ISSN 2277-2685
IJESR/April. 2026/ Vol-16/Issue-2s1/256-260
Munaf Yezdani *et. al.*, /International Journal of Engineering & Science Research

Full Length Article

Enhanced Yolov11m For Real-Time Multi-Scale Traffic Detection Under Haze Conditions

Munaf Yezdani[1], Abid Sohail Bin Mazi[2], Abdul Rafe Obaid[3], Dr. K. Upendra Babu[4], Mr.Syed Shah Mehmood Sarmast[5]
[1,2,3]B.E.Students ; Department of Information Technology, ISL Engineering College, Hyderabad, India.
[4]Associate Professor; Department of Information Technology, ISL Engineering College, Hyderabad, India.
[5]Head of the Dept; Department of Information Technology, ISL Engineering College, Hyderabad, India.
Mail Id; manafyazdani100@gmail.com, abidmazi14@gmail.com, abdulrafeobaid@gmail.com
Accepted 27-04-2026

Abstract
Traffic object detection is a fundamental component of intelligent transportation systems, autonomous vehicles, and smart surveillance applications. However, environmental conditions such as haze and fog significantly reduce image visibility, weaken object boundaries, and introduce background noise, thereby decreasing the performance of traditional object detection algorithms. Lightweight object detectors are suitable for real-time deployment on edge devices, but they often struggle to maintain detection accuracy under adverse weather conditions due to limited contextual understanding and insufficient multi-scale feature representation. This project proposes an enhanced YOLOv11m framework for real-time multi-scale traffic detection under haze conditions. The proposed model introduces three major architectural improvements: Attention-Gate Convolution (AGConv), Multi-Dilation Sharing Convolution (MDSC), and Cross-Channel Feature Fusion Module (CCFM). AGConv improves spatial attention and suppresses irrelevant background information, MDSC enhances receptive field diversity for better multi-scale feature extraction, and CCFM dynamically recalibrates channel-wise feature importance for efficient feature fusion. The model is trained and evaluated on traffic datasets containing foggy and hazy scenes using standard performance metrics including mean Average Precision (mAP), Intersection over Union (IoU), and Frames Per Second (FPS). Experimental results demonstrate that the proposed YOLOv11m achieves improved detection accuracy compared with YOLOv11n while maintaining lightweight architecture and real-time processing capability. The system achieves approximately 376 FPS with only 2.6 million parameters while improving mAP@0.5 and mAP@0.5:0.95 under challenging weather conditions. The proposed framework provides an efficient and practical solution for intelligent transportation systems operating in degraded visual environments.

Keywords
YOLOv11m, Intelligent Transportation Systems, Traffic Object Detection, Haze Detection, Multi-Scale Feature Extraction, Deep Learning, Computer Vision, AGConv, MDSC, CCFM, Real-Time Detection, Autonomous Driving.

Introduction

Intelligent transportation systems (ITS) have become one of the most important research areas in modern smart city development. The rapid growth of urban populations and the increasing number of vehicles on roads have created significant challenges in traffic management, road safety, and transportation efficiency. To overcome these issues, intelligent traffic monitoring systems are widely used for automatic vehicle detection, pedestrian monitoring, accident prevention, and traffic flow analysis. Computer vision and deep learning technologies have become essential tools in achieving these objectives because they provide accurate and real-time analysis of traffic scenes.

Object detection is a core component of intelligent traffic systems. It involves identifying and locating vehicles, pedestrians, cyclists, traffic signs, and other objects within images or video frames. In recent years, deep learning-based object detection methods have significantly improved detection accuracy and speed. Among these methods, the YOLO (You Only Look Once) family of object detection models has gained widespread popularity because of its balance between real-time performance and detection accuracy. YOLO models perform object localization and classification simultaneously in a single forward pass, making them highly efficient for practical applications.
Although YOLO-based systems perform effectively in clear weather conditions, their performance

degrades significantly in adverse environmental conditions such as haze, fog, rain, smoke, and low illumination. Haze is one of the most challenging weather conditions for computer vision systems because it reduces image contrast, blurs object boundaries, and weakens texture information. As a result, the visibility of distant and small objects becomes severely limited. In intelligent transportation systems, such limitations can lead to missed detections, false alarms, and unsafe driving decisions.

