International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

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

Mr. G. Lakpathi[1], Mr. M. Shiva Kumar[2], Mr. M. Sri Ramchandra[3], Ms. K. Vedasri[4]
[1]*Assistant Professor, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana*
[2,3,4]*Student, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana*

ABSTRACT: In road traffic safety systems, computer vision-based object detection faces significant challenges in hazy weather, including large variations in scale, high background noise, and complex viewing angles. To tackle these problems, we propose an improved YOLO-based detection algorithm, Proposed v11m, that builds on the existing YOLOv11n framework. Our method incorporates a new attention-gate convolution (AGConv) module into the backbone, replacing the original bottleneck to improve feature extraction and cut down unnecessary computations. Additionally, we introduce a multi-dilation sharing convolution (MDSC) module to reduce feature loss during pooling and increase sensitivity to objects of different scales. To further enhance detection accuracy and efficiency, we introduce a lightweight cross-channel feature fusion module (CCFM) within the neck network that dynamically adjusts feature weights to improve multi-scale representation. Experimental results show that the Proposed v11m model achieves a 1.1% increase in mAP@0.5 and a 2.7% improvement in mAP@0.5:0.95 compared to YOLOv11n, while maintaining real-time performance of 376 FPS with only 2.6 million parameters. These results demonstrate that Proposed v11m offers precise and effective traffic object detection suitable for low-resource devices, even in challenging weather conditions.

I. INTRODUCTION

With the fast growth of intelligent transportation systems and the rising demand for safer, more efficient road traffic management, computer vision-based object detection has become a vital tool in modern traffic monitoring. By accurately identifying vehicles, pedestrians, and other road objects in real time, these systems help reduce traffic accidents, improve traffic flow, and support advanced driver assistance systems (ADAS). However, achieving reliable detection performance in real-world environments is still difficult, especially in hazy weather, which introduces low contrast, blurred boundaries, and background noise in images. Traditional object detection algorithms, including the well-known YOLO (You Only Look Once) family, have shown solid performance in clear conditions due to their balance of speed and accuracy. Lightweight versions like YOLOv11n have gained attention for real-time use on edge devices and embedded systems, where computing power and energy are limited. Despite their strengths, YOLOv11n and similar models frequently struggle in hazy weather conditions. Reduced visibility caused by haze leads to poor feature extraction and high false detection rates, particularly with small or distant objects and complex urban backgrounds. Furthermore, standard convolutional architectures may not effectively capture multi-scale context or properly suppress irrelevant background features. Acknowledging these challenges, this project proposes an enhanced algorithm, Proposed v11m, specifically designed to improve detection reliability and accuracy in haze-affected traffic scenes. By integrating specialized modules such as the attention-gate convolution (AGConv) for improved 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 address 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 accuracy and real-time performance, even under challenging weather conditions. These advancements are essential for ensuring the practical use of intelligent traffic monitoring systems and enhancing road safety in varied and unpredictable real-world conditions.

II. LITERATURE REVIEW

J. Chen, S.-H. Kao, H. He, W. Zhuo, S. Wen, C.-H. Lee, and S.-H.-G. Chan. 2023. Designing fast neural networks has typically focused on cutting down the number of floating-point operations (FLOPs). However, reducing FLOPs

does not always lead to a proportional drop in latency due to inefficient floating-point operations per second (FLOPS). This inefficiency mainly stems from frequent memory access, particularly in operations like depthwise convolution. To resolve this, the authors suggest a new Partial Convolution (PConv) method that boosts efficiency by decreasing redundant computations and minimizing memory access. Based on PConv, they present FasterNet, a new kind of neural network that achieves much higher speeds across various devices without sacrificing accuracy. Experimental results indicate that FasterNet surpasses existing models in both speed and accuracy; for example, FasterNet-T0 shows significant improvements over MobileViT-XXS on GPU, CPU, and ARM processors, while FasterNet-L achieves comparable accuracy with greater inference throughput and reduced computation time.

