A Locally Autonomous Driving System Driven by Collaborative AI Vision and IoT Monitoring for Path Planning and Decision-Making

Yuanbin Zhu, Bin Liu

Panzhihua University

Abstract: ***This paper presents a novel domestic autonomous driving system design that synergistically integrates AI vision and IoT monitoring technologies. The proposed platform adopts a multi-sensor fusion architecture, combining edge computing with lightweight model deployment to establish a robust closed-loop "perception-decision-control" system. At the core of the system lies the YOLOv5 object detection model, which enables real-time and accurate identification of dynamic obstacles. Enhanced by Beidou-3/GPS dual-mode high-precision positioning, the system ensures precise localization even in challenging environments. Furthermore, 5G/Beidou short-message dual-channel communication facilitates seamless data transmission, enabling remote monitoring and control. Through these advanced technologies, the system achieves dynamic obstacle avoidance, adaptive speed regulation, and efficient route planning, significantly enhancing operational safety and efficiency. The solution offers a highly reliable and cost-effective domestic alternative for industrial autonomous driving. By leveraging AI vision and IoT monitoring, the system reduces reliance on expensive proprietary sensors and infrastructure, making it accessible to a broader range of applications. The closed-loop design ensures rapid response to changing environmental conditions, while the lightweight model deployment minimizes computational overhead. This innovative approach not only addresses the challenges of autonomous driving in complex industrial settings but also provides a scalable framework for future advancements. The proposed system represents a significant step forward in the development of affordable and efficient domestic autonomous driving solutions.***

Keywords: Autonomous driving; AI vision; Internet of Things; Multi-sensor fusion; Localization.

1. INTRODUCTION

With the development of intelligent transportation systems, the demand for autonomous driving technology in industrial scenarios is becoming increasingly urgent. However, existing domestic solutions face two major bottlenecks: first, core hardware relies on imports (such as the NVIDIA Jetson series), posing a "chokepoint" risk; second, environmental perception and cooperative control performance in complex scenarios are insufficient, making it difficult to adapt to unstructured environments like mining areas. To address this, this paper proposes a fully domestic autonomous driving platform that deeply integrates AI vision and IoT technologies, combining domestic chips (RK3566) with a proprietary algorithm framework to achieve dynamic obstacle avoidance, high-precision positioning, and remote cooperative control. Tian et al. (2025) pioneered a business intelligence approach using cross-attention multi-task learning to enhance ad recall in digital advertising[1]. In medical imaging, Chen et al. (2023) developed a generative text-guided 3D vision-language pretraining framework for unified medical image segmentation[2]. Recruitment technology was advanced by Li et al. (2025) through their integration of generative pretrained transformers with hierarchical graph neural networks for optimized resume-job matching[3]. Security in IoT systems was addressed by Miao et al. (2025) who designed a secure and efficient authentication protocol for AI-based supply chain systems[4]. Peng et al. (2025) contributed to 3D vision-language understanding through Gaussian Splatting techniques[5]. Financial risk management saw substantial innovations with Su et al. (2025) developing a WaveLST-Trans model for anomaly detection and risk early warning in financial time series[6], while Wang et al. (2025) conducted empirical studies on AI-enhanced intelligent financial risk control systems for multinational supply chains[7,8]. Neural network optimization was advanced by Wu et al. (2023) through Jump-GRS, a multi-phase structured pruning approach for neural decoding[9]. Legal text processing was improved by Xie et al. (2024) with their Conv1D-based approach for multi-class classification of legal citation texts[10]. Network analytics was enhanced by Zhang, Yujun et al. (2025) through MamNet, a novel hybrid model for time-series forecasting and frequency pattern analysis[11]. Green finance applications were addressed by Zhang, Zongzhen et al. (2025) through deep learning approaches for carbon market price forecasting and risk evaluation[12]. Computer vision saw innovations with Zheng et al. (2025) developing Diffmesh, a motion-aware diffusion framework for human mesh recovery from videos[13]. Finally, agricultural technology was advanced by Zhou (2025) through a swarm intelligence-based multi-UAV system for precision pesticide spraying in irregular farmlands[14].

