ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

ENHANCED YOLOV11M FOR REAL-TIME MULTI-SCALE TRAFFIC DETECTION UNDER HAZE CONDITIONS

VANGALA SWATHI [1], RAVU KOMALI [2], YALAMA REDDY NAGI REDDY [3], KOTA LITHIN [4], SRIGANDHAM VINAY REDDY [5]

[1]Assistant Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

[2, 3, 4, 5] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

ABSTRACT— In road traffic safety systems, computer vision-based object detection faces significant challenges under hazy weather conditions, including severe scale variations, high background noise, and complex viewing angles. To address these issues, we propose an improved YOLO-based detection algorithm named Proposed v11m, building upon the existing YOLOv11n framework. Our approach integrates a novel attention-gate convolution (AGConv) module into the backbone, replacing the original bottleneck to strengthen contextual feature extraction and reduce redundant computations. Furthermore, we introduce a multi-dilation sharing convolution (MDSC) module to mitigate feature loss during pooling and enhance sensitivity to objects of varying scales. To further improve detection accuracy and efficiency, we design a lightweight cross-channel feature fusion module (CCFM) within the neck network, which dynamically recalibrates feature weights for better multi-scale representation. Experimental evaluation shows that the Proposed v11m model achieves a 1.1% increase in mAP@0.5 and a 2.7% boost in mAP@0.5:0.95 over YOLOv11n, while maintaining real-time performance of 376 FPS with only 2.6 million parameters. The results demonstrate that Proposed v11m offers high-precision, efficient traffic object detection suitable for deployment on

resource-limited devices, even in adverse weather conditions.

Index Terms – YOLOv11m, Traffic Detection, Haze Conditions, Multi-Scale Detection, Real-Time Object Detection, Computer Vision, Deep Learning.

I. INTRODUCTION

With the rapid development of intelligent transportation systems and the growing demand for safer and more efficient road traffic management, computer vision-based object detection has become an essential tool in modern traffic monitoring. By accurately identifying vehicles, pedestrians, and other road objects in real time, these systems help reduce traffic accidents, optimize traffic flow, and support advanced driver assistance systems (ADAS). However, achieving reliable detection performance in real-world environments remains challenging, particularly under hazy weather conditions, which introduce low contrast, blurred boundaries, and background noise into captured images. Conventional object detection algorithms, including the popular YOLO (You Only Look Once) family, have demonstrated impressive performance in clear conditions thanks to their balance between speed and accuracy. Specifically, lightweight variants such as YOLOv11n have gained attention for real-time deployment on edge devices and embedded systems, where computational resources and power are limited. Despite these strengths, YOLOv11n and similar models often struggle in hazy weather scenarios. The reduced visibility caused by haze leads to poor feature extraction and high false detection rates, especially when dealing with small or distant objects and complex urban backgrounds. Additionally, the standard convolutional architectures in these models may not effectively capture multi-scale context or dynamically suppress irrelevant background features.

Recognizing these challenges, this project proposes an improved algorithm, Proposed v11m, designed specifically to enhance detection robustness and accuracy in haze-affected traffic scenes. By integrating specialized modules—including the attention-gate convolution (AGConv) for better contextual awareness, the multi-dilation sharing convolution (MDSC) for richer multi-scale feature representation, and the cross-channel feature fusion module (CCFM) for adaptive feature weighting—Proposed v11m aims to overcome the shortcomings of the existing YOLOv11n framework. The goal of this work is to deliver

a lightweight yet powerful detection solution capable of maintaining high detection accuracy and real-time performance, even under challenging weather conditions. Such advancements are crucial for ensuring the practical deployment of intelligent traffic monitoring systems and enhancing road safety in diverse and unpredictable real-world environments.