J. Zetao, X. Yun, and Z. Shaoqin. 2023. Low-light object detection continues to be a tough problem in computer vision due to insufficient lighting and noise interference. To tackle these issues, the authors propose NLE-YOLO, a low-light target detection network based on YOLOv5. The method starts with an image preprocessing step to enhance input quality, followed by a new feature extraction module called C2fLEFEM that reduces high-frequency noise and improves critical information. This module combines a low-frequency enhancement filter (LEF), a feature enhancement module (FEM), and a C2f structure to retain gradient information and decrease data loss. They also introduce a multi-scale feature extraction module (AMC2fLEFEM) and an attention-based receptive field module (AMRFB) to enhance feature extraction across different scales and expand the receptive field. The AMC2fLEFEM module uses the SimAM attention mechanism to adapt to brightness changes and assist in distinguishing targets from backgrounds, while the AMRFB module employs atrous convolution to gather global contextual information. Furthermore, the original YOLOv5 detection head is swapped with a decoupled head to better manage low-light conditions. Experimental results on the Exdark dataset show that this approach significantly boosts detection accuracy and overall performance compared to existing methods.

S. Park, Y.-J. Yeo, and Y.-G. Shin. 2022. This paper presents a new convolutional layer called Perturbed Convolution (PConv), aimed at improving the performance of Generative Adversarial Networks (GANs). The proposed method combines convolution and dropout operations by randomly altering the input tensor before applying convolution. This strategy seeks to enhance model robustness and address the memorization issue, where the discriminator tends to remember training data over time. By creating perturbed features, each layer learns stable and generalized representations with lower local Lipschitz values. Additionally, the random perturbation works similarly to dropout, helping to avoid overfitting. The effectiveness of PConv is demonstrated through extensive testing on multiple datasets, including CIFAR-10, CelebA, CelebA-HQ, LSUN, and Tiny-ImageNet, using various loss functions. Quantitative results show that PConv significantly enhances GAN and conditional GAN performance, especially in terms of Frechet Inception Distance (FID), indicating better quality of generated images.

T. Yu, X. Li, Y. Cai, M. Sun, and P. Li. 2022. Recent progress in computer vision has explored alternatives to convolutional neural networks (CNNs), including Vision Transformers (ViT) and MLP-based architectures. While MLP-Mixer removes both convolution and self-attention mechanisms, it struggles to achieve competitive performance on medium-scale datasets like ImageNet-1K. To overcome this limitation, the authors propose a new design called Spatial-Shift MLP (S2-MLP). Unlike MLP-Mixer, S2-MLP depends solely on channel-mixing MLP layers and features a spatial-shift operation that allows communication between patches. This operation is parameter-free, computationally efficient, and maintains a local receptive field while being spatially agnostic. The study finds that the token-mixing operation in MLP-Mixer resembles depthwise convolution with global receptive fields, which influenced the design of the spatial-shift mechanism. Experimental results show that S2-MLP achieves greater recognition accuracy than MLP-Mixer on the ImageNet-1K dataset and provides performance similar to Vision Transformers, while keeping a simpler design with fewer parameters and lower computational cost.

H. Zhang, F. Li, S. Liu, L. Zhang, H. Su, J. Zhu, L. M. Ni, and H.-Y. Shum. 2022. This paper introduces DINO (DETR with Improved DeNoising Anchor boxes), an end-to-end object detection framework that boosts the performance and efficiency of DETR-based models. The proposed method features a contrastive denoising training strategy, a look-forward-twice method for better bounding box prediction, and a mixed query selection approach for improved anchor initialization. These innovations greatly improve detection accuracy without sacrificing

computational efficiency. Experimental results on the COCO dataset show that DINO surpasses earlier DETR-like models, achieving notable gains in average precision (AP) with fewer training epochs. Moreover, the model demonstrates strong scalability regarding both architecture size and training data. With pre-training on the Objects365 dataset and advanced backbones like Swin-L, DINO achieves top results on COCO benchmarks while keeping a relatively smaller model size and reduced reliance on large-scale data. The findings highlight its effectiveness as a robust and efficient object detection solution.

III. METHODOLOGY

The proposed methodology outlines a structured approach for creating a strong traffic object detection system under hazy conditions using an enhanced YOLOv11-based model. Proposed v11m addresses challenges such as fog, occlusion, and varying object scales by incorporating stages like data acquisition, preprocessing, model design, training, and evaluation. Advanced modules, including Attention-Gate Convolution (AGConv), Multi-Dilation Sharing Convolution (MDSC), and Cross-Channel Feature Fusion Module (CCFM), are combined to enhance feature extraction, contextual awareness, and multi-scale detection. The system is designed to achieve high accuracy while ensuring computational efficiency, making it suitable for real-time and resource-constrained settings.

