

Research on Lightweight Environmental Perception Model Based on Deep Convolutional Generative Adversarial Network for Complex Traffic Scenarios

Pinghua Wang^{1*}

¹ School of Automotive Engineering, Hunan Mechanical & Electrical Polytechnic, Changsha, China

*Correspondence: Pinghua Wang

Received: 4 August 2026; **Accepted:** 26 August 2026; **Published:** 9 September 2026

Abstract:

To address the challenges of low recognition accuracy under severe weather conditions, prominent missed and false detection of occluded and small-sized targets, and excessive network parameters that hinder deployment on low-computing-power onboard terminals in mass-produced intelligent connected vehicles, this paper proposes a lightweight environmental perception model based on a deep convolutional generative adversarial network (DCGAN). First, a localized mixed-traffic dataset is constructed, covering complex scenarios such as rain and snow weather, backlighting, and target occlusion. Second, a deep-shallow feature fusion module is incorporated into the generative adversarial network to enhance feature representation for complex targets. Finally, network lightweighting is achieved through parameter pruning and sparsification. Comparative analyses are conducted from four dimensions—precision, recall, miss-detection rate, and inference time—via simulation platform tests and real-vehicle road tests. Experimental results demonstrate that, compared with traditional algorithms such as YOLO-v3 and SSD, the proposed model achieves a 7.62% increase in recognition precision, an 8.35% reduction in miss-detection rate, and a 42.7% decrease in model parameter volume in complex traffic scenarios. The single-frame inference time is significantly shortened, satisfying the real-time requirements of low-cost onboard terminals. Additionally, the model provides algorithmic case support for practical training in intelligent connected vehicle technology programs at higher vocational colleges.

Keywords:

Deep Convolutional Generative Adversarial Network, Environmental Perception, Lightweight, Feature Fusion, Complex Traffic Scenarios, Onboard Perception

1. Introduction

Environmental perception serves as the core prerequisite for intelligent connected vehicles (ICVs) to achieve obstacle detection, road condition assessment, and autonomous decision-making [1]. Current mainstream solutions adopt a multi-sensor fusion architecture integrating cameras, millimeter-wave radars, and LiDAR. Although hardware technologies have matured, software algorithms still face practical

limitations. Under harsh conditions such as rain, snow, haze, and strong backlighting, the signal-to-noise ratio of camera images decreases, and LiDAR is susceptible to interference from suspended particulate matter. On urban mixed-traffic roads, non-motorized vehicles, occluded obstacles, and small traffic targets frequently suffer from missed and false detection [2]. Moreover, existing high-performance perception models contain massive parameters and impose high computational demands on onboard terminals, which impedes their large-scale deployment in ordinary mass-produced civilian vehicles.

Deep convolutional generative adversarial networks can perform sample augmentation and image restoration through generators, effectively alleviating overfitting caused by scarce training samples for extreme scenarios and imbalanced dataset distributions [3]. Nevertheless, most existing studies on generative adversarial networks focus primarily on simulated sample generation, lacking in-depth feature mining for occluded targets, lightweight design oriented toward low-cost onboard hardware, and real-vehicle road test validation. Furthermore, research on practical implementation for vocational education training scenarios remains insufficient.

To bridge these gaps, this paper constructs a localized multi-scenario traffic dataset, improves the generator–discriminator architecture of the deep convolutional generative adversarial network, introduces a deep-shallow feature fusion module, and implements network lightweighting via pruning and sparsification. Multiple sets of experimental data are collected through simulation and real-vehicle tests for quantitative analysis to verify the effectiveness of the improved model in terms of recognition accuracy, anti-interference capability, and real-time inference performance.

2. Dataset Construction and Data Preprocessing

2.1. Multi-scenario Sample Collection

The dataset is composed of field-collected samples and open-source public datasets. Field sampling was conducted on congested urban roads and unpaved suburban roads in Changsha, capturing image samples under various operating conditions, including mixed pedestrian–vehicle traffic, non-motorized vehicle crossing, obstacle occlusion, rain and snow weather, backlighting, and dusk low-light environments. A common onboard monocular camera was adopted for sampling to match the hardware specifications of civilian vehicles.

Six categories of targets were collected: pedestrians, non-motorized vehicles, sedans, trucks, traffic signs, and road barriers. Open-source autonomous driving datasets were screened to supplement the samples.

