A Survey on Deep Learning-Based Integrated Perception-Cognition-Control Systems for Autonomous Mobile Robots

Wenqi Han

DIPLOMA Hochschule-Private Fachhochschule Nordhessen, An Hegeberg 2, 37242 Bad-Sooden-Allendorf, Germany
vickyhan129@outlook.com

Abstract: The mobile robots currently widely adopted in industrial and logistics sectors integrate various technologies, with deep learning emerging as a new driving force and breakthrough for enhancing robotic efficiency and precision. Integrating its technology into each component module of the robotic system's mobility drives interaction across the three dimensions of perception, cognition, and control. This paper divides the robotic system architecture into two major processes: “input” and “output,” with the cognitive layer serving as the hub, corresponding to “perception-cognition” and “cognition-control,” respectively. Within this framework, we outline the current state of module integration in mobile robots, adaptive solutions incorporating deep learning, deployment platforms, and relevant examples. The current bottlenecks in multi-sensor fusion under deep learning are discussed, and future development directions in this field are explored. This research aims to provide a theoretical framework and practical approach for the integration of deep learning with mobile robotic systems, thereby enhancing the autonomy and reliability of intelligent robots in complex dynamic environments.

Keywords: Mobile Robot, Deep Learning, Motion Control, Path lanning.

1 Introduction

Driven by Industry 4.0 and Made in China 2025, global industrial and logistics sectors are advancing toward intelligent transformation. Robots powered by artificial intelligence and deep learning are being widely deployed in smart manufacturing, intelligent logistics, security, and other fields, demonstrating immense value. Among these, Autonomous Mobile Robots (AMR) and Automated Guided Vehicles (AGV) are increasingly becoming pivotal forces propelling industrial development.

Traditional robot designs often involve fragmented development of subsystems, lacking coordination and information sharing, which limits their intelligent evolution. In complex scenarios, their environmental adaptability and task transfer capabilities significantly decline, becoming key bottlenecks for overall performance improvement. In contrast, deep learning excels in feature extraction and modeling. It supports single

K. Subramanian (ed.), *Proceedings of the International Workshop on Advances in Deep Learning for Image Analysis and Computer Vision (IWADIC 2025)*, Advances in Computer Science Research 128,
https://doi.org/10.2991/978-94-6239-648-7_20

tasks like object detection, trajectory prediction, and control signal generation, while also integrating multiple modules into a unified network through end-to-end training to achieve system-level intelligence. This "cross-module integration" paradigm offers a viable solution for enhancing the intelligence of autonomous mobile robots.

This paper systematically reviews cutting-edge advances in deep learning for cross-module integration in mobile robots, analyzes key challenges and practical applications in control system deployment and optimization, and aims to provide comprehensive and valuable theoretical references and methodological insights for constructing highly adaptive, intelligent robotic systems.

2 Overview

2.1 Common Logistics Mobile Robots

Logistics mobile robots serve as critical equipment across diverse scenarios. Equipped with onboard perception, decision-making, and motion systems, they achieve autonomous navigation, task execution, and path planning in complex environments, thereby enhancing material handling and operational efficiency. Current mainstream technological approaches in smart manufacturing, warehousing, and logistics primarily fall into two categories: AGVs and AMRs.

AGVs, as the first wheeled mobile robots implemented in industrial production, primarily handle logistics transportation in factories and warehouses. Due to their simple principles, low cost, and ease of deployment, AGVs saw widespread adoption during the early stages of industrial automation. However, constrained by external guidance systems, they lack flexibility and adaptability. Their market share has gradually been replaced by more intelligent AMRs. Today, AGVs are predominantly used in fixed-route or highly repetitive workshop logistics scenarios to leverage their cost and reliability advantages.

Compared to AGVs, AMRs demonstrate more pronounced characteristics in environmental adaptability, intelligent judgment, and operational flexibility. They represent a pivotal shift in mobile robotics toward "intelligent navigation" and "cognitive autonomy," possessing environmental awareness, route decision-making, action execution, and interactive capabilities. These robots can autonomously achieve work objectives within complex, obstructed spaces. They integrate cutting-edge advancements from disciplines including computer science, control theory, and artificial intelligence. [1]

2.2 Architecture of Autonomous Mobile Robot Systems

Autonomous mobile robots consist of three major modules, namely perception, cognition, and motion. The perception module collects environmental and self-related information—such as position, obstacles, and targets—through a multi-sensor array, enhancing data reliability through fusion. The cognition module processes this information to perform mapping, path planning, and obstacle avoidance. The motion module then translates the paths or decisions from the cognition layer into motor drive commands.