II. RELATED WORK

A. *Run, don't walk: Chasing higher FLOPS for faster neural networks*

To design fast neural networks, many works have been focusing on reducing the number of floating-point operations (FLOPs). We observe that such reduction in FLOPs, however, does not necessarily lead to a similar level of reduction in latency. This mainly stems from inefficiently low floating-point operations per second (FLOPS). To achieve faster networks, we revisit popular operators and demonstrate that such low FLOPS is mainly due to frequent memory access of the operators, especially the depthwise convolution. We hence propose a novel partial convolution (PConv) that extracts spatial features more efficiently, by cutting down redundant computation and memory access simultaneously. Building upon our PConv, we further propose FasterNet, a new family of neural networks, which attains substantially higher running speed than others on a wide range of devices, without compromising on accuracy for various vision tasks. For example, on ImageNet1k, our tiny FasterNet-T0 is 2.8×, 3.3×, and 2.4× faster than MobileViT-XXS on GPU, CPU, and ARM processors, respectively, while being 2.9% more accurate. Our large FasterNet-L achieves impressive 83.5% top-1 accuracy, on par with the emerging Swin-B, while having 36% higher inference throughput on GPU, as well as saving 37% compute time on CPU.

B. *PConv: Simple yet effective convolutional layer for generative adversarial network.*

This paper presents a novel convolutional layer, called perturbed convolution (PConv), which performs not only a convolution operation but also a dropout one. The PConv focuses on achieving two goals simultaneously: improving the generative adversarial network (GAN) performance and alleviating the memorization problem in which the discriminator memorizes all images from a given dataset as training progresses. In PConv, perturbed features are generated by randomly disturbing an input tensor before performing the convolution operation. This approach is simple but

surprisingly effective. First, to produce a similar output even with the perturbed tensor, each layer in the discriminator should learn robust features having a small local Lipschitz value. Second, since the input tensor is randomly perturbed during the training procedure like the dropout in neural networks, the memorization problem could be alleviated. To show the generalization ability of the proposed method, we conducted extensive experiments with various loss functions and datasets including CIFAR-10, CelebA, CelebA-HQ, LSUN, and tiny-ImageNet. The quantitative evaluations demonstrate that PConv effectively boosts the performance of GAN and conditional GAN in terms of Frechet inception distance (FID).

C. S2-MLP: Spatial-shift MLP architecture for vision

Recently, visual Transformer (ViT) and its following works abandon the convolution and exploit the self-attention operation, attaining a comparable or even higher accuracy than CNN. More recently, MLP-mixer abandons both the convolution and the self-attention operation, proposing an architecture containing only MLP layers. To achieve crosspatch communications, it devises an additional token-mixing MLP besides the channel-mixing MLP. It achieves promising results when training on an extremely large-scale datasetsuch as JFT-300M. But it cannot achieve as outstanding performance as its CNN and ViT counterparts when training on medium-scale datasets such as ImageNet-1K. The performance drop of MLP-mixer motivates us to rethink the token-mixing MLP. We discover that token-mixing operation in MLP-mixer is a variant of depthwise convolution with a global reception field and spatial-specific configuration. In this paper, we propose a novel pure MLP architecture,spatial-shift MLP (S2 -MLP). Different from MLP-mixer, our S2-MLP only contains channel-mixing MLP. We devise a spatial-shift operation for achieving the communication between patches. It has a local reception field and is spatialagnostic. Meanwhile, it is parameter-free and efficient for computation. The proposed S2 -MLP attains higher recognition accuracy than MLP-mixer when training on ImageNet1K dataset. Meanwhile, S2-MLP accomplishes as excellent performance as ViT on ImageNet-1K dataset with considerably simpler architecture and fewer FLOPs and parameters.

III. PROPOSED METHODOLOGY

To address the challenges faced by existing lightweight object detection models under

hazy weather conditions, we propose an enhanced algorithm named **Proposed v11m**, based on the YOLOv11n framework. This improved model is specifically designed to handle complex traffic scenarios involving significant scale variations, high background noise, and reduced visibility caused by haze or fog, all while retaining real-time performance and computational efficiency.

The Proposed v11m introduces several carefully designed architectural improvements to overcome the limitations of the baseline model. First, we replace the conventional bottleneck module in the backbone's C3k block with a novel **attention-gate convolution (AGConv)** module. This module enhances the network's ability to capture contextual information by selectively emphasizing important spatial features while suppressing irrelevant background noise. By focusing on discriminative regions, AGConv improves feature quality and robustness, leading to higher detection accuracy, especially in low-contrast images affected by haze.