2.2. Data Annotation, Cleansing and Sample Augmentation

Target bounding-box annotation was completed using the LabelImg tool. Blurred, duplicated, and invalid images with severely missing targets were eliminated through data cleansing. The generator of the deep convolutional generative adversarial network was employed for sample augmentation to generate simulated samples with rain–snow–fog noise, image noise, and partial target occlusion. This approach mitigates sample scarcity and imbalanced distribution under harsh and occluded scenarios. The final dataset contains 12,600 images, which are split into training, validation, and test sets at a ratio of 7:2:1.

3. Design of Improved Deep Convolutional Generative Adversarial Network Model

3.1. Overall Network Architecture

The proposed model is composed of three main components: a generator, a discriminator, and a lightweight detection sub-network. The generator performs image noise reduction, restoration, and sample generation for complex scenarios. The discriminator distinguishes real samples from generated simulated samples. The detection sub-network implements recognition and detection of traffic targets.

Two core improvements are made to the original generative adversarial network. First, a deep-shallow feature fusion module is embedded within the generator. Shallow-layer networks extract low-level texture features such as edges and contours, while deep-layer networks extract high-level semantic features of objects [4]. Feature maps of different scales are concatenated by the feature fusion module to strengthen feature representation for occluded targets and small-sized obstacles, thereby reducing the miss-detection probability. Second, network lightweighting is performed. Parameter pruning and sparsification are applied to the detection sub-network. Convolution kernels with low weight contributions are removed, and some standard convolutions are replaced with lightweight convolutions to reduce the total parameter count and computational overhead during inference, enabling adaptation to low-computing-power onboard terminals [5].

3.2. Loss Function Design

The total loss function combines adversarial loss, detection-classification loss and feature-fusion loss, as expressed below:

$$L_{total} = L_{adv} + \alpha L_{det} + \beta L_{fusion}$$

Where L_{adv} denotes the generation-discrimination loss of the generative adversarial network; L_{det} represents the classification and localization loss for target detection; L_{fusion} is the loss of the feature-fusion module; and α and β are balance weight coefficients. Joint-loss constraints are adopted to balance image generation quality and target-detection accuracy and suppress model over-fitting.

4. Experimental Environment and Evaluation Indicators

4.1. Software and Hardware Environment

PyTorch is selected as the simulation training framework, and an NVIDIA RTX 3090 GPU is used for model training. A civilian-grade onboard computing terminal is adopted for real-vehicle tests. YOLO-v3 and SSD, two widely used industrial algorithms, are selected as comparative baselines.

4.2. Evaluation Indicators

Five core quantitative indicators in target-detection domain are adopted:

(1) Precision (P): the proportion of true-positive samples among all predicted positive samples;

(2) Recall rate (R): the proportion of successfully detected real targets among all ground-truth targets;

- (3) Miss- detection rate (Miss): the proportion of real targets that fail to be detected;
- (4) Single- frame inference time (t): an indicator for evaluating model real-time performance, measured in milliseconds;
- (5) Model parameter quantity (Params), measured in millions (M).

5. Experimental Results and Data Analysis

5.1. Comparative Tests on Simulation Datasets

The test set was fed into the three models. The indicators under simulation conditions are summarized in Table 1.

Table 1. Performance comparison of different algorithms in simulation tests.

Model	Precision P /%	Recall R /%	Miss- detection rate Miss /%	Params /M	Single- frame inference time t /ms
SSD	78.13	75.26	14.71	26.4	47.3
YOLO- v3	82.47	80.15	11.24	61.2	32.6
Proposed model	90.09	87.81	2.89	35.1	21.4

From the simulation data, the following conclusions can be drawn: (1) The proposed model outperforms SSD and YOLO-v3 in both precision and recall. Its miss-detection rate drops by 8.35 percentage points compared with YOLO-v3, verifying that the feature-fusion module effectively improves recognition performance for occluded and small-sized targets. (2) Benefiting from lightweight pruning, the parameter volume decreases by 42.7% compared with YOLO-v3. The single-frame inference time drops from 32.6 ms to 21.4 ms, achieving a significant improvement in real-time performance.

5.2. Comparative Analysis on Harsh- Condition Sub-datasets

Sub-datasets covering rain–snow–fog weather, strong backlighting, and target occlusion were extracted from the test set to simulate harsh traffic environments. The recall rate comparison of the three models is listed in Table 2.