2.3 Overview of Deep Learning Technologies

Over the past decade, deep learning has emerged as the fastest-growing AI discipline, dominating intelligent perception, image processing, and behavioral prediction. In AMR, its capabilities align with the "perception-understanding- response" framework: CNN-based visual perception enables target recognition, semantic understanding, and obstacle detection, enhancing environmental representation; Strategic cognition based on reinforcement and imitation learning supports autonomous decision-making and path planning in complex environments, overcoming limitations of traditional algorithms; Control execution based on deep neural networks is used for trajectory prediction and adaptive control in nonlinear systems, enabling the construction of end-to-end perception-control frameworks.

Leveraging deep learning hardware acceleration platforms such as NVIDIA Jetson, an increasing number of enterprises are deploying models on physical AMRs for on-site perception and decision-making control. However, this trend also presents greater challenges in engineering aspects like system integration and embedded deployment, necessitating more refined technical and application designs.

3 Perception-Cognition-Control Functional System Architecture

3.1 Perception Layer

The perception layer is responsible for converting external environments and internal bodily states into processable data. Individual standalone sensors have inherent limitations [2]. Most mobile robots employ integrated multi-sensor systems to receive external perception information more directly and holistically. Relevant sensors include LiDAR, visual sensors, and IMUs.

LiDAR serves as the core sensor for SLAM, performing ranging through laser scanning. Typically utilizing infrared light as the emission source, it illuminates targets and receives their reflected signals. By analyzing the relationship between emission and reception, it estimates the relative position of targets and constructs maps based on this positional data. This process generates high-precision two-dimensional or three-dimensional point clouds with accuracy reaching the centimeter level.

Visual sensors are commonly used for obstacle recognition and detection. In recent years, RGB-D (3D imaging) has seen widespread adoption; compared to traditional RGB (2D imaging), it not only outputs color images but also provides a single-channel depth matrix, significantly expanding its application scope. Additionally, visual sensors possess target recognition and scene semantic segmentation capabilities, with performance at the perception layer primarily enhanced by deep learning advancements.

3.2 Cognitive Layer

The cognitive layer is responsible for transforming "data" into "action strategies," positioned between the perception and control layers. Its primary purpose is to utilize environmental data for inference, prediction, and decision-making.
SLAM technology is a primary method for autonomous positioning in robotics. Simultaneous localization and mapping enables robots not only to navigate autonomously but also to avoid obstacles. This approach offers high precision and real-time performance, making it widely applicable across various indoor and outdoor environments.

After the map is created, the robot performs global or local path planning. At this stage, the robot possesses a complete environmental map and does not require real-time updates [3]. Global planning is executed when a complete map is available, commonly employing Dijkstra's algorithm, A* algorithm, or D* algorithm. Local planning is suited for dynamic environments, utilizing real-time sensor information for obstacle avoidance and safe navigation to the target. Common methods include the artificial potential field method and the Time-Elastic Band (TEB) algorithm [4,5].

In recent years, deep reinforcement learning has been progressively applied to the cognitive layer, enabling efficient learning from perception to control strategies. For instance, combining convolutional neural networks with policy gradients allows robots to directly learn path selection strategies using only sensor data, significantly reducing the need for manual rule design.

3.3 Control Layer

The drive layer, as the core execution component of the robot, primarily converts motion commands issued by the understanding layer into actual physical actions. It must also execute precise movements and ensure stable, reliable motion processes to guarantee the dependable operation of the robotic system.

The changes in the robot's body posture and spatial relative position are primarily determined by motion control. Motion control is the core of autonomous navigation and operation for mobile robots, encompassing trajectory tracking, posture/velocity control, and obstacle avoidance. This integrated technology dictates both movement speed and efficiency while also influencing the robot's adaptability to complex environments and operational flexibility.

Trajectory tracking builds upon cognitive-level trajectory planning by employing kinematic control and model analysis during operation to ensure the robot moves precisely along a predefined trajectory/path while achieving accurate posture and position control [6]. A common approach involves fusing IMU and encoder data to correct posture angular errors, thereby maintaining operational stability and enhancing straight-line travel accuracy.

Closed-loop control, as the mainstream method in modern control systems, has been widely applied in AMR. The classic PID (Proportional–Integral–Derivative) controller, a mature closed-loop controller, demonstrates excellent performance across multiple scenarios. Its mechanism involves continuously comparing the output with the desired value and adjusting the control quantity to ensure system stability and precision.

