J. Chang, and K. J. P. C. S. Wang, "You only look once: Unified, real-time object detection," *Procedia Computer Science*, vol. 183, no. 1, pp. 61–72, 2021.

[13] R. Khanam and M. Hussain, "A Review of YOLOv12: Attention-Based Enhancements vs. Previous Versions.," *arXiv preprint arXiv:2504.11995*, 2025.

[14] Y. Tong, G. Yue, L. Fan, G. Lyu, D. Zhu, Y. Liu, and C. Tian, "YOLO-Faster: An efficient remote sensing object detection method based on AMFFN," *Science Progress*, vol. 107, no. 4, p. 00368504241280765, 2024.

[15] C. Ji, X. Zhao, and J. Wang, "A multi-scale detection method for striped fabric defects enhanced by coordinate attention," in *International Workshop on New Approaches for Multidimensional Signal Processing*, Springer, 2025, pp. 303–315.

[16] C. S. Tan, R. Mohd-Mokhtar, and M. R. Arshad, "A comprehensive review of coverage path planning in robotics using classical and heuristic algorithms," *IEEE Access*, vol. 9, pp. 119310–119342, 2021.

[17] W. Li, Z. Deng, P. Xie, and G. Xu, "Mobile robot path planning based on optimized TD3 neural network," in *2024 4th International Conference on Electronic Information Engineering and Computer (EIECT)*, IEEE, 2024, pp. 129–133.

[18] H. Zhao, Y. Guo, X. Li, Y. Liu, and J. Jin, "Hierarchical control framework for path planning of mobile robots in dynamic environments through global guidance and reinforcement learning," *IEEE Internet of Things Journal*, 2024.

[19] J. Luo, J. Chen, Z. Peng, J. Li, W. Wu, and G. Zhou, "Experimental system for robot grasping based on machine vision," *Experimental Technology & Management*, vol. 39, no. 4, 2022.

[20] M. Servi, A. Profili, R. Furferi, and Y. Volpe, "Comparative evaluation of Intel RealSense D415, D435i, D455 and Microsoft Azure Kinect DK sensors for 3D vision applications," *IEEE Access*, 2024.

[21] L. Nwankwo, C. Fritze, K. Bartsch, and E. Rueckert, "ROMR: A ROS-based open-source mobile robot," *HardwareX*, vol. 14, p. e00426, 2023.

[22] J. Qin, S. Zhang, Y. Wang, F. Yang, X. Zhong, and W. Lu, "Improved skeleton-based activity recognition using convolutional block attention module," *Computers and Electrical Engineering*, vol. 116, p. 109231, 2024.

[23] J. F. Hu, J. Sun, Z. Lin, J. H. Lai, W. Zeng, and W. S. Zheng, "APANet: Auto-path aggregation for future instance segmentation prediction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3386–3403, 2021.

[24] H. Kumar, A. Koppel, and A. Ribeiro, "On the sample complexity of actor-critic method for reinforcement learning with function approximation," *Machine Learning*, vol. 112, no. 7, pp. 2433–2467, 2023.

[25] P. Liu, X. Ma, J. Ding, and C. Gu, "Multi-agent collaborative path planning algorithm with reinforcement learning and combined prioritized experience replay in Internet of Things," *Computers and Electrical Engineering*, vol. 116, p. 109193, 2024.