Engineering Technology Trends Vol 4 Issue 4 2026

ORIGINAL RESEARCH ARTICLE

Open Access

Design of a Lightweight Road Pedestrian Detection System for Complex Environments

Yong Tang*, Jun-Hao Zhang, Miao Sun

Hainan Vocational University of Science and Technology, Haikou, Hainan Province, 571126, China

***Correspondence to:** Yong Tang, Hainan Vocational University of Science and Technology, Haikou, Hainan Province, 571126, China, Email: 18907582886@189.cn

Abstract: To address the challenges posed by complex outdoor operating conditions—such as poor visibility, occlusions from rain gear, glare on water surfaces, and cluttered backgrounds—which cause traditional pedestrian detection models to suffer from parameter redundancy, a high deployment threshold, and imbalanced detection accuracy, this paper presents a lightweight road pedestrian detection system for complex environments, with a core focus on the deployment requirements for embedded terminals. By strategically selecting a baseline model, optimizing the feature extraction module, and streamlining the network architecture, and by incorporating strategies for rain-fog noise suppression and rain-glare reflection, the system strikes a balance between its computational load and recognition performance. The optimized system effectively adapts to a variety of complex road environments, providing technical support for pedestrian detection tasks on mobile devices under adverse weather conditions.

Keywords: Harsh lighting conditions; stormy weather conditions; lightweight network; object detection; model selection

Introduction

In the field of intelligent perception, pedestrian detection is a critical foundational task, frequently utilized in applications such as intelligent traffic management and vehicle environmental perception. Current mainstream detection solutions predominantly rely on deep, large-scale networks. While these models offer excellent feature resolution, they are often too complex and computationally expensive for resource-constrained terminal devices. Furthermore, their detection stability can be compromised by adverse conditions, including challenging lighting, standing water, and occlusions from rain gear. This paper addresses these challenges by selecting a suitable network architecture and pruning redundant parameters to build a lightweight detection model, thereby overcoming the various obstacles to its real-world field deployment.

1. Characteristics of Complex Environments and System Design Requirements for Road Pedestrian Detection

The diverse and variable real-world outdoor road conditions pose significant challenges to the practical deployment of lightweight pedestrian detection technology. Before setting up the complete detection system, it is crucial to first systematically identify

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

the specific characteristics of various environmental interferences within the operational scenarios, and then finalize the corresponding design specifications for the entire system. The interferences in road scenarios can be broadly classified into two main categories: those arising from the external environment and those inherent to the pedestrian targets themselves. External interferences include heavy precipitation during rainstorms, which reduces visibility; strong glare from puddles on the road surface; and wind-driven airborne debris. Such conditions can significantly degrade the image quality and introduce artifacts like rain streaks and fog, which obscure the target's outline. Target-inherent interferences include factors such as a pedestrian's umbrella obstructing their torso, a raincoat altering their silhouette, mutual occlusion in crowded rainy-day scenes, and the loss of detail in distant pedestrians through rain and fog.

From a practical application standpoint, the small-scale, on-device hardware deployed on roadsides typically has limited computational power, which constrains its storage capacity, processing speed, and battery life. Consequently, conventional deep neural network (DNN) models cannot be directly deployed and executed on these devices. Therefore, our approach is to simplify the overall network architecture, prune redundant model parameters, and reduce the computational resource consumption for image recognition, all while ensuring detection accuracy in heavy rain conditions. The pedestrian detection system developed in this study must strike a balance among three key criteria: adaptability to adverse environments, a high degree of model efficiency, and high-speed image processing. This is to establish a dedicated detection architecture tailored for road conditions during torrential rain [1].