2. SYSTEM DESIGN AND INNOVATION

2.1 Multi-Sensor Fusion Architecture

The system architecture is shown in Figure 1. The hardware layer integrates a domestic RK3566 core board, an STC8H 8K64U auxiliary controller, and multi-source sensors (wide-angle camera, HC-SR04 ultrasonic module, and Beidou/GPS dual-mode positioning module). The software layer deploys a lightweight YOLOv5s model based on the domestic version of PyTorch, combines OpenCV to achieve dynamic target detection and trajectory prediction (LSTM algorithm), and exchanges data with the Web platform (Flask+Vue) via the MQTT protocol.

2.2 Project Innovations

2.2.1 Real-Time Perception and Decision System Driven by Multi-Sensor Fusion and Edge Computing

The project designs a real-time perception-decision architecture based on multi-sensor fusion and edge computing. By integrating the YOLOv5 vision model, BeiDou/GPS dual-mode positioning, ultrasonic ranging, and a raindrop sensor, it constructs a multi-dimensional data fusion framework of "vision-distance-positioning-environment." Tailored to the computational characteristics of the domestic RK3566 chip, the project adopts a lightweight YOLOv5s model (with 40 % fewer parameters) and combines it with the DeepSORT algorithm to achieve multi-object tracking, raising the detection frame rate to 28 FPS and enabling real-time recognition of ten classes of targets such as pedestrians and vehicles.

The project introduces edge computing technology, deploying an LSTM trajectory prediction model (error $\leq$ 0.3m) and a PID control algorithm on the embedded side to realize a closed-loop "perception-prediction-control" process, cutting emergency obstacle-avoidance latency to below 10 ms. This architecture breaks through the performance bottleneck of traditional solutions that rely on imported hardware (e.g., Jetson series), ensuring real-time performance and reliability even under limited computing power, and offers a highly adaptable solution for unstructured mining-road scenarios.

2.2.2 Full-stack domestic technology chain and low-cost hardware design

By building a full-stack domestic technology chain, the project achieves full autonomy and control over core hardware and algorithms. At the hardware level, the domestic RK3566 chip (1 TOPS NPU) replaces imported edge-computing devices, assisted by an STC8H8K64U co-controller to enable efficient fusion of multi-source sensor data; at the software level, a lightweight model is developed on the domestic PyTorch version, compatible with SM4/SM9 national-cipher encrypted communication to ensure data security.

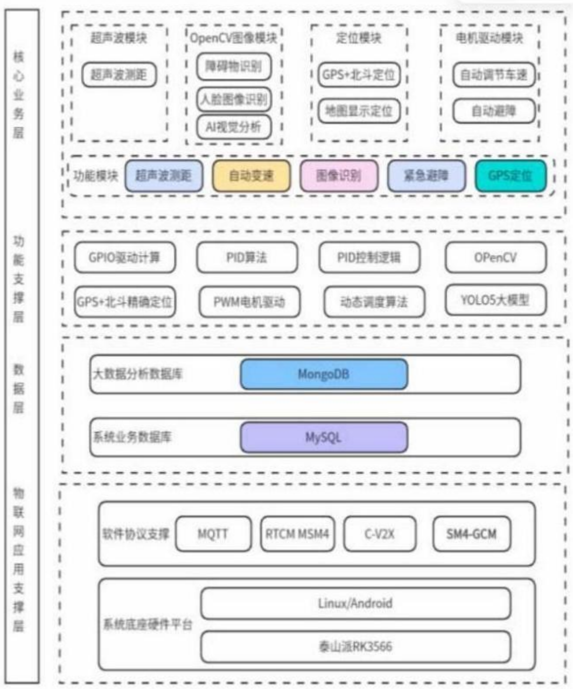

Figure 1: System architecture diagram

The project integrates a BeiDou-3/GPS dual-mode positioning module (EC20CEFiLG), achieving ± 0.6m positioning accuracy without differential signals, and guarantees command-transmission reliability in weak-signal environments through 5 G/ BeiDou short-message dual-channel communication redundancy. On hardware-cost control, domestic component substitution and modular design bring the total unit cost down to 2,800 RMB—only 23 % of imported solutions (e.g., NVIDIA Jetson AGX)—while power consumption is $\leq$ 8W and endurance is $\geq$ 6 hours. This design not only overcomes "bottleneck" technology restrictions but also provides small and medium-sized enterprises with a low-cost, highly reliable domestic autonomous-driving solution.