4 Deep Learning-Based Cross-Module System Integration

4.1 Perception-Cognition Collaborative Approach

The essence of perception-cognition fusion lies in establishing communication channels and semantic associations between environmental perception and task decision-making. Its objective is to enable cognitive modules to directly utilize high-dimensional features generated by perception, thereby avoiding information loss caused by low-level representations. This approach enhances response speed in dynamic environments while preserving high-level characteristics to support more accurate decision-making.

Object detection and image semantic segmentation based on deep learning are essentially a fusion of perception and cognition. The former focuses on identifying the types and locations of objects within an image, while the latter precisely distinguishes and extracts the pixel regions of objects from their background. [7] Taking AMR as an example, its navigation does not require predefined paths and often operates in complex dynamic environments, thus typically relying on deep learning for training. Specifically, deep neural networks (such as CNN, RNN, AE) perform multi-layer feature abstraction learning on raw data, with common methods including YOLO, point cloud detection, and others.

Object detection fusion is primarily categorized into three levels: low-level, mid-level, and high-level. Low-level fusion is data-level fusion [8], which combines visual information from 2D images with raw point clouds to perform detection through 3D feature extraction—such as projecting camera images onto LiDAR. Mid-level fusion is feature-level fusion [9], where features are first extracted separately from image and point cloud detectors, then integrated and fed into a 3D detector. High-level fusion is decision-level fusion [9], where image and point cloud detectors operate independently before combining results. This avoids complex cross-modal computations, yielding higher efficiency.

Semantic segmentation differs from object detection or image classification in that it not only identifies the categories of objects within an image but also precisely determines their spatial locations and contours. At its core, it elevates object recognition to the pixel level, explicitly labeling each pixel [10]. Semantic segmentation significantly enhances a mobile robot's ability to recognize its environment. Given an image, it can provide more accurate environmental perception results, enabling the robot to perform complex tasks more effectively.

The aforementioned technology relies on multi-sensor fusion, commonly employing Kalman filters and their variants to fuse multi-source observations in an optimal or suboptimal manner under noisy and uncertain conditions, thereby obtaining an optimal estimate of the robot's state. Concurrently, it requires completing temporal and spatial calibration: achieving cross-sensor alignment through time synchronization and determining their pose in the robot's coordinate system to accomplish spatial calibration.

The primary component of robotic workflow is the perception-cognition interaction. This step involves data collection and preliminary processing, providing essential

support for subsequent decision-making and autonomous driving to enhance overall operational effectiveness and stability.

4.2 Cognition-Control Integration Approach

In traditional architectures, the perception layer handles global and local path planning, while the control layer performs tracking independently. In dynamic environments, this separation often leads to response delays and trajectory discontinuities. Deep learning-based perception-control integration couples high-level planning with low-level control, enabling real-time, smooth, and adaptive motion.

SLAM serves as a critical decision-making component, ensuring robots operate effectively in real-time environments. Multi-sensor fusion SLAM mitigates the biases, accuracy limitations, and map overlay issues inherent in single-laser SLAM [11], while also addressing the sensitivity to lighting conditions and reduced accuracy in snowfall or haze encountered by pure vision-based SLAM [12].

After the map is constructed, the global optimal path from start to end can be calculated based on offline or online maps, generating a continuous, obstacle-free sequence of waypoints as the navigation trajectory. In global path planning, the A* algorithm is most commonly adopted due to its efficient search capabilities, scalability, and broad applicability.

Global path planning typically employs the A* algorithm to evaluate node priorities using "deterministic cost + heuristic estimation." In moderately dynamic scenarios, local planning can perform online trajectory corrections. In practical applications, global planning often first generates a primary path, followed by local planning for obstacle avoidance and refinement. To enhance responsiveness to local disturbances, an artificial potential field (APF) can be integrated into the heuristic. This combines the gravitational potential toward the goal and the repulsive potential away from obstacles as weighted components of the energy potential. Together with the deterministic cost, this forms an evaluation criterion that explicitly incorporates local obstacle avoidance into the global search, thereby improving real-time performance and trajectory smoothness [13,14].

After path generation is complete, various control methods can be employed to ensure the robot travels stably along the planned path. Beyond MPC and traditional PID for path following, deep learning-based end-to-end closed-loop policy learning (Actor–Critic + PPO) can be employed. During trajectory tracking, the policy network (actor) outputs continuous actions, while the value network (critic) reduces variance in policy updates through value estimation to enhance training stability; PPO further constrains update steps through objective function clipping and KL divergence constraints, suppressing excessive updates and policy collapse to enhance robustness. As the most comprehensive form of cross-module integration, this end-to-end architecture abandons traditional layered structures, directly mapping raw sensor signals to actuator commands. Deep neural networks permeate feature representation, task perception, and motion planning, playing a dominant role in the control chain.

