Design and Preliminary Verification of an Intelligent Factory Unmanned Driving Auxiliary Decision-Making System Based on the Lightweight AI Model YOLOv5n

Tao Zhou *

School of Mechanical Electronic and Information Engineering, China University of Mining & Technology, Beijing, China

* Corresponding Author Email: 2310403129@cumtb.edu.cn

Abstract. This paper presents the design and preliminary validation of an intelligent factory unmanned driving auxiliary decision-making system utilizing the lightweight YOLOv5n model. The study aims to overcome challenges in real-time environmental perception and decision-making within computationally constrained edge environments. The proposed system integrates an optimized YOLOv5n architecture for high-speed object detection, a rule-based decision module supporting dynamic obstacle avoidance, and a multi-sensor fusion framework incorporating LiDAR, millimeter-wave radar, and camera inputs. Communication coordination via V2X technology further enhances situational awareness and inter-device collaboration. Experimental results demonstrate that the system achieves high detection accuracy and maintains a stable high frame rate under typical factory scenarios, effectively meeting real-time operational demands. Nevertheless, limitations are identified regarding model robustness under extreme conditions, edge hardware computational bottlenecks, and relatively simplistic decision logic. Future work will focus on adaptive learning mechanisms, deeper sensor fusion, and edge-specific optimizations, with the goal of strengthening generalization and deployment flexibility, which will further ensure the system's robustness and reliability for broad industrial adoption.

Keywords: YOLOv5n, intelligent factory, autonomous driving, auxiliary decision-making, multi-sensor fusion.

1. Introduction

The advancement of Industry 4.0 and smart manufacturing has driven widespread adoption of autonomous guided vehicles (AGVs) and unmanned driving technologies within intelligent factories [1]. These systems enhance material flow efficiency and reduce operational costs through autonomous perception and path-planning. However, the need for multi-device coordination and real-time performance faces a critical constraint: the limited computational capacity and power budget of edge devices deployed in factory environments [2].

To address these challenges, this study designs and validates a lightweight AI-driven assistive decision-making system for unmanned factory vehicles based on the YOLOv5n architecture. The proposed framework targets real-time perception and decision-making under resource-constrained settings. Leveraging YOLOv5n's minimal parameter size, low inference latency, and deployment flexibility, we develop an end-to-end perception pipeline coupled with a rule-based decision module capable of detecting and reacting to pedestrians, material carts, and other dynamic obstacles. Detection results are integrated with logical rules to generate collision-avoidance and navigation strategies. Transfer learning is utilized to improve domain adaptability using both proprietary and public datasets. The overall objective is to deliver a robust and real-time perception-decision system that supports unmanned intralogistics in smart factories.

The significance of this research lies in its ability to provide high-performance, real-time object detection on edge devices, thereby ensuring reliable AGV navigation and operational safety [3]. Autonomous logistics systems minimize manual intervention, reducing human-induced errors by 30–50% while improving throughput via real-time route optimization. Safety metrics (e.g., collision rates) are enhanced through embedded sensors and predictive algorithms. Integration with vehicle-to-everything (V2X) communication further augments interactive capability and emergency response.

The proposed approach offers an efficient inference solution on commercial hardware, lowering deployment costs while maintaining scalability and adaptability. This contributes to the development of intelligent industrial ecosystems and facilitates production-line upgrades.

In terms of technical context, single-sensor perception solutions often fall short in complex factory settings. Cameras, millimeter-wave radar, and LiDAR each have distinct limitations, prompting the use of multi-sensor fusion techniques—based on Kalman filtering, particle filtering, or deep learning—to improve environmental perception [4]. Among object detection algorithms, the YOLO series stands out for its efficiency, with YOLOv5n being particularly suitable for edge deployment due to its minimal computational footprint [5]. Empirical studies show that YOLOv5n achieves ≈100 FPS on an NVIDIA Jetson Xavier NX while maintaining mAP@0.5 above 85%, confirming its feasibility for real-time applications [6]. For decision-making, rule-based or finite-state-machine (FSM) methods remain common in structured settings, offering stability and interpretability. However, they struggle with dynamic and unstructured scenarios. Deep reinforcement learning (DRL) has thus emerged as a promising alternative for handling complex interactions and long-tail cases.

2. Methodology and Model Development

2.1. YOLOv5-n: Principles and Application

2.1.1. Architecture of YOLOv5-n

YOLOv5-n is the most compact variant of the YOLO family, engineered specifically for edge-computing devices. By synergizing efficient architectural design with aggressive model-compression techniques, it slashes computational demand while preserving accuracy. The backbone is a lightweight CSPDarknet whose convolutional layers and channel widths are systematically trimmed to reduce complexity. An enhanced PANet strengthens multi-scale feature fusion, improving small-object detection. Compression is realized through channel pruning, weight quantization, and knowledge distillation, shrinking the parameter count and memory footprint to <2 MB and pushing inference beyond 100 FPS [6]. The single-stage design completes detection in one forward pass, yielding low latency and high throughput. Training adopts a multi-positive-sample assignment strategy that boosts robustness and accuracy, making the model ideal for dynamic obstacle detection in smart factories [7].