Lightweight models such as YOLOv11n are specifically designed for real-time deployment on embedded devices and edge computing platforms. However, these lightweight architectures often contain fewer parameters and reduced feature extraction capability, which limits their robustness in degraded environments. Traditional convolution operations may fail to extract sufficient contextual information from hazy images, leading to poor object representation and inaccurate predictions.

To address these problems, this project proposes an enhanced YOLOv11m model for real-time multi-scale traffic detection under haze conditions. The proposed architecture introduces three specialized modules designed to improve feature extraction, contextual awareness, and feature fusion efficiency. The Attention-Gate Convolution (AGConv) module enhances spatial attention by focusing on important regions while suppressing irrelevant background noise. The Multi-Dilation Sharing Convolution (MDSC) module captures both local and global features using multiple dilation rates, thereby improving multi-scale representation. Additionally, the Cross-Channel Feature Fusion Module (CCFM) strengthens channel-wise feature interaction and improves semantic feature integration.

The proposed YOLOv11m framework aims to improve detection accuracy while preserving the lightweight and real-time characteristics required for edge deployment. Experimental evaluation demonstrates that the model achieves superior performance under haze conditions while maintaining computational efficiency. These advantages make the proposed system highly suitable for autonomous driving, traffic surveillance, and intelligent road safety applications.

Challenges in Hazy Traffic Detection

Traffic object detection in adverse weather conditions is a highly challenging task because environmental degradations significantly affect image quality and feature visibility. Haze and fog reduce scene clarity and create low-contrast environments, making object detection more difficult than in normal conditions.

One major challenge in hazy traffic detection is reduced visibility. Haze particles scatter light in the atmosphere, causing distant objects to appear blurred and faded. This scattering effect weakens important visual features such as edges, textures, and color information. As a result, deep learning models may fail to extract discriminative features required for accurate detection.

Another important challenge is the detection of small and distant objects. In urban traffic scenes, pedestrians, bicycles, motorcycles, and distant vehicles occupy only a small number of pixels. Under haze conditions, these objects become even harder to recognize because their visual characteristics are partially lost. Traditional convolutional layers often fail to preserve such fine-grained information during feature extraction.

Traffic scenes also contain significant scale variations. Large objects such as buses and trucks may appear close to the camera, while pedestrians and small vehicles may appear at far distances. Effective traffic monitoring therefore requires strong multi-scale feature representation. Lightweight models frequently struggle to capture contextual information across multiple scales due to limited receptive field diversity.

Background clutter and noise introduce additional complexity. Urban environments contain buildings, road signs, trees, shadows, and illumination variations that can confuse detection systems. Hazy conditions further increase ambiguity between foreground objects and the background. Consequently, object detectors may generate false positives or inaccurate localization results.

Another important issue is computational efficiency. Real-time traffic monitoring systems are typically deployed on embedded platforms, roadside cameras, drones, or edge computing devices with limited hardware resources. Therefore, object detection models must achieve high accuracy while maintaining low memory usage and fast inference speed. Balancing these requirements remains a difficult research problem.

These challenges highlight the need for advanced feature extraction and feature fusion techniques capable of preserving fine details, improving contextual understanding, and maintaining lightweight computational complexity under degraded visual environments.

Existing YOLOv11n Framework

YOLOv11n is a lightweight object detection framework designed for real-time applications. The architecture follows the traditional YOLO pipeline consisting of three major components: the backbone, neck, and detection head.

The backbone network is responsible for extracting hierarchical visual features from input images. Early convolutional layers capture low-level features such as edges, corners, and textures, while deeper layers extract semantic representations related to object categories. Lightweight backbones reduce

computational complexity and improve inference speed, making them suitable for edge devices.

The neck network combines feature maps from different stages of the backbone using feature fusion mechanisms such as feature pyramids. This process enables the model to detect objects at multiple scales by integrating both high-resolution spatial features and deep semantic features.