5 Deployment and Self-Testing of the Control System

Currently, AMR control systems have evolved from isolated “planning-execution” units into full-lifecycle automation systems integrating perception and cognition. As applications expand, operational environments grow increasingly complex, with external uncertainties such as continuous target movement and changing operating conditions; internally, they are affected by degradation of execution units, cumulative sensor errors, and communication delays. Pre-execution control alone cannot ensure high reliability, necessitating the introduction of online health monitoring and self-optimization mechanisms during operation.

Building upon this foundation, control system research diverges into two distinct approaches. The first focuses on secure and trustworthy deployment, leveraging digital twins to perform real-time verification and optimization of virtual control systems during operation, thereby reducing risks associated with physical debugging. Post-deployment, continuous online monitoring and parameter updates are maintained. The second approach involves Fault Detection and Diagnosis (FDD), which continuously monitors operational status, analyzes potential fault signals, and predicts occurrence probabilities to enable preemptive adjustments or seamless switching to redundant systems. Synergy between the two approaches: FDD enhances robustness and safety, while digital twins reduce operational risks and improve maintenance efficiency.

5.1 Control Platform Based on Digital Twin

The deployment of mobile robot platforms directly impacts task effectiveness and reliability. However, installing newly built or upgraded controllers directly on the equipment carries significant risks, encompassing safety hazards, equipment wear, and downtime costs. Digital twins mitigate these risks by creating virtual entities equivalent to physical counterparts. Through deep integration with perception-cognition-control systems, they provide real-time mirroring of sensors, actuators, and environmental signals, forming a virtual-physical closed-loop system. This significantly reduces deployment risks and optimizes processes. In intelligent logistics scenarios, AMR-based digital twins differ functionally and visually from those in industrial production lines. The former focuses on the distribution, scheduling, and coordination of autonomous mobile units, while the latter emphasizes end-to-end monitoring and 3D visualization of complex processes and equipment status.

Beijing Geek+ develops mobile robots and their simulation deployment platforms; with industry-leading AMR technology and high-quality, flexible integrated solutions, it holds the top position in the global AMR market. [15]; Geek+ 's Robot Management System (RMS) visualizes warehouse environments, route layouts, and robot scheduling through two-dimensional maps. Its simulation platform constructs digital twins based on 1:1 digital models and algorithmic simulations for planning tests and operational analysis. Consequently, Geek+ delivers end-to-end services spanning real-time monitoring to virtual validation, enhancing warehouse operational autonomy and system stability.

5.2 Fault Detection and Diagnosis (FDD)

Traditional FDD consists of two modules: system modeling and system redundancy, further classified by fault type [16]. Modeling methods primarily include model-based and signal-processing-based approaches: the former constructs mathematical models and compares them with current states, while the latter analyzes sensor signals against normal operating conditions. However, these methods lack adaptability for high-dimensional, multi-source, and strongly nonlinear data. In recent years, data-driven artificial neural networks and deep learning have demonstrated strong pattern recognition capabilities, enabling automatic feature extraction and potential fault identification from massive operational data [17]. In mobile robots, typical scenarios include: health monitoring of drive motors and servo systems, consistency comparison between encoders and IMUs, redundancy detection and switching between LiDAR and cameras, as well as real-time monitoring of communication latency and packet loss rates.

To address the limitations of traditional FDD in flexibility and scalability, a digital twin-based remote monitoring solution is proposed. By comparing virtual models with actual operational data through a mobile intelligent platform, it enables real-time deviation detection and predictive analysis of potential failures [18,19]. This significantly reduces operational risks, shortens troubleshooting and fault localization times, and enhances continuous operational capability in harsh environments.

6 Research Challenges and Future Directions

Currently, deep learning-based mobile robots still face numerous challenges in many aspects.

The perception layer must enhance robustness against environmental light, noise, uneven ground, and other disturbances, while addressing issues such as latency, data loss, and data invalidation in multi-sensor fusion. Failure to do so will weaken perception-cognition layer coordination and cause robotic response delays.

In dynamic environments, path planning at the cognitive layer still suffers from slow optimization solutions and inadequate responses to unexpected events. Simultaneously, deep learning models exhibit weak interpretability and “black box” characteristics, undermining the transparency and credibility of decision-making and limiting their effective application in high-risk scenarios.