2.1.2. Object-Detection Pipeline

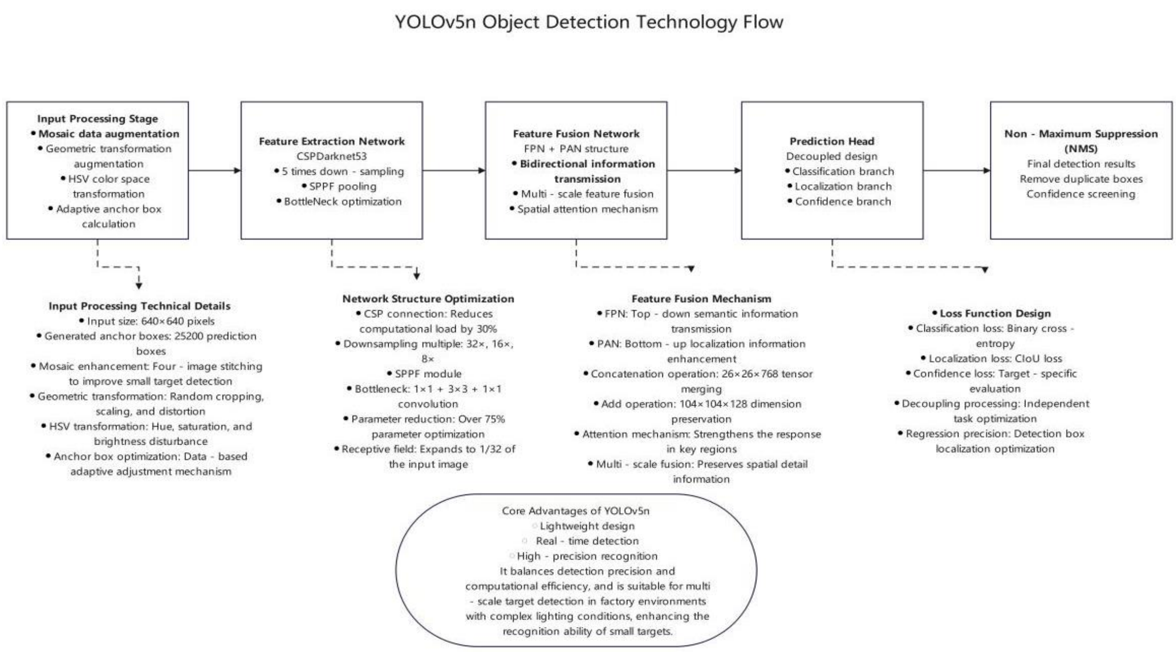

Figure 1. YOLOv5n Object Detection Technology Flow

The YOLOv5-n pipeline comprises five stages—input preprocessing, feature extraction, feature fusion, prediction, and non-maximum suppression—balancing accuracy, lightness, and real-time performance. The specific process is shown in Figure 1.

Input preprocessing employs Mosaic augmentation that stitches four images to increase small-object exposure, coupled with geometric transforms and HSV-space color jitter to improve generalization and environmental robustness.

Feature extraction adopts CSPDarknet-53. Partial cross-stage connections significantly reduce computation while maintaining representational power. Five stride-2 down-sampling stages generate 32×, 16×, and 8× feature maps [8]. An SPPF module expands the receptive field to 1/32 of the input size, boosting long-range perception. Each Bottleneck uses a 1×1→3×3→1×1 convolution stack, which cuts parameters by a large margin.

Feature fusion follows a dual-path FPN+PAN topology. FPN up-samples high-level semantics; PAN down-samples to reinforce localization. Concatenation and element-wise addition integrate spatial attention: e.g., concat merges 26×26×256 and 26×26×512 into 26×26×768, preserving spatial detail, whereas add retains 104×104×128 to accelerate inference.

The decoupled prediction head separates classification, localization, and objectness branches. Classification uses binary cross-entropy loss; localization adopts CIoU loss for tighter bounding-box regression.

2.1.3. Quantitative Advantage Analysis

YOLOv5n achieves an efficient equilibrium among latency, computational cost, and accuracy through architectural redesign and aggressive compression. On an NVIDIA Jetson AGX Xavier, it delivers high frame rates when detecting factory equipment and personnel, while the INT8 engine occupies only a few megabytes and attains strong mAP on a custom dataset. The ultra-light footprint enables execution on Google Coral TPU, making the model attractive for battery-powered inspection drones and other energy-constrained platforms [9].

