

# Research on Unmanned-Vehicle Safety Early-Warning System Based on Machine Vision

Shuai Wang, Shengbo Qi, Yajun Han\*

School of Automotive and Traffic Engineering, Jiangsu University of Technology, Changzhou 213001, Jiangsu, China

*\*Author to whom correspondence should be addressed.*

**Copyright:** © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

**Abstract:** The capability to deliver timely early warnings and take countermeasures during driving constitutes one of the most critical competencies for unmanned vehicles, and early-warning solutions based on machine vision are among the most widely-discussed ones nowadays. When travelling, unmanned vehicles need to carefully perceive their surrounding environment. This paper introduces early-warning responses of unmanned vehicles supported by machine vision under various scenarios. Some of these early-warning responses apply to most real-life situations, which provides certain references for research on unmanned-vehicle early-warning systems.

**Keywords:** Unmanned-vehicles; Machine vision; Early-warning research

**Online publication:** October 26, 2026

## 1. Algorithm research on early-warning systems

### 1.1. Early warning without LiDAR

Millimeter-wave radar serves as the core sensing equipment for on-board early-warning systems. It delivers stable velocity and range measurements, yet it has inherent recognition drawbacks and cannot sustain stable and precise performance over long periods. Lacking semantic image-recognition capacity, radars struggle to distinguish guardrails, pedestrians and vehicles. In congested road sections, target confusion and signal overlap frequently occur, giving rise to false early warnings and missed target detection. It also yields poor detection performance for low-lying crossing objects. To address the inherent defects of radar-based early warning, David Nistér<sup>[1]</sup> set up Tesla Autopilot's pure-vision algorithm team. Rejecting radar as the primary sensing hardware, he built an eight-camera surround-view visual early-warning architecture and adopted the CNN deep-learning model to replace traditional manual feature-extraction algorithms. This model can accurately identify hundreds of road targets, effectively reduce misjudgments and missed judgments in early warning, and improve the overall reliability of the early-warning system.

## 1.2. Single-vehicle visual early warning

Compensating for long-distance perception defects of single-vehicle vision represents the mainstream development direction of current visual early-warning systems for unmanned vehicles. Targeting this direction, Li Li<sup>[2]</sup> proposed a dynamic-game-based risk early-warning algorithm. He optimized visual prediction models for China-specific scenarios such as intersection vehicle-meeting, lane cutting-in and road-crossing behaviors, and improved hierarchical early-warning decision-making logics. This algorithm is gradually replacing overseas solutions. The early-warning system integrates vehicle-road collaborative visual perception to realize linkage between single-vehicle early warning and section-wide global early warning.

## 1.3. Quantified dynamic-risk early warning

Domestic urban traffic scenarios are characterized by mixed pedestrian-vehicle flow, frequent non-motor-vehicle crossing, numerous sudden operating conditions and complex road disturbances. Mature foreign early-warning models can hardly adapt to local road conditions. Prominent false and spurious early-warning problems seriously impair the practicality of unmanned-vehicle early-warning systems. Aiming at this industry pain point, Yu Rongjie<sup>[3]</sup> optimized and improved traditional visual risk-quantification models using real domestic urban-road data and mixed-traffic characteristics. He readjusted weight coefficients for hazard-level judgment and dynamic-risk thresholds, and enhanced the system's capacity to identify typical domestic traffic risks such as sudden road-crossing behaviors of pedestrians and non-motor vehicles. Meanwhile, he innovatively put forward a human-machine co-driving early-warning alignment mechanism, which makes machine early-warning logic better conform to drivers' cognitive habits. It effectively cuts down false alarms and frequent alerts on complex roads, and greatly boosts the stability and adaptability of visual early-warning systems under complicated domestic traffic conditions.

# 2 Machine-vision-based environmental perception technology

## 2.1. Object detection for machine vision

Object-detection technology is the key to practical application of machine vision. Environmental recognition can boost algorithm accuracy in decision-making. Nevertheless, affected by environmental factors and complex real-world scenarios, recognition accuracy and response speed still fall far short of practical requirements, and improvements need to be made from algorithmic and other perspectives.

According to research conducted by Zhou Rui et al.<sup>[4]</sup> on LabVIEW-based visual-detection systems, image feature segmentation, feature-quantity learning and processing are adopted to implement recognition based on point-line-surface features of images. However, it has limitations and struggles to process information from complex environments.

To solve the above-mentioned problems, Zhao Nannan et al.<sup>[5]</sup> applied Fourier-transform filtering technology to enhance raw images, which enables more prominent identification of image features and effectively improves the robustness of recognition results.

## 2.2. Deep learning for machine vision

As one of the core technical approaches of AI, deep learning can greatly improve the recognition capability of machine vision. Still, deep learning is confronted with theoretical and data bottlenecks, which means optimizations should be carried out from algorithm-related perspectives.