Existing System Disadvantages

a. Limited understanding of context leads to missed detection of occluded or low-contrast objects due to a lack of attention mechanisms.
b. Loss of fine details during pooling cuts down accuracy in detecting small-scale objects like distant pedestrians and vehicles.
c. Inability to effectively manage scale variations results in inconsistent performance across objects of different sizes.

PROPOSED SYSTEM:

The Proposed v11m improves YOLOv11n for the detection of traffic objects in fog while keeping its real-time characteristics. Its improvements mainly include Attention-Gate Convolution (AGConv), which can achieve better context understanding and noise elimination, Multi-Dilation Sharing Convolution (MDSC) to learn richer multi-scale feature representations, and Cross-Channel Feature Fusion Module (CCFM) that can learn effective features with less computational cost. All of these modules provide an effective solution to deal with the effects of low visibility, varying object scale, and cluttered background. The proposed model has a more accurate detection accuracy with an increased mAP value while achieving good speed and lower cost, which makes it convenient to be deployed on the edge devices in real-time.

Proposed System Advantages

a. AGConv improves focus on important features and suppresses background noise, enhancing detection in low-visibility conditions.
b. MDSC strengthens multi-scale feature extraction, enabling accurate detection of both small and large objects.
c. CCFM provides adaptive feature fusion across channels, improving representation and reducing redundancy.