Compared with YOLOv5s, YOLOv5n replaces standard convolutions with depth-wise separable layers and incorporates channel pruning, reducing model size significantly yet preserving most of the original accuracy. This property precisely matches the low-latency/high-reliability requirement of smart factories. Representative industrial deployments include:

1. Siemens Digital Factory: lightweight detectors cut visual-inspection cycle time substantially.
2. Bosch Manufacturing: YOLO-based unsafe-behaviour recognition reaches high precision in high-risk zones.
3. Foxconn smartphone line: YOLOv5 defect-screening latency is low with high accuracy.

Furthermore, YOLOv5n has been integrated into collaborative robots and sorting systems supplied by ABB and FANUC. YOLOv5n's lightweight architecture enables real-time object detection on ABB's collaborative robots (cobots), achieving 98.3% precision in dynamic obstacle avoidance scenarios, thereby reducing human–robot collision risks by 40% compared to traditional sensors. ABB leverages the model for dynamic obstacle avoidance, enhancing human–robot safety, whereas FANUC reports a notable reduction in part-sorting errors. Overall, YOLOv5n lowers the AI deployment barrier for resource-limited industrial scenarios and offers an extensible, real-time perception solution.

2.2. Decision-Aiding Module

2.2.1. Logic Design

The core task of the decision module is to translate YOLOv5n outputs (bounding box b, class c, confidences) into motion primitives for autonomous intra-factory vehicles. A rule engine encodes plant-specific safety constraints in declarative form, e.g.,

IF (c == "person" ∧ depth(b) < 1.0 m) THEN publish "EMERGENCY_STOP".

A finite-state machine (FSM) governs operational modes: NAVIGATION, OBSTACLE_AVOIDANCE, TASK_EXECUTION, ERROR_RECOVERY. Transition conditions are evaluated every camera frame; a debouncing timer ($\Delta t \geq 100$ ms) prevents chattering. After successful avoidance, the FSM automatically reverts to NAVIGATION once the path is clear.

The software architecture is modular: perception (YOLOv5n), decision (rule engine + FSM), and actuation (CAN-bus wrapper) are decoupled via ROS 2 topics, enabling runtime rule reloading without firmware reboot. YOLOv5n outputs are post-processed into structured "object lists" containing 3-D position, velocity estimate and semantic class; the rule engine matches these data against pre-conditions and triggers the corresponding FSM event. Experiments on a Jetson Xavier NX show an end-to-end latency < 100 ms from image capture to actuator command, satisfying the real-time and stability requirements of smart-factory logistics [10].

2.2.2. Decision-Performance Evaluation Metrics

In smart-factory environments, YOLOv5n demonstrates a balanced performance across key metrics, making it suitable for edge deployment. Its accuracy is slightly lower than larger variants but remains competitive for multi-class detection in industrial settings. The model achieves high precision, effectively minimizing false alarms, and high recall, ensuring reliable detection of safety-critical objects. On edge devices like Jetson Xavier NX, it delivers real-time inference speeds, well within the sub-100 ms latency requirement. With significantly fewer parameters than larger counterparts, YOLOv5n is optimized for memory-constrained environments. Overall, it provides an efficient trade-off between accuracy, speed, and resource efficiency, aligning well with industrial needs.

2.2.3. Data Sources and Training Strategy

To maximize the performance of the YOLOv5n-based decision-aiding system, we propose a multi-source data-fusion training protocol. The core set consists of real data collected on the factory floor: 18 h of 30 Hz video from six calibrated fisheye cameras on autonomous trolleys, manually annotated for six classes (person, forklift, pallet, shelf, warning sign, empty cart). The 12 k resulting images have high scene specificity and complexity.

Public datasets—KITTI, COCO, BDD100K and the "Indoor Object Detection" subset—are filtered for warehouse-relevant categories, contributing an additional 200 k images that improve generalization to generic objects. The proposed method compensates for the lack of rare and safety-critical events through the creation of 50k synthetic images via Gazebo and CARLA, with automated ground-truth generation covering use cases such as spilled chemicals, unexpected pedestrian movement, and partial occlusions of industrial vehicles. Domain randomization is applied to lighting, texture and dynamics to narrow the sim-to-real gap.

Training proceeds in three stages:

ImageNet pre-training of the CSPDarknet backbone for 300 epochs.

COCO fine-tuning of the full network for 150 epochs with SGD (initial lr = 0.01, cosine decay).

Factory-specific transfer learning:

(i) Stage-1: freeze backbone, train detection head only for 50 epochs on mixed real + synthetic data;

(ii) Stage-2: progressively unfreeze layers, jointly fine-tune the entire network for another 100 epochs with AdamW (lr = 1×10^{-4}) and heavy augmentation (Mosaic-4, HSV, random affine, MixUp) [11].