2. Analysis of Lightweight Pedestrian Detection Techniques

2.1 Mainstream Lightweight Neural Network Architectures

At present, the core optimization strategies for lightweight network architectures in the field of object detection are centered on innovating convolutional modes and decoupling channel structures. Depthwise separable convolution architectures decouple the channel fusion and spatial feature extraction steps of standard convolutions, drastically reducing the parameter count. While widely employed for basic recognition tasks on mobile devices, they exhibit limitations in their capacity for fine-grained feature extraction, particularly under complex conditions such as heavy rain. Channel-shuffling-based networks enhance the efficiency of feature fusion by scrambling the channel distribution of feature maps. With negligible computational overhead, they are well-suited for ultra-low-power devices; however, they are prone to feature discontinuity when faced with occlusion from objects like rain gear. Inverted Residual Networks (IRNs), leveraging inverted residual blocks, expand the channel capacity for shallow-layer features, striking a balance between model accuracy and computational volume. Consequently, they offer the most well-rounded overall performance, making them the preferred architecture for lightweight detection tasks in extreme scenarios at present.

2.2 Model Selection for Detection in Complex Environments

In the process of constructing a lightweight network, the selection of a baseline detection network is pivotal. Under adverse road conditions, such as heavy rain, pedestrians are subject to various interferences, including occlusion by rain streaks, distortion of their silhouette, and variations in target size. Consequently, the detection network must possess fundamental capabilities, such as multi-scale feature fusion and rapid inference. A comparative analysis was conducted across several YOLO-based frameworks. Two-stage networks, such as Faster R-CNN, suffer from a high volume of redundant computations. Early single-stage networks, like YOLOv3, exhibited a significant number of missed detections when identifying small, overlapping pedestrians. While YOLOv8 demonstrated improved detection accuracy, its C2f module consumes more memory, making its deployment on low-computational-power devices more challenging than that of YOLOv5. Furthermore, YOLOv12 achieved superior detection performance; however, its attention module involves a complex computational process, rendering it unsuitable for the real-time detection requirements of edge terminals. In contrast, YOLOv5s was found to strike an optimal balance among parameter size, detection accuracy, and hardware compatibility. Its network architecture is well-suited for the operating conditions

on roads during heavy rain. This network was therefore designated as the baseline model. The backbone was subsequently reconfigured for a lightweight design by restructuring it with inverted residual modules. Serving as the foundation for optimization, these inverted residual modules reduce the number of parameters and enhance the model's robustness against environmental interferences, thereby ensuring stable, low-latency pedestrian detection on computationally constrained devices.

2.3 Image Preprocessing Techniques for Complex Environments

Raw footage captured in adverse road scenarios, such as under intense sun exposure or heavy rain, often suffers from issues like imbalanced exposure and widespread rain-induced noise and artifacts. Feeding these raw images directly into a neural network would not only increase the computational load but also degrade the accuracy of target object recognition. An adaptive algorithm for rain, haze, and noise reduction is employed to counteract two primary types of interference: rain streaks and road glare. The intensity of the filtering is automatically adjusted based on the density of rain streaks and fog across different regions of the image, allowing for the effective removal of various types of noise while preserving key features, such as the silhouettes of pedestrians and the contours of vehicles. A multi-scale color balancing technique is utilized to address issues like localized overexposure and the loss of detail in shadows, which are caused by glare from puddles and reflections. Brightness and contrast are adjusted on a per-layer basis, which significantly suppresses road glare. Consequently, the overall sharpness and the integrity of the useful information in the image are enhanced. Detection fluctuations caused by variations in target object size can be mitigated, and the entire system's data processing speed and recognition stability are concurrently improved [2].