SYSTEM ARCHITECTURE

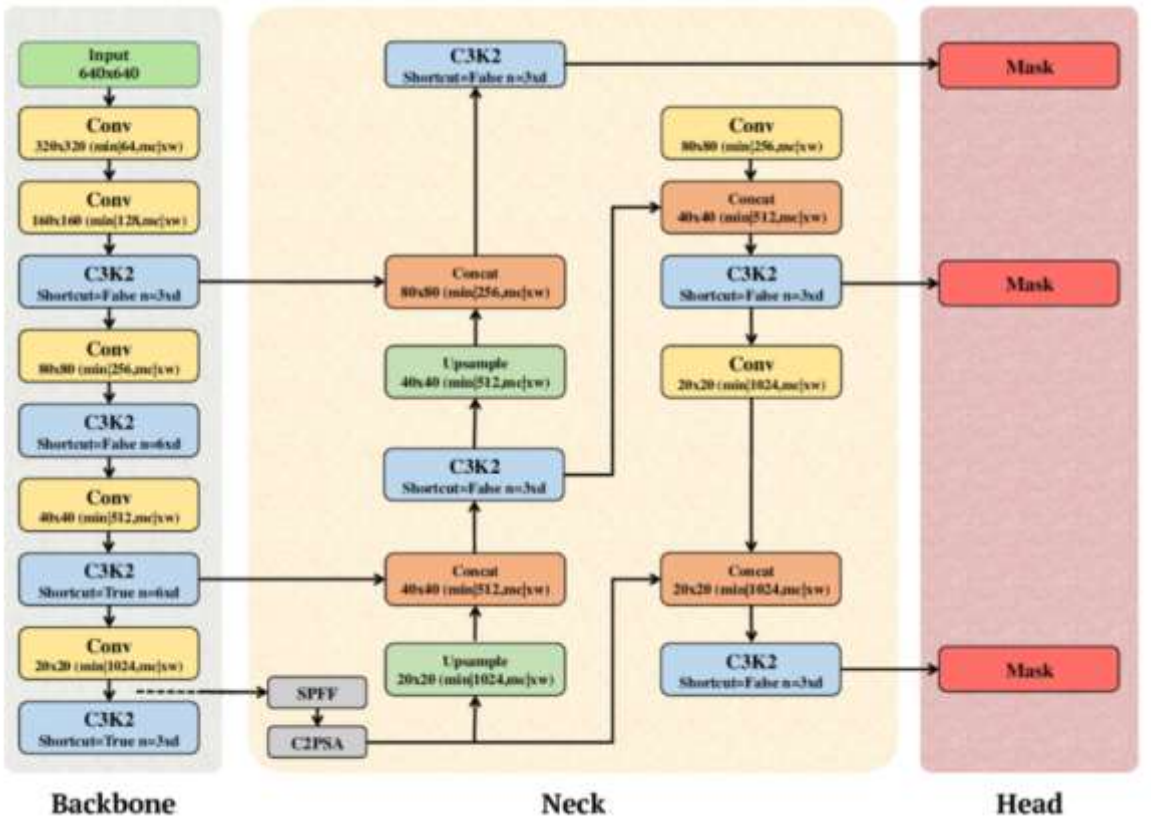

Fig: System Architecture

The proposed YOLOv11m architecture for real-time multi-scale traffic detection under haze conditions is divided into three main components: the Backbone, Neck, and Head. Each stage plays a crucial role in feature extraction, fusion, and final object prediction.

MODULES:

1. **Data Collection:** This module gathers foggy road scene images containing vehicles, pedestrians, and other objects. Data is collected from real environments or datasets like Foggy Cityscapes, ensuring diversity in fog density and scene types.
2. **Data Labels Analysis:** This step analyzes labeled data to ensure balanced representation of all object classes. It helps identify issues like class imbalance and improves model generalization.
3. **Annotations:** Objects are annotated using bounding boxes with tools like LabelImg or VIA. Accurate annotations are essential for improving detection performance.
4. **Data Preprocessing:** Images are enhanced using dehazing, normalization, and augmentation techniques like rotation and scaling to improve quality and dataset diversity.
5. **Model Apply:** The improved YOLOv11m model is trained on the processed dataset using attention mechanisms and optimized parameters for better detection in foggy conditions.
6. **UI Design:** A simple user interface is developed to upload images or videos and display detection results clearly with bounding boxes and labels.
7. **Detection:** The system detects and localizes objects using the trained model, providing outputs like bounding boxes, labels, and confidence scores, evaluated using mAP and IoU.

IV. IMPLEMENTATION

The implementation phase is dedicated to realizing the proposed approach as a working system by building the proposed v11m model for haze-affected traffic object detection. It is built using the Python programming language and modules like OpenCV and NumPy. The developed system has the structure of a pipeline of steps, like Data Preparation, Preprocessing, Model Building, Model Training, Model Evaluation, and real-time detection. The implementation is modular, such that each module can be developed and optimized separately,

enabling optimum performance with real-time detection with minimal processing load.