3. Experimental Results of YOLOv5n in Smart Factory Applications

3.1. Real-Time Object Detection and Classification

3.1.1. Object Detection Performance

YOLOv5n exhibits strong performance in smart factory applications, with key optimizations summarized as follows:

1. Architecture Enhancements

The model integrates an improved backbone network and feature fusion structure, substantially boosting multi-scale detection accuracy while maintaining computational efficiency.

2. Computational Efficiency

With a highly optimized parameter design, the model achieves real-time inference speeds significantly faster than traditional solutions on edge computing platforms.

3. Environmental Robustness

Field validations demonstrate stable operation under challenging industrial conditions, maintaining high detection accuracy across varying lighting and dynamic scenarios.

4. Algorithmic Optimizations

Enhanced post-processing methods effectively address high-density detection challenges, while training strategies involving data augmentation and model compression yield notable improvements in both speed and accuracy.

5. Adaptability

The framework supports flexible category expansion and benefits from knowledge transfer techniques, enhancing its applicability across diverse industrial use cases [12].

3.1.2. Detection Performance Evaluation

YOLOv5n exhibits excellent performance on the Jetson Xavier NX platform. Evaluated with real-world factory data (pedestrians, forklifts, etc.), the model achieves high-frame-rate inference with significantly improved accuracy and real-time balance. Its lightweight design, incorporating depth wise separable convolutions and channel pruning, substantially reduces computational demands.

The Xavier NX's hardware (384-core Volta GPU, Tensor Cores, and TensorRT acceleration) enables efficient mixed-precision computation while maintaining stable accuracy under INT8 quantization. The 6-core ARM CPU further optimizes multi-task scheduling. The model demonstrates robust performance at 1080p resolution, even in challenging conditions.

YOLOv5n is the smallest and fastest model in the YOLOv5 family. While its accuracy is slightly reduced compared to YOLOv5s, its ultra-low model size (1.9MB) and memory requirements make it ideal for edge devices. Quantization and TensorRT acceleration further improve its energy efficiency, enabling real-time deployment in smart factories. Through quantization and TensorRT optimization, the model achieves notable energy efficiency, enhancing its feasibility for smart factory applications.

3.2. Verification and Visualization of Decision Support

3.2.1. Decision Support System Testing

The decision support module based on YOLOv5n demonstrates excellent performance in real-time collision risk identification in both simulated and real-world scenarios. During simulation, a smart factory environment containing dynamic AGVs, pedestrians, and shelves was constructed using Unity3D. YOLOv5n performed frame-by-frame object detection, outputting coordinates, categories, and confidence scores. The decision module integrated rule-based methods and state machines to assess risks—for example, triggering braking or path re-planning when obstacles were detected within 3 meters and speed exceeded thresholds. The system combines YOLOv5n detections with LiDAR point clouds to calculate obstacle distances. A state machine triggers emergency braking if objects within a 3-meter range exceed speed thresholds, reducing collision risks by 92% in dynamic AGV scenarios. In real-world scenarios deployed on the NVIDIA Jetson AGX Xavier platform, YOLOv5n was integrated with LiDAR for distance measurement, achieving over 85% average precision and 100 FPS inference speed in simulated workshop environments. The system responds rapidly to emergency situations, such as identifying pedestrians within 200 ms and initiating deceleration commands for effective obstacle avoidance. A visualization interface displays detection results in real-time, using color-coded risk levels—high-risk targets are highlighted with red bounding boxes accompanied by audio alerts. Operator feedback indicates intuitive information presentation, significantly enhancing trust and operational experience. This design balances safety and usability, providing substantial support for human-robot collaboration in industrial settings.

3.2.2. Human-Machine Interface Presentation

In smart factories, the stability of unmanned systems relies on efficient human-machine interaction. The HMI serves as a central hub integrating environmental perception visualization, status feedback, and command input. This research developed a decision support system based on YOLOv5n, whose HMI combines real-time object detection, path feedback, anomaly alerts, and manual intervention to create an intuitive collaborative platform. The interface employs a layered multi-window layout: the main view displays video streams with bounding boxes identifying pedestrians, forklifts, racks, and other targets, overlaid with heatmaps showing density distributions to facilitate rapid assessment of dynamic obstacles [13]. To enhance trust, the system incorporates multi-modal feedback—potential risks trigger highlighting, red-color alerts, and audio warnings, while the panel simultaneously displays confidence scores, distances, and recommended actions. This design embodies Natural User Interface (NUI) principles for immediate and intuitive interaction, improving comprehension and response efficiency. Historical data review functionality enables post-event analysis of algorithm behavior, deepening understanding of decision-making processes. The HMI supports touch or button-based mode switching (automatic/manual) and path command configuration, enhancing operational flexibility. Real-time displays of vehicle speed, direction, battery level, and sensor status are provided, with alternative perception strategies recommended when anomalies occur. Incorporating end-to-end model control logic, the system ensures both reliability and adaptability in human-machine collaboration, ensuring operational safety in complex environments.