2.4 Core Strategies for Model Lightweighting and Optimization

Network lightweighting optimization involves more than simply reducing the number of network layers. A comprehensive set of streamlining techniques eliminates redundant modules at a manageable and controlled trade-off in accuracy, thereby striking a balance between the model's computational speed and recognition performance. This makes the model suitable for deployment in resource-constrained environments, such as mobile and edge devices. The three mainstream optimization approaches are parameter pruning, feature knowledge transfer, and network architecture restructuring. Each of these approaches offers distinct advantages and can be synergistically combined. Parameter pruning removes nodes with minimal impact to compress the model's size; common techniques include weight thresholding and gradient-based iterative pruning. However, excessive pruning can disrupt feature pathways, leading to a significant drop in accuracy, which can be mitigated through re-training and gradual fine-tuning. Feature knowledge transfer involves distilling the learned feature patterns from a large-scale model into a smaller one, leveraging techniques like soft-target labels and intermediate-layer features to facilitate this knowledge transfer. This method increases the training overhead and its effectiveness is constrained by the base model and the specific transfer strategy employed. Structural restructuring re-engineers the feature processing modules by replacing resource-intensive standard convolutions with more efficient alternatives, such as depthwise separable convolutions and bottleneck modules. This approach achieves a lightweight transformation at the foundational architecture level, thereby reducing the computational burden and enhancing the model's overall adaptability. As the primary optimization method for this system, this approach, when combined with automated network architecture search (NAS) and hardware-aware design, significantly boosts overall energy efficiency.

3. Overall Design of a Lightweight Pedestrian Detection System for Complex Environments

3.1 System Overall Architecture Design

The entire system is structured in a four-layer architecture: image acquisition, preprocessing, feature detection, and result output. Each layer can operate independently while also being logically interconnected, which simplifies module updates, iterations, and hardware-software debugging. On-board and roadside surveillance cameras handle the image acquisition, with the capacity for multi-channel video signals to be synchronously integrated. The acquisition

frame rate and resolution automatically adjust based on lighting and environmental conditions such as rain and snow, ensuring the data quality and real-time performance of the raw images. The preprocessing layer integrates adaptive functions for rain and fog removal, color balancing, and image standardization. Its algorithms automatically identify and filter out various types of visual interference, effectively removing noise while enhancing the distinctiveness of target features. This results in clean, well-defined image data being passed to the detection stage. The system's core is the feature detection layer, which is built upon an optimized YOLOv5s network. Its architecture is primarily composed of a lightweight backbone, feature enhancement, and a detection and classification module. The backbone network maintains recognition accuracy while reducing computational overhead. Multi-scale fusion is leveraged by the feature enhancement module to improve the detection of small-sized and occluded pedestrians. An attention mechanism is embedded within the detection module to boost the precision of target classification and localization. The output layer uniformly presents visualized information, including object coordinates, count, type, and confidence score. It is fully compatible with functions such as bounding box annotation, real-time statistics, and trajectory plotting. The system can adapt to various terminal data formats and fulfills all requirements for data integration and visualization across different platforms [3].

3.2 Optimized Design of a Lightweight Backbone Network

To adapt to complex road conditions and the operational constraints of low-performance edge devices, this study builds upon the YOLOv5s model by introducing an inverted residual structure into the backbone network for lightweight optimization, achieving an effective balance between computational efficiency and detection accuracy while meeting the deployment requirements of edge computing scenarios. The residual unit structure is prioritized for reconstruction, with high-dimensional redundant feature channels reasonably pruned to significantly reduce memory usage and floating-point computation. A channel attention screening module is integrated into the network architecture, which autonomously assesses the importance of each feature channel, dynamically suppressing the weight of background interference features such as rain streaks and floating debris, while simultaneously enhancing pedestrian target feature information, thereby improving the model's feature discrimination performance in complex interference scenarios. The network hierarchy is further optimized by removing deep, redundant, and ineffective convolutional layers as a whole, effectively mitigating overfitting and computational resource waste. Original large-size convolutional layers are decomposed into lightweight units combining depthwise separable convolutions and point convolutions, with layer-by-layer feature extraction further compressing model parameters and storage size. To address the issue of missed detection of small targets in rainy and foggy scenes, a multi-scale feature fusion branch is added, integrating a feature pyramid and cross-layer connections to merge shallow high-resolution features with deep semantic features, enhancing feature extraction for distant small pedestrians and alleviating missed detections caused by scale variations. By adopting a bidirectional strategy combining network simplification and feature enhancement, the model improves weak target recognition performance while controlling inference latency and computational overhead. Ablation experiments validate the effectiveness of each module, demonstrating that the optimized network achieves high detection accuracy and recall rates in complex scene datasets [4].