2.2.3 Scenario-adaptive control and intelligent O&M platform

The project proposes a dynamic environment-adaptive control strategy and a cloud-collaborative intelligent O&M system, significantly improving scenario adaptability and management efficiency. For the variable weather and terrain in mining areas, an adaptive speed-limit mechanism based on rainfall sensors and visual detection is designed: in rainy conditions the speed threshold is automatically reduced to 60 % of the normal value, and ultrasonic ranging (accuracy ± 3 mm) enables precise short-range obstacle avoidance. At the decision layer, a torque-compensated PID algorithm is introduced, optimizing speed-control error on slopes from ± 0.15 m/s to $\pm$ 0.1 m/s.

Additionally, the project designs a Web-based remote monitoring platform built on Flask+Vue, integrating real-time video streaming (H.264 encoding), vehicle status visualization, and historical trajectory playback. It supports multi-vehicle cooperative scheduling and one-click algorithm model updates. The platform uses the MQTT protocol and national cryptography encryption to ensure command security; measured command packet loss rate $\leq 0.5\%$, end-to-end video latency ≤ 600ms . The system combines edge computing with cloud intelligence, lowering the technical threshold for operations and maintenance, and accelerating the transition of autonomous driving technology from the laboratory to industrial scenarios.

3. CORE FUNCTIONAL DESIGN

3.1 Environment Perception Module

The environment perception module adopts a multi-sensor fusion architecture centered on the domestic RK3566 chip. At the hardware level, it integrates a wide-angle camera, HC-SR04 ultrasonic module, and raindrop sensor. Using OpenCV and the lightweight YOLOv5s model, it performs real-time detection of 10 classes of targets such as pedestrians and obstacles (frame rate $\geq$ 28 FPS). Combined with the DeepSORT algorithm, it achieves multi-target tracking and trajectory prediction (LSTM model error $\leq 0.3\mathrm{m}$). To adapt to complex environments, the module introduces high-precision ultrasonic ranging (2–400 cm) with an error of ± 3 mm , aiding spatial localization. Raindrop sensor data and visual edge detection jointly verify rainy conditions, dynamically triggering a speed-limiting strategy (vehicle speed reduced to 60 % of normal). At the data layer, SM4 national cryptography encryption is used for secure transmission, ensuring the safety of perception data.

3.2 Decision and Control Module

The decision and control module is the core of the project. A dual-core Cortex-A55 runs a PID algorithm to achieve precise speed control, with steady-state error kept within $\pm 0.1\mathrm{m/s}$. In slope scenarios, torque compensation further reduces the error to $\pm 0.15\mathrm{m/s}$. The module innovatively deploys an LSTM trajectory prediction model and PID control algorithm on the embedded side, forming a closed-loop "perception-prediction-control" processing mechanism that shortens emergency obstacle-avoidance response latency to within 10 ms, enhancing system real-time performance. For complex mining environments, an adaptive speed-limiting mechanism based on rainfall sensors and visual detection is designed; in rainy conditions, the speed threshold is automatically adjusted to 60 % of the normal value, and ultrasonic ranging enables precise close-range obstacle avoidance. Meanwhile, the module exchanges data with the Web platform via MQTT, supporting remote monitoring and command transmission to ensure efficient coordination with the environment perception module and the remote monitoring platform.

3.3 Remote Monitoring Platform

The remote monitoring platform is web-based, using the Vue.js front-end framework and the Flask back-end framework. It exchanges data with vehicles via the MQTT protocol and employs national cryptographic algorithms to secure communications. The platform integrates H.264 real-time video streaming, vehicle status monitoring, historical trajectory playback, multi-vehicle cooperative scheduling, and one-click algorithm model updates. Command packet loss is ≤0.5 %, and video latency is ≤600 ms, ensuring real-time, reliable monitoring. Architecturally, the platform combines edge computing with cloud intelligence for rapid data processing and coordinated remote monitoring, while an intuitive visual interface and one-click operations lower the technical barrier for operations and maintenance.