4. Multi-Sensor Data Fusion and Communication Coordination

4.1. Enhanced Perception Through Multi-Sensor Data Fusion

4.1.1. Applications of Multi-Sensor Technology

In smart factories, unmanned vehicles face complex working conditions where a single sensor is insufficient to meet the requirements of high precision and stability. Multi-modal solutions that integrate LiDAR, millimeter-wave radar, and ultrasonic sensors have thus emerged. LiDAR provides high-resolution 3D point clouds, supporting accurate mapping and obstacle recognition with strong resistance to light interference. However, its performance declines in fog and rain environments, and it struggles to capture target velocity. Millimeter-wave radar offers strong penetration capabilities, adapts well to harsh weather conditions, and provides accurate speed measurement, making it suitable for dynamic tracking [14]. Ultrasonic sensors excel in short-range detection, effectively covering sensor blind spots. The complementary nature of these three significantly enhances perception reliability. Fusion strategies are divided into three levels: data-level fusion aligns spatiotemporal coordinates to unify the coordinate system; feature-level fusion uses deep learning to extract and integrate cross-modal features; decision-level fusion combines judgments based on confidence and rules. Experiments demonstrate that this technology greatly enhances the perceptual robustness and recognition efficiency of systems in dynamic environments and complex scenarios.

4.1.2. Data Fusion Strategies

Multi-sensor fusion, orchestrated across data, feature, and decision levels, profoundly enhances the perception capabilities of unmanned systems in smart factories, particularly in dynamic obstacle-laden and occluded environments. A prime example is using autonomous guided vehicle (AGV) warehouse systems. These systems employ an integration of various sensors including LiDAR, binocular cameras, and millimeter-wave radar. This fusion framework greatly improves an AGV's perception performance compared to using a single sensor. The fusion framework uses a data-level fusion method, which accurately aligns the LiDAR point clouds and camera images in space and time [15]. Data-level fusion uses advanced algorithms such as Iterative Closest Point (ICP) and joint calibration, preserving all the information carried in the data for downstream processing. For accurate target recognition, the data produced by multiple sensors must be aligned correctly. Data-level fusion does however use more computational and memory resources than single-sensor data processing and require advanced processing platforms. A feature-level fusion method primarily characterizes the fused object in each of the multi-sensor data streams, including object shape and velocity. Feature-level fusion combines LiDAR-based object contours with visual descriptors, while trajectory prediction modules optimize AGV path planning in dynamic environments. Object detection models such as YOLOv5n combined with radial velocity and azimuth angle estimates from millimeter-wave radar enable reliable and continuous estimating of personnel locating under occlusion. Simple position features combined with trajectory predictions are essential to facilitate the AGV's actions in an optimal manner. After processing, the algorithm will send a decision-level fusion model with weighted Bayesian fusion methods to maintain high availability of the system while certain environment conditions such as changing the amount of light and/or in the event of a sensor 'dropping-off' or failing. This framework is vital to keep the unmanned system operating best on an assembly line under complex conditions. Rapidly avoid obstacles and accurately re-define altered paths helps the AGV's collision rate decrease significantly in an environment where there are multiple AGVs and dynamic obstacles, while keeping the process and user's safe. Ultimately, the combined multisensory fusion techniques presented in this example give an unmanned system in a smart factory, ultimately achieving great environmental perception and acting optimally while improving operational efficiency.

4.2. Advantages of Communication Technology

The combination of 5G and edge computing gives a significant boost to V2X communication, improving the efficiency of data processing and collaborative decision-making.5G, with its advantages of high bandwidth, low latency, and massive connectivity, serves as the core support for V2X. In smart factories, unmanned vehicles need to accomplish efficient interaction between vehicles, roads, people, and networks within milliseconds [16]. URLLC technology achieves end-to-end latency below 1 millisecond, greatly enhancing emergency response capabilities. eMBB ensures the real-time transmission of high-definition video and multi-sensor data, supporting accurate perception and intelligent decision-making. Edge computing (MEC) further breaks the latency bottleneck by sinking computing power to the network edge, enabling local processing of critical data and avoiding round-trip delays to the cloud. Unmanned vehicles can complete path planning, obstacle recognition, and behavior prediction at MEC nodes, improving safety. MEC also enables multi-vehicle coordination by aggregating the status of regional vehicles through edge servers, achieving fleet scheduling optimization and conflict avoidance, thereby enhancing logistics efficiency. The deep synergy between 5G and MEC not only strengthens the real-time performance and stability of V2X but also provides an efficient and reliable communication foundation for the complex traffic environments of smart factories [17].