Inspired by AlexNet, a novel convolutional-neural-network model proposed by Krizhevsky et al., Zhang Hui et al. <sup>[6]</sup> put forward two technical paths: one is proposal-based methods represented by the RCNN series; the other is proposal-free methods represented by YOLO and SSD. Even so, deep learning still suffers from difficulties, including imperfect theories and insufficient data. Nevertheless, they hold the view that the “Parallel Vision” framework (ACP approach) may bring breakthroughs in the future.

Tian Yonglin <sup>[7]</sup> further extended Parallel Vision to three-dimensional detection tasks. Its core innovation lies in the collaboration between artificial worlds and real worlds. Virtual data generate training samples with controllable and automatic annotation to construct the ACP framework: artificial scenarios → computational experiments → parallel execution. It facilitates iterative upgrading of data generation and perception algorithms. This paradigm breaks the traditional development mode of “static dataset + fixed model”. It transforms data from passive resources into active optimization tools and opens up feasible paths for the development of deep-learning theories.

### **2.3. Multimodal machine vision**

Against the backdrop of rapid-fire AI development, machine vision has achieved remarkable progress in fields such as intelligent driving.

Yet deep-learning visual models rely excessively on annotated data. Real-world data suffers from long-tail distribution, scarce samples and high costs, which restrict the development of machine-vision technologies.

Starting from parallel-system theory and the ACP approach, Zhang Hui et al. <sup>[8]</sup> constructed the Parallel-Image technical framework. It generates synthetic image data via regulated artificial scenarios and extreme environments for model training, and achieves continuous optimization through feedback from practical model deployment. It forms the workflow of “modeling — training — feedback — optimization”, which can effectively relieve difficulties including hard-to-obtain real-world data, high costs and imbalanced sample distribution, and improve the generalization ability and robustness of visual models in complex traffic environments. Meanwhile, Zhang Hui et al. <sup>[8]</sup> also discussed prospects for in-depth integration between Parallel Images and generative large-scale models. They pointed out that leveraging semantic-understanding and content-generation capabilities of multimodal large-scale models can further enhance the sense of reality and diversity of virtual scenarios, and offer strong support for the practical production-oriented application of machine-vision technologies.

## **3. Visual-language-model-driven decision-making**

### **3.1. Architecture of visual-language models**

Although visual early-warning systems based on deep-learning models such as CNN and YOLO mentioned above achieve high accuracy, machine-vision-based early warning faces difficulties in tracing root causes for optimization once accidents occur, which hinders further development of unmanned vehicles to a certain extent. Liu H et al. <sup>[9]</sup> proposed migrating the “instruction tuning” technology of large-language models (LLMs) to the multimodal domain. However, relevant integrations remain scarce, resulting in poor generalizability and higher costs. Against such a background, Liu H et al. <sup>[9]</sup> also proposed LLaVA (Large Language and Vision Assistant). They attempted to combine the instruction-following capacity of LLMs with the perception capacity of visual encoders, and generated multimodal instruction data via GPT-4. It provides core methodological and model foundations for subsequent domain-specific applications of Visual-Language

Models (VLMs) such as DriverVLM. Even so, problems including slow inference speed and hallucinations persist.

In response to the above issues, inspired by humans' step-by-step reasoning in problem-solving, Wei J et al. <sup>[10]</sup> proposed Chain-of-Thought prompting. It guides models to conduct analogous reasoning before giving answers: construct triplets (input, chain-of-thought, output) for examples in prompts to guide models toward correct answers. It boasts advantages such as decomposing complex problems, interpretable decision-making processes and strong generalizability. Though the correctness and real-time performance of Chain-of-Thought prompting cannot be fully guaranteed at present and costs remain high, it lays theoretical and methodological foundations for subsequent applications of Chain-of-Thought reasoning in complex decision-making, autonomous driving and other fields.

### **3.2. Integration between reasoning models and autonomous driving**

To make up for deficiencies of VLMs in spatial reasoning and real-time performance, Tian X et al. <sup>[11]</sup> proposed a dual-system mode working collaboratively with traditional autonomous-driving systems. By guiding models to output reasoning chains of "scene description → scene analysis → hierarchical planning" step-by-step, every step of early-warning decision-making becomes analyzable. In unmanned-vehicle scenarios, Chain-of-Thought reasoning can be specified as follows: first, parse current road images and identify key elements such as lane lines, traffic signs, pedestrians and vehicles together with their spatial relationships; second, carry out dynamic analysis combined with motion-state parameters; finally, implement hierarchical planning according to preset rules. Compared with models that directly output instructions, reasoning chains generated by Chain-of-Thought reasoning offer clear routes for accident traceback and model optimization.