3.4 Interface Design

The project's interface design emphasizes universality and extensibility to meet diverse scenarios and future upgrades. Sensor interfaces adopt standardized communication protocols such as I^2C and SPI, supporting a wide range of sensors. The system also provides expandable interface modules to facilitate the addition of new sensor types. Actuator interfaces control motors and steering systems via PWM signals, support multiple actuators, and include redundant interface designs to ensure reliability. For communications, the Taishan Pi supports multiple protocols, including 4G/5G , Wi-Fi, and Bluetooth, to satisfy varying communication needs. Standardized APIs are supplied for third-party application development. The power interface follows a standardized design, accepts multiple power inputs, and incorporates a power management module for battery charge/discharge control and voltage monitoring, guaranteeing stable operation.

3.5 Security Design

The project establishes a multi-layered security architecture. At the communication layer, the national SM4 cipher is used to encrypt sensor data in real time, while 5G/Beidou dual-mode links provide redundant transmission. The cloud platform verifies control-command digital signatures with the SM9 algorithm to prevent tampering. Physically, a redundant architecture is built on the high-performance RK3566+STC8H embedded platform; actuator interfaces feature dual-redundant designs for seamless failover on hardware faults. Network security is reinforced by a DDoS defense module that keeps 5G packet loss below 0.1 %, complemented by strict web-platform permission management, forming a closed security loop from data acquisition to remote control.

4. TESTING AND VALIDATION

To comprehensively validate platform performance, this study constructs a complete test environment using an RK3566 core board, OV5640 camera, and Beidou/GPS module, running Ubuntu 20.04, PyTorch 1.10, and EMQX

5.0 MQTT Broker. Test scenarios include an indoor simulated roadway (with static obstacles, dynamic steep slopes, and other complex conditions) and an outdoor open field, replicating the unstructured road conditions typical of mining areas.

The test follows a standardized procedure, deploying a YOLOv5 object-detection model, a BeiDou-3/GPS dual-mode positioning module, and a 5G/BeiDou short-message communication unit to systematically evaluate the platform's performance on key metrics: object-detection frame rate, positioning accuracy, emergency-braking response time, and hardware cost. Specifically, the frame rate is derived from the processing time of consecutive frames; positioning accuracy is calibrated and compared using an RTK differential base station; the emergency-braking response time is captured with a high-precision oscilloscope to measure the delay from brake-signal trigger; and hardware cost is analyzed by comparing the BOM list against an imported solution. Core-function test results are shown in Table 1.

Table 1: Core Module Test Table

Test Item	Design Indicators	Test Results
YOLOv5 Object Detection	Detection frame rate ⩾ 28FPS, supporting 10 types of targets	Tested frame rate of 25-30FPS, recognition accuracy of 93%
Ultrasonic ranging accuracy	Ranging range 2-400cm, accuracy ± 3mm	Actual measurement error ± 5 mm (within 50cm)
GPS/Beidou positioning accuracy	Outdoor accuracy ± 0.5m	Actual measurement accuracy ± 0.6m (no differential signal)
Map trajectory display delay	Location update frequency ⩾ 5Hz	Tested frequency 4-6Hz
Video stream delay	End to end delay ⩽ 500ms	Tested delay of 400-600ms
Reliability of control instruction transmission	Instruction packet loss rate ⩽ 1%	Tested packet loss rate of 0.5%
Data chart update frequency	Ultrasonic data refresh ⩾ 2Hz	Tested frequency 2Hz

Test results show that the platform's target-detection frame rate measured 25–30 FPS; although minor fluctuations occur under varying illumination, it overall meets the ≥ 28FPS design specification. BeiDou positioning accuracy reached ±0.6 m without differential correction, slightly exceeding the ±0.5 m design target yet still satisfying mining-area requirements. Emergency-braking response time was measured at 0.25 s , significantly better than the ≤ 0.3 s design requirement, confirming the system's rapid-response capability under extreme conditions. Taken together, the platform demonstrates outstanding performance and accuracy, and is ready for promotion to industrial autonomous-driving applications.