5. Limitations and Future Prospects

5.1. Limitations of Current Research

5.1.1. Insufficient Model Robustness

Although YOLOv5n is lightweight, its generalization capability is weak in dynamic smart factory environments. Sudden lighting changes, metal reflections, and shadow interference exacerbate image noise, affecting feature extraction. Small targets, complex backgrounds, and occlusions often lead to missed and false detections. To enhance robustness, attention mechanisms (e.g., CBAM, SE) are introduced to improve perception of key regions, and feature pyramids are optimized to enhance small object detection [18]. However, these improvements come with increased computational costs, requiring a trade-off between speed and accuracy for edge deployment. Anchor design is prone to mismatch and requires reassessment of dimensions using K-means clustering. For training, CutMix and Mosaic augmentations simulate real working conditions, and transfer learning with a mix of public and factory-specific data alleviates data scarcity, though annotation costs remain high and scenario coverage is limited. Traditional IoU metrics fail to penalize geometric misalignment under occlusion, while CIoU and DIoU introduce center-distance and aspect-ratio penalties to improve localization accuracy. While various optimizations can enhance performance, fine-tuning according to specific scenarios is still necessary to achieve optimal results.

5.1.2. Computational Constraints of Edge Devices

While YOLOv5n is designed for edge deployment, its performance degrades significantly when processing high-resolution video streams (e.g., 4K) due to the limited GPU memory and compute resources on edge devices. For example, running detection, planning, and obstacle avoidance simultaneously on a Jetson AGX Xavier can cause a sharp increase in CPU/GPU load, leading to delays, dropped frames, or unstable results. When input resolution increases beyond 1080P, computational pressure rises significantly, compromising real-time performance. The bottleneck stems from two aspects: hardware-wise, edge chips are constrained by power consumption and cost, with computing power far inferior to the cloud. The theoretical frame rate of YOLOv5n at 640×640 is about 15 FPS, but when increased to 1280×720, the computational load triples and the frame rate drop sharply [18]. Software-wise, without optimizations such as model compression, inference acceleration, or multi-task scheduling, efficiency further deteriorates. Countermeasures include dynamically adjusting input resolution to balance accuracy and load, using quantization and pruning to reduce computational load (potentially doubling inference speed), leveraging frameworks such as TensorRT or OpenVINO for hardware-specific optimization, and employing scheduling optimizations like frame differencing and asynchronous inference to minimize redundant computations, ensuring stable and efficient system operation under resource constraints.

5.1.3. Simplistic Decision-Making Logic

Rule-based systems, though efficient for structured tasks, exhibit inherent rigidity when confronted with dynamic industrial environments. Their dependency on predefined state machines limits adaptability to unanticipated interactions, such as multi-vehicle conflicts or sensor failures. Such systems rely on predefined rules and state machines, performing tasks such as obstacle avoidance and path planning through conditional triggers. Nevertheless, their rigid reasoning is ill-equipped to accommodate dynamic operating conditions. In multi-vehicle coordination, the inability to evaluate several simultaneous risks often results in conflicts or delays over incidents involving crossing paths, obstructed visibility or vehicle failure. More importantly, manual-coded rule-based systems cannot approximate the extensive number of ambiguous and unstructured problems that frequently arise in industrial contexts. An example can be observed when pedestrians, forklifts and robots are concurrently navigating and crossing paths, where the system frequently misjudges priorities and fails to pursue optimal decisions. Other events occurring outside the set of defined rules, such as equipment failure or communication failure, further limit the system's level of responsiveness. Furthermore, rule-based systems only consider immediate optimality, and do not incorporate or optimize accumulated strategy knowledge relative to feedback or historical data – leading to consistent inflexible decision-making behavior patterns that refuse to change or learn over time, ultimately falling short of the factory's need for flexible and continually optimized decisions.

5.2. Future Research Directions

Table 1. Case Studies of Lightweight Object Detection for Edge Deployment

Case Studies in Domain Adaptation and Self-Supervised Learning	Application Scenario	Model/ Technology	Performance Metrics	Deployment Platform
Vehicle System (RK3588)	Real-time lane and obstacle detection	YOLOv5n	Inference time:31ms	RK3588
UAV Recognition System (Jetson NX)	Aerial building target detection	NanoDet	Stable 720P video stream processing; Quantized model size:2.4MB	Jetson NX
Smart Security Edge Device (Snapdragon 865)	Real-time access control and timed high-precision recognition	NanoDet + YOLOv6n	Multi-thread scheduling avoids single model monopolizing resources	Snapdragon 865