In addition, in-depth integration of VLMs with the Parallel-Vision ACP framework proposed by Tian Yonglin <sup>[7]</sup> and Zhang Hui et al. <sup>[8]</sup> can be implemented. Large-scale annotated data generated by artificial scenarios within parallel systems is adopted to fine-tune VLMs for transportation domains, which can effectively improve model robustness and accuracy in complex traffic scenarios.

### **3.3. VLM-based early-warning-system framework and localized adaptation**

Mao J et al. <sup>[12]</sup> attempted to use large-language models as decision-makers for autonomous driving. LLMs receive textualized environmental and vehicle information to instruct MPC. High-level decision-making of LLMs and high-frequency decision-making of MPC are utilized to implement concrete planning.

Facing complicated domestic urban-traffic conditions including mixed pedestrian-vehicle flow, frequent non-motor-vehicle crossing and numerous sudden operating conditions, VLM-based early-warning systems require targeted adaptation. Xu Z et al. <sup>[13]</sup> attempted to construct a multimodal large-language model that can not only interpret driving behaviors and perform reasoning, but also directly predict underlying vehicle-control signals. Combined with the human-machine co-driving early-warning alignment mechanism put forward by Yu Rongjie <sup>[3]</sup>, a cognitive-alignment module can be added at the output end of VLM reasoning. It makes interpretation of machine-reasoning logic better conform to drivers' cognition, facilitates human-machine mutual understanding, and effectively reduces false alarms and frequent alerts under complicated domestic road conditions.

## **4. Research on generational development of machine-vision-based unmanned-vehicle early-warning systems**

### **4.1. Domestic research status**

Domestic research on unmanned-vehicle early-warning systems focuses on breakthroughs in three major directions: on-board perception early warning, vehicle-road collaborative beyond-visual-range early warning, and cloud-based global-risk early warning. Early-warning latency, recognition accuracy, and adaptability under extreme environments keep improving, and some technologies have reached internationally-advanced levels.

#### **4.1.1. Focus on perception algorithms**

Construct adversarial-risk early-warning and human-like decision-making safety frameworks. Tackle core challenges such as long-distance detection, recognition under bad weather, and defense against malicious interference. The team led by Academician Li Jun from Tsinghua University<sup>[14]</sup> addressed detection failures of traditional autonomous-driving visual perception under long-distance conditions, camera shake, rain and fog. They proposed a dual-modeling network based on height-space modeling, which adapts to dual scenarios of low on-board perspectives and high roadside perspectives. Obstacle-detection performance is improved by 28 %, and early warning for long-distance obstacles and special-shaped obstacles can be realized in advance.

#### **4.1.2. Large-scale deployment of early-warning systems**

Huawei ADS intelligent-driving early-warning system integrates multiple sensors including LiDAR, millimeter-wave radar and cameras. It identifies pedestrians, non-motor vehicles, falling rocks and crashed vehicles at millisecond-level speed, and realizes forward-collision early warning (AEB), blind-lane lane-change early warning and risk early warning for unprotected left-turn maneuvers. Another example is Xpeng XNGP full-domain early-warning system. It adopts the integration scheme of pure vision plus lightweight LiDAR and multi-frame temporal Transformer algorithms. It can give warnings for pedestrians and non-motor-vehicles darting out at intersections as well as oncoming vehicles around bends, and predict risks for merging-in and merging-out on highway ramps three seconds in advance.

#### **4.1.3. Smart highways and urban intelligent-connected-vehicle demonstration zones**

In this regard, Yao Danya from Tsinghua University and her team<sup>[15]</sup> put forward a map-based road-network early-warning system in cooperation with the China Academy of Safety Science and Technology, an achievement of the National Key R&D Program. Integrating massive anonymous driving big data and map spatiotemporal data, it can identify multiple types of high-risk scenarios. Millimeter-wave radars, high-definition cameras and edge-computing units are deployed along the Jingxiong Expressway. Targeting highway-ramp scenarios, speed-control and hazard-avoidance early warnings are pushed to autonomous-driving vehicles via vehicle-road-cloud communication, solving the beyond-visual-range perception blind-spot problem of on-board sensors.

### **4.2. Foreign research status**

Research on unmanned-vehicle early warning started earlier abroad. Systems such as the United States' pure-vision / multi-sensor-fusion route and Germany's vehicle-grade safety-redundancy early warning have taken shape.

#### 4.2.1. Research on visual-scheme models

Tesla sticks to the pure-vision scheme. It realizes forward-collision early warning (AEB), blind-lane early warning and lane-departure early warning relying on cameras and neural networks. It self-developed HW3.0 and HW4.0 for low-cost large-scale deployment. David Nistér<sup>[1]</sup> proposed abandoning millimeter-wave radar as primary perception hardware and building an eight-camera surround-view visual early-warning architecture.