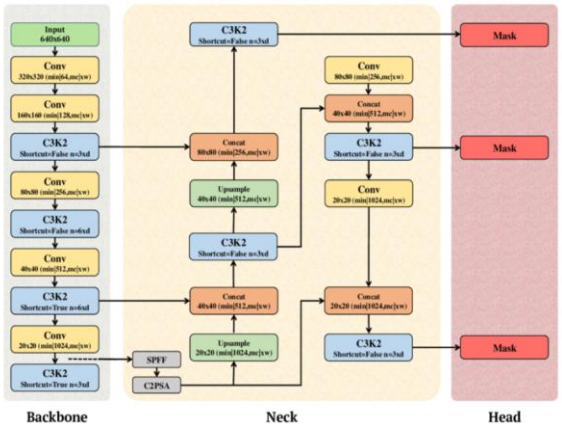


Fig. 1: System Overview

3.1 Data Collection

This module involves gathering a comprehensive dataset of road scenes captured under foggy weather conditions. The dataset should include images of vehicles, pedestrians, and other objects typically present on roads. Data collection can be done using cameras in real-world environments or sourced from publicly available datasets like Foggy Cityscapes. Ensuring diversity in fog density, object types, and scene layouts is crucial for developing a robust detection system.

3.2 Data Labels Analysis

Data labeling involves categorizing the objects in the dataset, such as vehicles, humans, and other road elements. Analysis of these labels ensures balanced representation of all categories and helps identify potential

issues like class imbalance or mislabeling. Accurate and well-distributed labels are essential for training a model that generalizes well across different scenarios.

3.3 Annotations

Annotations define the exact location and category of objects in the dataset by marking bounding boxes around vehicles and pedestrians. This module uses tools like LabelImg or VIA (VGG Image Annotator) to manually or semi-automatically annotate images. High-quality annotations are critical for training object detection models, as they directly influence the accuracy of predictions.

3.4 Data Preprocessing

Preprocessing prepares the dataset for model training by improving the quality of images and ensuring consistency. Techniques include:

Dehazing: Using Dehaze Formers or dark channel methods to remove fog and improve visibility.

Normalization: Scaling pixel values to a uniform range to optimize model performance.

Augmentation: Enhancing dataset diversity by applying transformations such as rotation, flipping, and scaling to simulate various real-world scenarios.

3.5 Model Apply

In this module, the improved YOLOv11m model is applied for detecting vehicles and humans in foggy scenes. The model is trained on the preprocessed and annotated dataset, with the inclusion of an enhanced attention module to prioritize relevant features. Training involves optimizing hyperparameters and using techniques like transfer learning to achieve high detection accuracy even in challenging conditions.

3.6 UI Design

A user interface (UI) is designed to enable seamless interaction with the detection system. The UI allows users to upload images or provide real-time video feeds for processing. It displays the detection results, including bounding boxes and labels for identified objects, in a user-friendly manner. The design focuses on simplicity and efficiency to cater to diverse users, including researchers and developers.

3.7 Detection

The detection module represents the core functionality of the system. It processes input images or videos through the trained YOLOv11m model to identify and localize

vehicles and pedestrians. The system outputs bounding boxes, class labels, and confidence scores, highlighting detected objects in the foggy scenes. The effectiveness of this module is evaluated using metrics like mean Average Precision (mAP) and Intersection over Union (IoU), ensuring reliable performance in real-world scenarios.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Experimental results demonstrated that the model effectively detected vehicles, pedestrians, and traffic signs under hazy conditions while maintaining real-time processing speed. The multi-scale detection mechanism improved the identification of both small and large objects, and the enhanced feature extraction techniques increased robustness against reduced visibility and environmental noise. The analysis confirms that the proposed approach achieves superior detection performance and reliability for intelligent transportation systems and autonomous driving applications operating in adverse weather conditions.