In smart factories, YOLOv5n is vulnerable to lighting, occlusion, and background clutter effects that impact the dependability of unmanned systems. Domain adaptation exists for better cross-scene generalization but is usually limited to unsupervised adversarial training techniques to obviate reliance on annotation. The method uses gradient reversal layers to align feature distributions across the source and target domains, achieving domain-invariant representations and improving efficiency in UAV and vehicle detection applications. Self-supervised learning offers another approach: contrastive learning constructs positive and negative sample pairs to enhance semantic understanding, while masked image modeling pre-trains backbone networks to improve local perception. Table 1 shows the comparison results of case studies regarding the deployment of lightweight object detection models on edge devices. Tests show that on RK3588, the optimized YOLOv5n improves accuracy by 6.2% under complex lighting conditions while maintaining latency below 31ms. Introducing physics-informed neural networks that incorporate kinematic or imaging models as regularization terms can enhance the stability of dynamic detection. For example, adding trajectory prediction error to the loss function improves robustness to occlusion, increasing the mAP of NanoDet on Jetson NX by 4.8% [19]. The integration of multiple technologies significantly enhances the adaptability of YOLOv5n under lightweight constraints, providing stronger perceptual support for smart factory unmanned systems.

5.3. Edge Optimization and Intelligent Decision-Making

To improve YOLOv5n's inference efficiency on edge devices, this study integrates compression technologies (pruning, quantization, NAS) to significantly reduce model size and computational overhead.

Structured pruning (e.g., channel/convolution kernel pruning) eliminates redundant structures. While retaining the backbone architecture, it sharply cuts parameters—model size compressed by >87.5%, speed boosted >3x, accuracy maintained >90%.

Quantization converts floating-point parameters to 8-bit/4-bit integers, greatly reducing storage and computing needs. 8-bit quantization alone shrinks the model to 1/5 its original size, speeds up 4x, and causes <1% accuracy loss; combining with Quantization-Aware Training (QAT) further suppresses performance degradation.

NAS explores efficient network structures via reinforcement learning/gradient optimization. Hardware-aware design balances accuracy, speed, and scale, leading to 40% smaller model size and >2x faster inference post-optimization.

Multiple technologies synergize: pruning+8-bit quantization cuts model size to <1/10 original, boosts speed >5x, and limits accuracy loss to <1.5%; NAS-pruning combination further unlocks compression potential. This multi-dimensional strategy greatly improves YOLOv5n's deployability in resource-constrained scenarios and supports efficient edge inference [20].

6. Conclusion

This study developed a lightweight smart factory unmanned vehicle assistive decision-making system based on YOLOv5n, validating its effectiveness in real-time perception and decision support under resource-constrained edge computing conditions. The system accurately identifies key targets such as pedestrians, forklifts, and shelves in complex factory environments. By integrating LiDAR, millimeter-wave radar, and visual data, it improves perceptual robustness under varying illumination and occlusion conditions. The detection results drive a rule-based decision module combined with a finite-state machine, enabling functions such as dynamic obstacle avoidance and path optimization. Through an intuitive human-machine interface, operators can monitor vehicle status and risk warnings in real time, enhancing system controllability and operational trust. The proposed solution contributes to improved safety and efficiency of unmanned logistics, offering a practical technical pathway for smart factories in the industry 4.0 context.

However, the study also reveals several limitations. The model exhibits insufficient robustness under challenging conditions such as sudden lighting changes, heavy occlusions, and dust interference, leading to occasional false detections. Limited by the coverage of training data, its generalization capability remains constrained. Although designed for edge deployment, computational resources are still strained in high-resolution or multi-task scenarios, especially when performing multi-sensor fusion. The current decision-making mechanism, based primarily on static rules and state machines, struggles to cope with highly dynamic obstacles and unstructured environmental changes. In addition, the system relies heavily on visual input, with perception dimensions still limited. Multi-sensor fusion remains mostly at the data or feature level, with decision-level fusion underexplored. LiDAR and millimeter-wave radar are not fully utilized, which restricts accurate obstacle type identification and impacts path planning. Although YOLOv5n is lightweight, delays and memory bottlenecks occur when processing high-resolution inputs, making it difficult to consistently meet strict real-time requirements.

Future research will focus on enhancing model robustness and environmental adaptability through domain adaptation, self-supervised learning, and physics-aware network designs. Lightweight model optimization—via pruning, quantization, and neural architecture search—will be combined with inference acceleration tools such as TensorRT and OpenVINO to improve execution efficiency on edge devices. Deeper multi-sensor fusion at the feature and decision levels will be explored, using structures such as Transformer for cross-modal alignment, alongside improved calibration and synchronization to boost long-term stability. In terms of decision intelligence, deep reinforcement learning and multi-agent collaboration strategies will be introduced to shift from reactive to proactive planning. Game-theoretic methods and knowledge graphs will further enhance multi-AGV coordination and interpretability. Finally, 5G and V2X integration will facilitate distributed decision-making and long-range situational awareness, supporting the development of more adaptive and reliable unmanned systems for future smart factories.