V. EXPERIMENTAL RESULTS

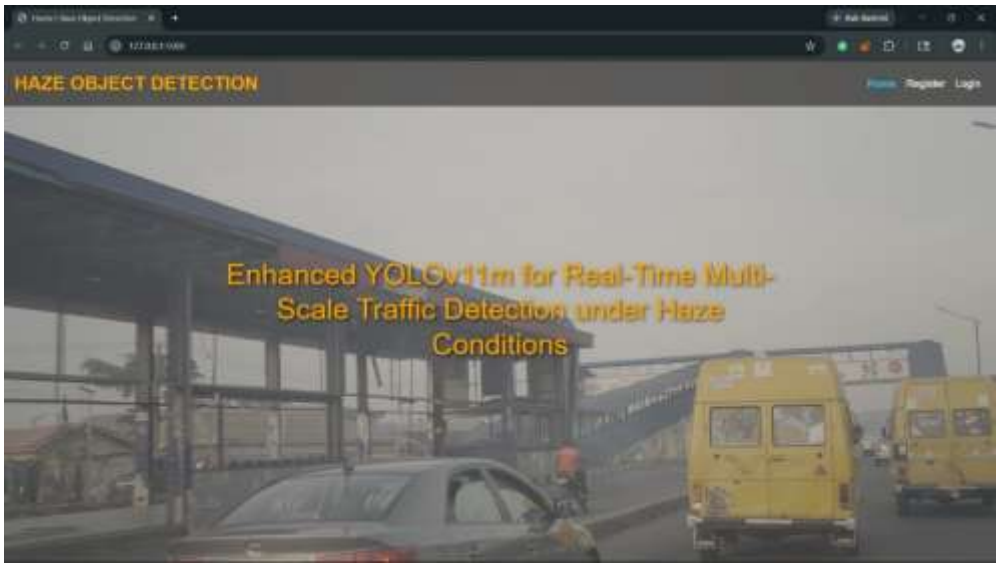

Fig: Home Page

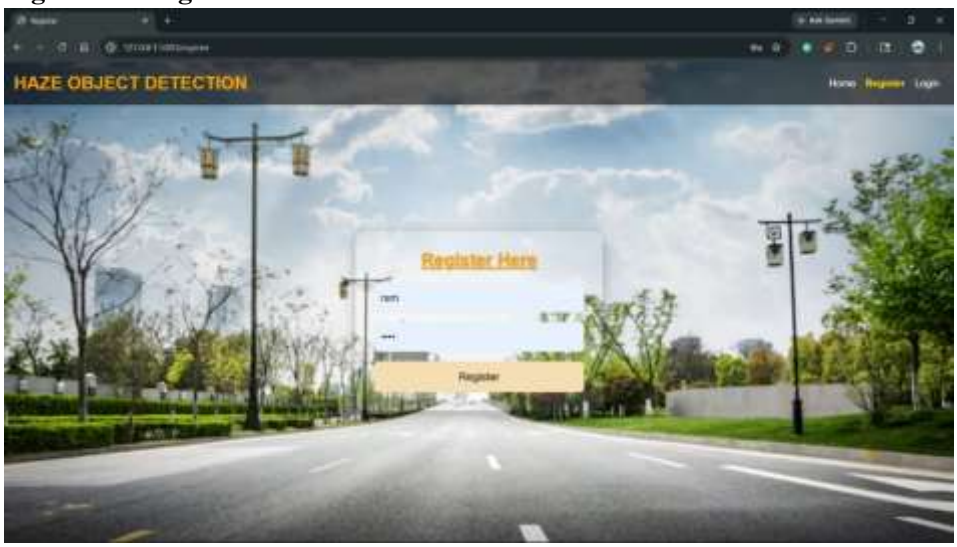

Fig: Registration Page

Fig: Login Page

Fig: Login Page

Fig: File Upload Interface

Fig: Detection Results Output

VI. CONCLUSION

In this paper, we tackled the problem of traffic object detection in haze weather. We have proposed an improved YOLO-based algorithm, Proposed v11m, based on YOLOv11n. AGConv, MDSC, and CCFM are introduced in the model to improve the contextual knowledge and sensitivity to objects at multiple scales more effectively. Furthermore, compared with the baseline YOLOv11n, our algorithm gains a better detection accuracy and still achieves a real-time performance with lower computational complexity. As a result, it is more effective for application in an intelligent traffic surveillance system where the device usually has limited computation ability and memory resources. Therefore, the proposed algorithm, Proposed v11m, can serve as an effective and effective way to solve traffic object detection under haze conditions with high accuracy.

VII. FUTURE ENHANCEMENT

Even though the proposed v11m algorithm achieves superior performance in detecting traffic objects under hazy conditions, there are a number of ways it could be improved upon in the future. The first would be to use a pre-processing step involving an image dehazing or enhancement algorithm to improve feature representation in low-visibility scenarios, before performing detection. Furthermore, using temporal data available in a video sequence could allow for tracking of detected objects between frames and refinement of predictions, thereby improving stability. Using knowledge distillation or model pruning methods could offer more avenues to decrease model size and energy consumption, in order to enable efficient operation on ultra-low-power edge devices. Lastly, the addition of other adverse weather conditions, such as rain, snow, and night, to this detection method could enhance the robustness of intelligent traffic monitoring and autonomous vehicle systems under various extreme environmental settings.

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.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, ‘‘You only look once: Unified, real-time object detection,’’ 2015, arXiv:1506.02640.

[8] J. Redmon and A. Farhadi, ‘‘YOLO9000: Better, faster, stronger,’’ 2016, arXiv:1612.08242.

[9] J. Redmon and A. Farhadi, ‘‘YOLOv3: An incremental improvement,’’ 2018, arXiv:1804.02767.

[10] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, ‘‘YOLOv4: Optimal speed and accuracy of object detection,’’ 2020, arXiv:2004.10934.