The detection head performs final object localization and classification. It predicts bounding box coordinates, object confidence scores, and class probabilities for each detected object. YOLO-based detection heads are highly efficient because they perform these tasks simultaneously in a single forward pass.

Despite its efficiency, YOLOv11n suffers from several limitations under haze conditions. The lightweight architecture reduces the model's ability to extract strong contextual information from degraded images. Pooling operations may also remove fine-grained features required for detecting small and distant objects. Furthermore, conventional convolution layers may not effectively suppress haze-related noise or adapt to varying object scales. As a result, YOLOv11n may produce inaccurate predictions, missed detections, and reduced robustness in adverse weather conditions. These limitations motivate the development of the enhanced YOLOv11m architecture.

Proposed YOLOv11m Model

The proposed YOLOv11m architecture introduces several improvements designed specifically for haze-aware traffic object detection. The model enhances feature extraction, contextual representation, and multi-scale feature fusion while maintaining lightweight computational complexity.

The first enhancement is the Attention-Gate Convolution (AGConv) module. This module introduces an attention mechanism into the convolution operation to focus on important spatial regions while suppressing irrelevant background information. Hazy traffic scenes often contain noisy and low-contrast regions that can confuse the network. AGConv dynamically assigns higher importance to informative areas containing vehicles or pedestrians while reducing the influence of haze-distorted regions. This selective attention mechanism improves feature quality and enhances low-visibility object detection.

The second enhancement is the Multi-Dilation Sharing Convolution (MDSC) module. Conventional convolutions use fixed receptive fields, which may not effectively capture objects of different sizes. MDSC solves this problem by incorporating multiple dilation rates within shared convolution kernels. Small dilation rates capture local details, while larger dilation rates capture global contextual information. This multi-scale feature extraction mechanism improves the detection of both nearby and distant objects while preserving computational efficiency.

The third enhancement is the Cross-Channel Feature Fusion Module (CCFM). This module dynamically recalibrates feature importance across channels and strengthens feature interaction between different layers. By emphasizing important semantic information and suppressing redundant features, CCFM improves overall feature representation. The module also enhances the integration of low-level spatial details with high-level semantic information, resulting in more accurate localization and classification.

The proposed YOLOv11m model replaces conventional bottleneck structures with these specialized modules. Despite the architectural enhancements, the model maintains lightweight design and real-time inference speed, making it suitable for deployment on embedded systems and smart surveillance devices.

Methodology

The methodology of the proposed system includes multiple stages such as dataset collection, annotation, preprocessing, model training, and performance evaluation.

The first stage involves collecting traffic datasets captured under hazy and foggy conditions. Both publicly available datasets and real-world traffic surveillance footage are used to ensure diversity in environmental conditions and traffic scenarios. The datasets include various traffic objects such as cars, buses, trucks, motorcycles, bicycles, and pedestrians.

After data collection, bounding box annotations are created for all traffic objects. Annotation tools are used to label object categories and spatial coordinates accurately. Proper annotation is essential for supervised learning because it provides the ground truth required for training the object detection model.

Image preprocessing techniques are then applied to improve data quality and model generalization. These techniques include image resizing, contrast enhancement, normalization, and noise reduction. Data augmentation methods such as rotation, scaling, flipping, brightness adjustment, and random cropping are also applied to increase dataset diversity and improve robustness against varying environmental conditions.

The YOLOv11m model is trained using optimized hyperparameters including learning rate, batch size, optimizer selection, and epoch count. Transfer learning techniques may also be used to accelerate convergence and improve feature learning.

The model is evaluated using standard object detection metrics such as mean Average Precision (mAP), Intersection over Union (IoU), precision,

recall, and Frames Per Second (FPS). These metrics provide comprehensive evaluation of detection accuracy, localization performance, and real-time capability.

Experimental Results

Experimental analysis demonstrates that the proposed YOLOv11m model achieves superior performance compared with YOLOv11n under haze conditions. The integration of AGConv, MDSC, and CCFM modules significantly improves contextual understanding and feature representation.

The proposed model improves mAP@0.5 by approximately 1.1% and mAP@0.5:0.95 by approximately 2.7%. These improvements indicate stronger localization accuracy and more reliable object classification under degraded visual environments.