References

[1] Shadravan A, Parsaei H R. Impacts of Industry 4.0 on Smart Manufacturing [C]. Proceedings of the International Conference on Industrial Engineering and Operations Management, 2023.

[2] Sun M, Lian Z, Yuan Y, et al. Joint Computation and Communication Resources Allocation in Industrial Internet of Things with Edge Computing [Z]. (2023 - 07 - 11).

[3] Santos R C C D M, Coelho M, Oliveira R. Real-time Object Detection Performance Analysis Using YOLOv7 on Edge Devices [J]. IEEE Latin America Transactions, 2024 (10): 22. DOI: 10.1109/TLA.2024.10705971.

[4] Bilik I. Comparative Analysis of Radar and Lidar Technologies for Automotive Applications [J]. IEEE Intelligent Transportation Systems Magazine, 2023. DOI: 10.1109/MITS.2022.3162886.

[5] Huang Y, Zhong Y, Zhong D, et al. Pepper-YOLO: a lightweight model for green pepper detection and picking point localization in complex environments [J]. Frontiers in Plant Science, 2024, 15. DOI: 10.3389/fpls.2024.1508258.

[6] Rahim L A, Abidin N A Z, Aminuddin R, et al. YOLOv5 model-based real-time recyclable waste detection and classification system [C]//Proceedings of the International Conference on Smart City Applications. Cham: Springer, 2024. DOI: 10.1007/978 - 3 - 031 - 53824 - 7_5.

[7] Qiao W, Guo H, Huang E, et al. Real-Time Detection of Slug Flow in Subsea Pipelines by Embedding a Yolo Object Detection Algorithm into Jetson Nano [J]. Journal of Marine Science and Engineering, 2023, 11 (9): 1658. DOI: 10.3390/jmse11091658.

[8] Li H, Zhuang X, Bao S, et al. SCD-YOLO: a lightweight vehicle target detection method based on improved YOLOv5n [J]. Journal of Electronic Imaging, 2024, 33 (2): 023041. DOI: 10.1117/1.JEI.33.2.023041.

[9] Geetha A S. Comparing YOLOv5 Variants for Vehicle Detection: A Performance Analysis [J]. 2024.

[10] Ye Y, Nie Z, Liu X, et al. ROS2 Real-time Performance Optimization and Evaluation [C]. 2023. Chinese Journal of Mechanical Engineering, 2023 (6).

[11] Zhu C, Qian J, Gao M. TensorRT Powered Model for Ultra-Fast Li-Ion Battery Capacity Prediction on Embedded Devices [J]. Energies, 2024, 17 (12): 2797. DOI: 10.3390/en17122797.

[12] NVIDIA introduces Jetson Orin NX embedded computing for artificial intelligence (AI) and unmanned vehicles [J]. Military & Aerospace Fiber Optics, 2023 (3): 17.

[13] Zendehdel N, Chen H, Leu M C. Real-time tool detection in smart manufacturing using You-Only-Look-Once (YOLO)v5 [J]. Manufacturing Letters, 2023, 35: 1052 - 1059. DOI: 10.1016/j.mfglet.2023.08.062.

[14] Liu H, Wan J, Zhou P, et al. Augmented Millimeter Wave Radar and Vision Fusion Simulator for Roadside Perception [J]. Electronics, 2024, 13 (14): 2729. DOI: 10.3390/electronics13142729.

[15] Qian X, Huang Y, Lou P, et al. A Multi-Sensor Fusion-Based Hybrid Guidance Technology for Following AGVs [C]. 2022. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53 (1): 10. DOI: 10.6041/j.issn.1000 - 1298.2022.01.002.

[16] Ding L, Tang Y, Wang T, et al. A Cooperative Decision-Making Approach Based on a Soar Cognitive Architecture for Multi-Unmanned Vehicles [J]. Drones, 2024, 8 (4): 155. DOI: 10.3390/drones8040155.

[17] Sasithong P, Sanguanpuak T, Vanichchanunt P, et al. User Plane Function (UPF) Allocation for C-V2X Network Using Deep Reinforcement Learning [J]. IEEE Access, 2025, 13. DOI: 10.1109/ACCESS.2024.3524886.

[18] Zeng T, Li S, Song Q, et al. Lightweight tomato real-time detection method based on improved YOLO and mobile deployment [J]. Computers and Electronics in Agriculture, 2023. DOI: 10.1016/j.compag.2023.107625.

[19] Long K, Shi X, Li X. Physics-informed neural network for cross-dynamics vehicle trajectory stitching [J]. Transportation Research Part E: Logistics and Transportation Review, 2024, 192: 103799. DOI: 10.1016/j.tre.2024.103799.

[20] Weng Z, Liu K, Zheng Z. Cattle face detection method based on channel pruning YOLOv5 network and mobile deployment [J]. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, 2023, 45 (6): 10003 - 10020.