#### 4.2.2. Radar-driven early-warning research

Waymo adopts redundant perception combining LiDAR, millimeter-wave radar and cameras. It mainly develops Robotaxi unmanned vehicles and represents one of the most reliable L4-level early-warning schemes at present.

#### 4.2.3. Safety-redundancy architectures

European vehicle manufacturers adhere to the “safety-redundancy priority” principle. Bosch takes the lead in national-level research projects such as ConnRAD and STADT: to develop V2X vehicle-road collaborative early warning, left-turn-conflict early warning and bad-weather perception early warning. Mercedes-Benz Drive Pilot L3-level autonomous-driving system is equipped with multi-sensor-fusion early warning, driver-fatigue-monitoring early warning and hierarchical takeover-request early warning, complying with strict EU vehicle-grade safety certifications. Dr. Lutz Bürkle took the lead in putting forward technical routes for urban-scenario V2X vehicle-road collaborative early warning, left-turn-conflict risk prediction and multi-sensor-fusion early warning under bad weather. He advocated adapting to stringent EU vehicle-grade requirements with safety-redundancy architectures.

### 5. Conclusion

This paper sorts out core algorithms of unmanned-vehicle early-warning systems, machine-vision environmental-perception technologies and visual-language-model-driven decision-making systems. It analyzes advantages, disadvantages and optimization schemes of mainstream algorithms such as pure-vision early warning and quantified dynamic-risk early warning, explores the empowering effect of deep learning and Parallel Vision on environmental perception, and summarizes technical research and deployment status of unmanned-vehicle early-warning systems at home and abroad. On the whole, domestic technologies are better adapted to complicated local road conditions, while foreign technologies started earlier and are relatively mature. It provides theoretical references for subsequent optimization of unmanned-vehicle early-warning systems for complex traffic scenarios.

### Disclosure statement

The authors declare no conflict of interest.

### References

- [1] Nister D, Stewenius H, 2004, Linear Time Maximally Stable Extremal Regions. Computer Vision–ECCV 2004. Proceedings, Part II 8, 345–357.
- [2] Li L, Wang Y, Zheng N, et al., 2007, Research and Development of Intelligent Analysis for Driving Behaviors. Acta

Automatica Sinica, 10: 1014–1022.

- [3] Wang Y, Yu R, 2024, Analysis on Influencing Factors of Vehicle Longitudinal Interaction Safety Under Human–Machine Mixed Driving. *Journal of Transport Information and Safety*, 42(3): 11–19.
- [4] Zhou R, Wang M, Yang Z, 2016, Research on Visual Detection System Based on LabVIEW. *Technology Outlook*, 26(12): 166.
- [5] Zhao N, Deng C, Wen Z, et al., 2025, Fast Recognition Algorithm for Salient Image Targets Based on Machine Vision. *Journal of Jilin University (Engineering and Technology Edition)*, 55(10): 3346–3351.
- [6] Zhang H, Wang K, Wang Y, 2017, Application Progress and Prospect of Deep Learning in Object Visual Detection. *Acta Automatica Sinica*, 43(8): 1289–1305.
- [7] Tian Y, 2022, Research on Key Parallel-Vision Methods for 3D Object Detection in Traffic Scenarios. University of Science and Technology of China.
- [8] Zhang H, Tian Y, Wang Y, et al., 2026, Parallel Images: From Theoretical Framework to Multimodal Intelligent-Vision Applications. *Journal of Image and Graphics*, 31(7): 2358–2380.
- [9] Liu H, Li C, Wu Q, et al., 2023, Visual Instruction Tuning. *Advances in Neural Information Processing Systems (NeurIPS)*.
- [10] Wei J, Wang X, Schuurmans D, et al., 2022, Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *Advances in Neural Information Processing Systems (NeurIPS)*.
- [11] Tian X, Gu J, Li B, et al., 2024, DriveVLM: The Convergence of Autonomous Driving and Large Vision-Language Models. *Conference on Robot Learning, PMLR*.
- [12] Sha H, Mu Y, Jiang Y, et al., 2023, LanguageMPC: Large Language Models as Decision Makers for Autonomous Driving. *arXiv preprint arXiv:2310.03026*.
- [13] Xu Z, Zhang Y, Xie E, et al., 2024, DriveGPT4: Interpretable End-to-End Autonomous Driving via Large Language Model. *IEEE Robotics and Automation Letters*, 9(2): 1183–1190.
- [14] Li J, Zhang X, Lu Y, Gao X, Huang Y, Liu H, Wang Y, 2025, Vehicle-Road Cooperative Perception Technology and Its Development Trend for Intelligent Connected Vehicles. *Acta Automatica Sinica*, 51(2): 233–248.
- [15] Yao D, Zhang Y, Li L, et al., 2021, Vehicle-Road Collaboration and Intelligent Technology. *Journal of Transportation Systems Engineering and Information Technology*, 21(5): 40–51.

**Publisher's note**

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.