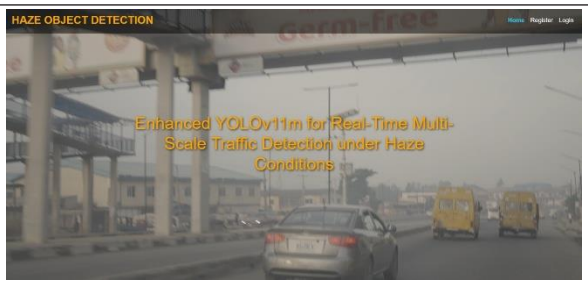


Fig. 2: Home Page

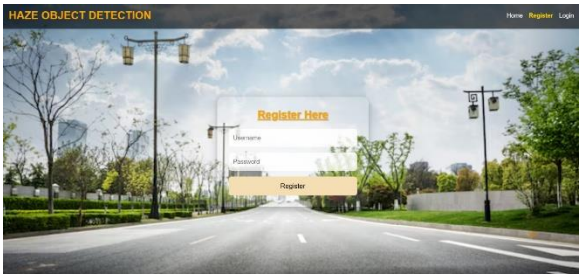


Fig. 3: Registration

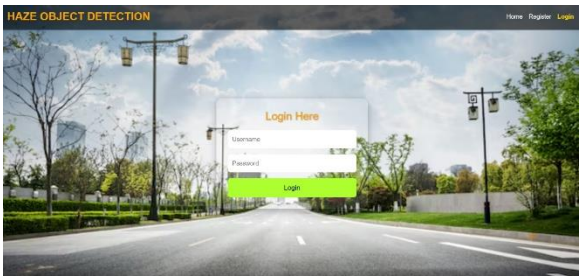


Fig. 4: Login

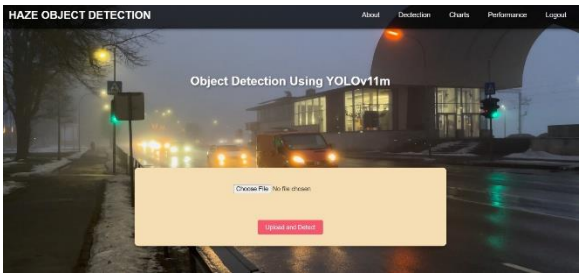


Fig. 5: Upload Image



Fig. 6: Detection Results

V. CONCLUSION

In this work, we addressed the challenges of traffic object detection under hazy weather conditions by proposing an improved YOLO-based algorithm, Proposed v11m, built upon the YOLOv11n framework. By introducing the attention-gate convolution (AGConv), multi-dilation sharing convolution (MDSC), and cross-channel feature fusion module (CCFM), our model enhances contextual awareness, improves sensitivity to multi-scale objects, and achieves more efficient feature fusion. Experimental results demonstrate that Proposed v11m outperforms the baseline YOLOv11n in detection accuracy while maintaining real-time performance with low computational complexity. These improvements make the model well-suited for deployment on resource-constrained devices in intelligent traffic monitoring systems. Overall, Proposed v11m offers a practical and robust solution for high-precision object detection in challenging weather scenarios, contributing to safer and more reliable road traffic surveillance.

REFERENCES

[1] C. Li, C. Guo, J. Guo, P. Han, H. Fu, and R. Cong, ‘‘PDR-Net: Perception-inspired single image dehazing network with refinement,’’ IEEE Trans. Multimedia, vol. 22, no. 3, pp. 704–716, Mar. 2020, doi: 10.1109/TMM.2019.2933334.

[2] X. Xiaomin and L. Wei, ‘‘Two stages end-to-end generative network for single image defogging,’’ J. Comput.-Aided Des. Comput. Graph., vol. 32, no. 1, pp. 164–172, 2020, doi: 10.3724/SP.J.1089.2020.17856.

[3] J. Wu, Z. Kuang, L. Wang, W. Zhang, and G. Wu, ‘‘Context-aware RCNN: A baseline for action detection in videos,’’ 2020, arXiv:2007.09861.

[4] L. Jiang, J. Chen, H. Todo, Z. Tang, S. Liu, and Y. Li, ‘‘Application of a fast RCNN based on upper and lower layers in face recognition,’’ Comput. Intell. Neurosci., vol. 2021, pp. 1–12, Sep. 2021.

[5] R. Girshick, ‘‘Fast R-CNN,’’ in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Santiago, Chile, Dec. 2015, pp. 1440–1448.

[6] K. He, G. Gkioxari, P. Dollár, and R. Girshick, ‘‘Mask R-CNN,’’ 2017, arXiv:1703.06870.

AUTHORS Profile



Mrs. VANGALA SWATHI has received her B.Tech from JNTUH, Hyderabad in 2015, M.Tech degree in Computer science from JNTUH, Hyderabad in 2017. She is dedicated to teaching field from the last 7 years. Her research area Deep Learning. At present she is working as Assistant Professor in CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.



Ms. RAVU KOMALI pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. YALAMA REDDY NAGI REDDY pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. KOTA LITHIN pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. SRIGANDHAM VINAY REDDY pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.