3.3 Anti-Interference Module Design for Complex Environments

To address the detection challenges posed by adverse road conditions, such as heavy rain and fog, a dedicated anti-interference module is embedded within the model's feature detection layer. This module is comprehensively adapted for visual detection tasks in harsh weather, and the entire architecture is composed of three synergistic sub-modules. The first is the Dynamic Rain/Fog and Glare Correction Sub-module, which utilizes data on the mean image grayscale and pixel variance to assess ambient lighting conditions and the intensity of road surface glare in real-time. Consequently, parameters for rain/fog removal and glare suppression can be adaptively adjusted, effectively mitigating detection failures caused by reflections from puddles and dim lighting. The second is the Local Feature Completion Sub-module, which fuses basic contour features with target motion information. This

allows for the precise prediction of the distribution patterns of occluded areas, such as those caused by rain gear or being enveloped by a raincoat. Prior knowledge of human pose is leveraged to fill in missing limb features, thereby significantly reducing the probability of missed detections of pedestrians caused by occlusion. The third is the Spatial Attention Filtering Sub-module, which employs time-series dynamic data to pinpoint regions of high-frequency motion targets. The weights of interfering features, such as rain streaks and flying debris, as well as the static background, are automatically down-weighted, leading to a significant enhancement in the spatial-level separation of target and background information. Working in concert, these three sub-modules—from the dimensions of image quality optimization, missing feature restoration, and background noise filtering—effectively counteract the adverse interference of harsh environments on the detection system, comprehensively enhancing the stability and reliability of the entire detection process.

3.4 Embedded Terminal Adaptation and Optimization

To address the incompatibility of lightweight models with embedded terminals, optimizations have been implemented from three perspectives: model encapsulation, compute scheduling, and power consumption management. At the model encapsulation level, redundant configurations and inactive nodes are stripped out, and the terminal adaptation interface is unified to ensure compatibility across diverse hardware platforms. At the compute scheduling level, a dynamic allocation mechanism is designed: under low load, the full detection mode is activated; under high load, non-core feature branches are streamlined to prioritize and guarantee baseline detection. At the power consumption management level, the system coordinates with the data acquisition (DAQ) layer to dynamically adjust the frame rate; it lowers the sampling frequency in static scenes and reverts to a high-frequency mode in high-risk, high-storm scenarios, thereby reducing the terminal's power consumption at the source.

Conclusion

Addressing the specific interference characteristics of road scenarios during heavy rain and the practical deployment challenges posed by the limited computational power of edge devices, this paper presents the overall design of a lightweight pedestrian detection system. By leveraging strategic improvements—such as judicious model selection, network architecture optimization, and the integration of an interference-robust module—this system effectively mitigates the shortcomings of conventional models, namely their resource-intensive nature, poor robustness against interference, and difficulty in deployment. Future research can be extended to encompass extreme scenarios, such as sandstorms and high winds, with a focus on continuously optimizing feature extraction and enhancing the system's overall comprehensive performance.

References

[1] Li Zhenlu, Huang Wei, Sun Kai. A Lightweight Road Object Recognition Algorithm for Complex Environments[J]. Computer Engineering, 2024, 50(4): 219-227.

[2] Li Chunhua, Wu Zeyu, Luo Ying. A Lightweight Real-time Road Traffic Monitoring System[J]. Fujian Computer, 2025, 41(12): 78-83.

[3] zhang, yinhui, zhang, pengcheng, & he, zifen. (2022). Lightweight real-time detection of infrared pedestrian targets with fine-scale embedding. *Chinese Journal of Photonics*, *51*(9), 258-268.

[4] Li Yongtan, Zhang Jinwei, Wang Junjie. An Improved Lightweight YOLOv5s-based Object Detection Method for Vehicles and Pedestrians[J]. Journal of Shandong Normal University (Natural Science), 2025, 40(1): 71-80.