Table 2. Recall- rate comparison under harsh- scenario sub-datasets (Unit: %).

Model	Rain- snow- fog scenario	Strong backlight scenario	Target- occlusion scenario
SSD	64.32	62.75	60.11
YOLO- v3	70.56	68.23	66.47
Proposed model	82.47	81.32	79.64

Experimental results reveal that the recall rates of traditional algorithms decline significantly under harsh environments. Benefiting from image restoration and sample augmentation by the generative adversarial network, the proposed model maintains relatively high recall rates under rain–snow–fog, strong backlight, and occlusion conditions, which proves that the DCGAN can effectively improve perception performance under adverse environments.

5.3. Real- Vehicle Road Test Results

The proposed model was deployed on an onboard terminal. A total of 2,000 valid real-vehicle image frames collected on urban and suburban roads were statistically analyzed, as shown in Table 3.

Table 3. Real-vehicle road test results.

Model	Precision P /%	Miss- detection rate Miss /%	Single- frame inference time t /ms
YOLO- v3	77.31	13.62	45.7
Proposed model	84.53	5.17	28.6

Real-vehicle tests indicate that, affected by real-world road noise and lens jitter, performance degradation can be observed for all models compared with simulation results. Nevertheless, the proposed model still maintains high recognition accuracy. Its inference time satisfies the real-time requirements for onboard perception (generally less than 50 ms per frame), enabling stable operation on low-cost onboard hardware terminals.

5.4. Error Analysis

Error samples from the real-vehicle tests were statistically analyzed. Two main error sources are summarized: (1) complete image overexposure caused by extreme strong light, leading to nearly total loss of target information; and (2) extremely distant small-sized targets with very low pixel proportions in the images. Further expansion of extreme-scenario samples will be conducted to optimize model robustness in follow-up research.

6. Discussion on Model Application in Teaching

The codes, datasets, and test cases of the improved model can be transformed into practical teaching resources for the intelligent connected vehicle technology major. In classroom practice, students can complete operations including dataset reading, model inference, and indicator calculation, thereby gaining an intuitive understanding of generative adversarial networks, model lightweighting methods, and evaluation metrics for target detection. This approach compensates for the deficiency that traditional teaching only adopts generic open-source models without localized improved cases, achieving an effective transformation from research achievements to teaching resources.

7. Conclusions

Aiming at complex mixed pedestrian–vehicle traffic scenarios in China, this paper constructs a deep convolutional generative adversarial network embedded with a deep-shallow feature fusion module and implements network lightweighting. Multiple groups of comparative experiments and quantitative data analyses are carried out on simulation datasets, harsh-condition sub-datasets, and real-vehicle road tests. The following conclusions are drawn: (1) The DCGAN with an embedded feature-fusion module can effectively boost environmental perception accuracy for occluded targets under harsh weather and reduce the miss-detection rate. (2) Lightweighting via parameter pruning and sparsification greatly reduces model parameters and inference time, enabling the model to adapt to low-computing-power onboard terminals of ordinary mass-produced civilian vehicles. (3) Performance degradation exists in real-vehicle tests compared with simulation results. Extreme-scenario samples will be further expanded to improve model robustness. Meanwhile, the experimental cases will be converted into teaching resources to support talent cultivation in the intelligent connected vehicle specialty group.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this article.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

- [1] Liu, Y.Q. Construction of anomaly detection model based on multi-sensor fusion. *Southern Metals*, 2026, 04, 84-85, 89.
- [2] Chen, Z.; Zheng, C. Research on intelligent agent for locomotive operating environment perception based on multimodal large model. *Modern Information Technology*, 2026, 10(05), 106-111.
- [3] Zhai, Z.P.; Shao, J.J. Small road target detection in complex environment based on visual feature guidance. *Chinese Journal of Engineering Science*, 2026, 48(04), 878-890.
- [4] Peng, Z.R.; Yang, X.Y. Energy consumption prediction model for reinforcement learning integrating multi-source environmental perception. *Control and Decision*, 2026, 41(06), 1731-1742.
- [5] Zhou, W.; Li, X.B. Intelligent perception and parameter optimization method for wireless channel environment based on large model. *Electronic Components and Information Technology*, 2026, 10(01), 150-153.



© 2026 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)