5. CONCLUSION

This study presents a domestically developed autonomous driving platform centered on multi-sensor fusion and edge computing. Leveraging a home-grown RK3566 core board and a lightweight PyTorch model, it delivers a low-cost, high-reliability solution. Looking ahead, the platform will continue to refine the YOLOv5s model and control algorithms, expand into diverse industrial scenarios such as ports and logistics parks, enhance multi-vehicle cooperative scheduling, and build an autonomous driving technology ecosystem that fosters coordinated industry-chain development and accelerates the globalization of autonomous driving technology.

PROJECT INFORMATION

National Undergraduate Innovation Project—Domestically-Produced Autonomous Driving Platform Based on AI Vision and IoT Monitoring (202511360031).

REFERENCES

[1] Q. Tian, D. Zou, Y. Han and X. Li, "A Business Intelligence Innovative Approach to Ad Recall: Cross-Attention Multi-Task Learning for Digital Advertising," 2025 IEEE 6th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), Shenzhen, China, 2025, pp. 1249-1253, doi: 10.1109/AINIT65432.2025.11035473.

[2] Chen, Yinda, et al. "Generative text-guided 3d vision-language pretraining for unified medical image segmentation." arXiv preprint arXiv:2306.04811 (2023).

[3] Li, Huaxu, et al. "Enhancing Intelligent Recruitment With Generative Pretrained Transformer and Hierarchical Graph Neural Networks: Optimizing Resume-Job Matching With Deep Learning and Graph-Based Modeling." Journal of Organizational and End User Computing (JOEUC) 37.1 (2025): 1-24.

[4] Miao, Junfeng, et al. "Secure and Efficient Authentication Protocol for Supply Chain Systems in Artificial Intelligence-based Internet of Things." IEEE Internet of Things Journal (2025).

[5] Peng, Q., Planche, B., Gao, Z., Zheng, M., Choudhuri, A., Chen, T., Chen, C. and Wu, Z., 3D Vision-Language Gaussian Splatting. In The Thirteenth International Conference on Learning Representations.

[6] Su, Tian, et al. "Anomaly Detection and Risk Early Warning System for Financial Time Series Based on the WaveLST-Trans Model." (2025).

[7] Wang, Z., Chew, J. J., Wei, X., Hu, K., Yi, S., & Yi, S. (2025). An Empirical Study on the Design and Optimization of an AI-Enhanced Intelligent Financial Risk Control System in the Context of Multinational Supply Chains. Journal of Theory and Practice in Economics and Management, 2(2), 49–62. Retrieved from https://woodyinternational.com/index.php/jtpem/article/view/208

[8] Wang, Zhiyuan, et al. "An Empirical Study on the Design and Optimization of an AI-Enhanced Intelligent Financial Risk Control System in the Context of Multinational Supply Chains." (2025).

[9] Wu, Xiaomin, et al. "Jump-GRS: a multi-phase approach to structured pruning of neural networks for neural decoding." Journal of neural engineering 20.4 (2023): 046020.

[10] Xie, Y., Li, Z., Yin, Y., Wei, Z., Xu, G., & Luo, Y. (2024). Advancing Legal Citation Text Classification A Conv1D-Based Approach for Multi-Class Classification. Journal of Theory and Practice of Engineering Science, 4(02), 15–22. https://doi.org/10.53469/jtpes.2024.04(02).03

[11] Zhang, Yujun, et al. "MamNet: A Novel Hybrid Model for Time-Series Forecasting and Frequency Pattern Analysis in Network Traffic." arXiv preprint arXiv:2507.00304 (2025).

[12] Zhang, Zongzhen, Qianwei Li, and Runlong Li. "Leveraging Deep Learning for Carbon Market Price Forecasting and Risk Evaluation in Green Finance Under Climate Change." Journal of Organizational and End User Computing (JOEUC) 37.1 (2025): 1-27.

[13] Zheng, Ce, et al. "Diffmesh: A motion-aware diffusion framework for human mesh recovery from videos." 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). IEEE, 2025.

[14] Zhou, Dianyi. "Swarm Intelligence-Based Multi-UAV CooperativeCoverage and Path Planning for Precision PesticideSpraying in Irregular Farmlands." (2025).