Despite the architectural enhancements, the model maintains high inference speed with approximately 376 FPS and only 2.6 million parameters. This demonstrates that the proposed framework successfully balances detection accuracy and computational efficiency.

The AGConv module effectively suppresses irrelevant background noise and enhances low-visibility object features. As a result, the model achieves improved detection performance for pedestrians and distant vehicles. The MDSC module enhances receptive field diversity and improves multi-scale feature extraction, enabling more accurate detection of objects with varying sizes. Meanwhile, the CCFM module strengthens channel-wise feature interaction and improves semantic feature fusion.

The proposed model also shows strong robustness in dense urban traffic scenes with heavy haze and cluttered backgrounds. False detections are reduced significantly compared with conventional lightweight detectors.

Applications

The proposed YOLOv11m framework has broad applications in intelligent transportation systems and smart city infrastructure. One major application is real-time traffic monitoring. The system can automatically detect and track vehicles, monitor traffic congestion, and analyze traffic flow patterns. Such capabilities help improve transportation efficiency and reduce road accidents.

Another important application is autonomous driving. Self-driving vehicles require reliable object detection systems capable of operating under adverse weather conditions. The proposed model improves environmental perception and enhances road safety by accurately detecting vehicles, pedestrians, and obstacles in hazy environments.

The framework is also suitable for Advanced Driver Assistance Systems (ADAS). Applications such as collision warning, pedestrian detection, lane monitoring, and adaptive cruise control depend heavily on robust object detection performance.

Because the proposed architecture is lightweight, it can be deployed on embedded systems, smart surveillance cameras, roadside monitoring devices, and edge computing platforms. This makes it highly practical for large-scale smart transportation infrastructure.

Future Enhancements

Several future improvements can further enhance the proposed system. One possible enhancement is the integration of image dehazing algorithms before object detection. Dehazing networks based on generative adversarial networks (GANs) or transformer architectures may improve image clarity and enhance feature visibility.

Temporal video analysis can also improve detection consistency across consecutive frames. By incorporating temporal information, future systems may achieve more stable tracking and reduce detection fluctuations.

Additional optimization techniques such as model pruning, quantization, and knowledge distillation can further reduce memory usage and computational requirements, enabling deployment on extremely resource-constrained devices.

Future research may also extend the framework to handle additional weather conditions such as rain, snow, and nighttime environments. Developing a generalized multi-weather traffic detection system would significantly improve real-world applicability.

Conclusion

The proposed YOLOv11m model provides an efficient and robust solution for real-time multi-scale traffic detection under haze conditions. By integrating AGConv, MDSC, and CCFM modules, the architecture improves contextual understanding, feature extraction capability, and adaptive feature fusion.

Experimental results demonstrate that the model achieves improved detection accuracy while maintaining lightweight architecture and real-time inference speed. The proposed framework successfully balances computational efficiency and robustness, making it highly suitable for intelligent transportation systems operating in degraded visual environments.

Its ability to detect small and distant objects under haze conditions makes it valuable for autonomous driving, smart surveillance, and road safety applications. Future enhancements involving image dehazing, temporal analysis, and multi-weather adaptation can further strengthen system performance and broaden practical usability.

References

1) J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.
2) A. Bochkovskiy, C. Y. Wang, and H. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv preprint arXiv:2004.10934, 2020.
3) J. Chen et al., "Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks," Proceedings of CVPR, 2023.
4) S. Park et al., "PConv: Simple Yet Effective Convolutional Layer for GAN," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.
5) H. Zhang et al., "DINO: DETR with Improved Denoising Anchor Boxes for End-to-End Object Detection," International Conference on Learning Representations, 2022.
6) K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Proceedings of CVPR, 2016.
7) T. Lin et al., "Feature Pyramid Networks for Object Detection," Proceedings of CVPR, 2017.
8) W. Liu et al., "SSD: Single Shot MultiBox Detector," European Conference on Computer Vision, 2016.
9) Z. Ge et al., "YOLOX: Exceeding YOLO Series in 2021," arXiv preprint arXiv:2107.08430, 2021.
10) A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," International Conference on Learning Representations, 2021.