Future efforts should focus on enhancing model interpretability and trustworthiness while ensuring decision-making performance, improving real-time optimization and rapid replanning capabilities, thereby increasing availability and safety assurance in high-risk operating conditions.

For control layer execution units, upgrading and optimizing physical hardware is crucial. Autonomous control represents the developmental trend for mobile robots; without enhancing performance and reliability at the hardware level—such as in actuators and controllers—it is difficult to achieve more intelligent, convenient, and adaptive control under real-time constraints.

7 Conclusion

This paper reviews the perception-cognition-control integrated system for autonomous mobile robots based on deep learning, detailing its technical architecture and key modules while summarizing cross-module integration and control system optimization/deployment processes. By examining module coordination and integration approaches, it outlines research progress and existing challenges. With the iterative advancement of deep learning, increased computational power, and the deep integration of digital twins with robotics technology, AMRs are poised to achieve safer, more reliable, and efficient autonomous operation in increasingly complex and dynamic environments.

References

1. Sun, M., Wang, S., Xie, Z., et al: Analysis of Autonomous Mobile Robots (AMR) for Intelligent Logistics. Popular Standardization (06), 151-153 (2021)
2. Li, C.: Research on Multi-Sensor Fusion for Indoor Robot Navigation. Master's thesis, Xi'an University of Technology (2025)
3. Wang, Z., Hu X., Li X., et al: A Survey of Global Path Planning Algorithms for Mobile Robots. Computer Science 48(10), 19-29 (2021)
4. Xin, P., Wang, Y., Liu, X., et al: Path Planning Algorithm Based on Optimized RRT and Artificial Potential Field Method. Computer-Integrated Manufacturing Systems 29(9), 2899-2907 (2023)
5. Chen, L., Liu, R., Jia, D., et al: Improvement of the TEB Algorithm for Local Path Planning of Car-like Mobile Robots Based on Fuzzy Logic Control. Actuators 14(1), 12 (2025)
6. Zhang, R., Jiang, W.: Mobile Robot Target Following System Based on Visual Tracking and Autonomous Navigation. Journal of Engineering Design 30(6), 687-696 (2023)
7. Zhong, Z.: Research and Application of Artificial Intelligence Image Processing Technology Based on Deep Learning. Internet Weekly (14), 50-52 (2025)
8. Zhou, W., Lu, L., Wang, J.: A Review of Multi-Sensor Information Fusion in Autonomous Driving. Automotive Digest (1), 7 (2022)
9. Shi, X., Yang, S.: A Review of Multi-Sensor Information Fusion Research. Communications and Information Technology (6), 34-41(2022)
10. Yao, X., Wang, X., Wang, S., et al: A comprehensive survey on convolutional neural network in medical image analysis. Multimedia Tools and Applications 81(29), 41361-41405 (2022)
11. Chen, Y., Chen, Q., Li, Y., et al: Research on Map-Optimized Fusion Positioning Based on Two-Dimensional Lidar and Multi-Ultra-Wideband. High Technology Communications 34(10), 1118-1126 (2024)
12. Gao, Q., Lu, K., Ji, Y., et al: A Review of Multi-Sensor Fusion SLAM Research. Modern Radar 46(8), 29-39 (2024)
13. Wu, D.: Improving the artificial potential field by A-star to solve the local minima problem. Applied and Computational Engineering 11, 34-39 (2023)
14. Zhang, W., Wang, N., Wu, W.: A hybrid path planning algorithm considering AUV dynamic constraints based on improved A* algorithm and APF algorithm. Ocean Engineering 285, 115333 (2023)

15. Li Y.: Geek+: Seeking Future Solutions for Warehouse Logistics Robots. China Entrepreneur (03), 98-101 (2023)
16. Abid, A., Khan, M.T., Iqbal, J.: A review on fault detection and diagnosis techniques: basics and beyond. Artificial Intelligence Review 54, 3639–3664 (2021)
17. Mercorelli, P.: Recent Advances in Intelligent Algorithms for Fault Detection and Diagnosis. Sensors 24(8), 2656 (2024)
18. Jiang, S., Xie, N., Yin, D., et al: Virtual-Physical Interaction System for Mobile Robots Based on Digital Twins. Industrial Instrumentation and Automation Devices (02), 69-75 (2025)
19. Xu, X., Yan, L., Xie, Y.: Design of a Mobile Terminal-Based Remote Monitoring and Fault Diagnosis System for Industrial Robots. Machine Tools and Hydraulics 49(23), 73-76 (2021)