IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 12, Issue 3, May-June 2025

Impact Factor: 8.152

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203050

Road Object Detection in Foggy Complex Scenes Based on Improved YOLOv10

Shaikh Mohamad[1], Srikanta[2], Shashivardhan[3], Vishal Kale [4]

Assistant Professor, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana, India[1]

Student, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana, India[2-4]

ABSTRACT: Foggy weather presents substantial challenges for vehicle detection systems due to reduced visibility and the obscured appearance of objects. To overcome these challenges, a novel vehicle and Humans detection algorithm based on an improved lightweight YOLOv10 model is introduced. The proposed algorithm leverages advanced preprocessing techniques, including data transformations, Dehaze Formers, and dark channel methods, to improve image quality and visibility. These preprocessing steps effectively reduce the impact of haze and low contrast, enabling the model to focus on meaningful features. An enhanced attention module is incorporated into the architecture to improve feature prioritization by capturing long-range dependencies and contextual information. This ensures that the model emphasizes relevant spatial and channel features, crucial for detecting small or partially visible vehicles in foggy scenes. Furthermore, the feature extraction process has been optimized, integrating an advanced lightweight module that improves the balance between computational efficiency and detection performance. This research addresses critical issues in adverse weather conditions, providing a robust framework for foggy weather vehicle and Humans detections.

KEYWORDS: Road Object Detection, Improved YOLOv10, Convolutional Neural Networks (CNNs)

I. INTRODUCTION

Foggy weather significantly hinders the performance of vehicle detection systems, as fog reduces visibility, obscures object details, and creates low-contrast conditions. Accurate detection of vehicles and pedestrians in such weather conditions is critical for improving the safety and reliability of autonomous driving systems. Traditional detection methods often fail to deliver satisfactory results under foggy conditions due to the blurring and reduction in contrast that fog causes. In this paper, we propose a novel vehicle and human detection algorithm based on an improved lightweight YOLOv10 model designed to overcome these challenges. By incorporating advanced preprocessing techniques such as Dehaze Formers and dark channel methods, we aim to enhance image quality and restore visibility, enabling the model to focus on critical features despite the fog. Furthermore, an enhanced attention module is introduced to prioritize important spatial and channel features, ensuring better performance in detecting small or partially visible objects, which are common in foggy scenes. Our approach leverages the power of YOLOv10, optimized for both computational efficiency and high detection accuracy, to deliver reliable results even in the most challenging conditions.

II. LITERATURE SURVEY

Chen, J., Kao, S.-H., He, H., Zhuo, W., Wen, S., Lee, C.-H., & Chan, S.-H.-G. (2023). 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.

Park, S., Yeo, Y.-J., & Shin, Y.-G. (2022)**.** 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.

Zetao, J., Yun, X., & Shaoqin, Z. (2023). Low-light object detection is an important research area in computer vision, but it remains a challenging problem due to insufficient illumination and noise interference. This study proposes NLE-YOLO, a low-light target detection network based on YOLOv5, specifically designed to address these issues. The network begins by preprocessing input images using an enhancement technique, followed by the application of C2fLEFEM—a custom feature extraction module that suppresses high-frequency noise and enhances key information. The C2fLEFEM module integrates the LEF (low-frequency enhancement), FEM (feature enhancement with dual input), and C2f (gradient-preserving) modules to reduce information loss while improving robustness. Additionally, a multi-scale feature extraction module, AMC2fLEFEM, is introduced, incorporating the SimAM attention mechanism to extract receptive-field-aware features that can adapt to brightness variations and improve target-background separation. The AMRFB module, employing atrous convolutions, further enhances the receptive field to capture global context and detect objects of various scales. Lastly, the original YOLOv5 detection head is replaced with a decoupled head better suited for low-light conditions. Experimental results on the ExDark dataset demonstrate that the proposed method significantly outperforms existing techniques in both accuracy and detection performance under low-light scenarios.

III. EXISTING SYSTEM

- YOLOv8 is an evolution of the YOLO object detection family, continuing its core design of a single-stage architecture that simultaneously predicts both object classes and bounding boxes. It is an end-to-end deep learning model that uses convolutional layers for feature extraction, followed by a series of dense layers for classification and localization.
- YOLOv8 builds on the improvements from previous versions by optimizing the backbone architecture, enhancing the detection head, and introducing various techniques to reduce computational costs while improving detection accuracy. The model has become widely used in real-time applications due to its efficiency and ability to handle high-speed processing tasks.

Existing System Disadvantages

- Despite improvements, YOLOv8 may still struggle with detecting small objects, especially when they are surrounded by large, complex backgrounds.
- YOLOv8 might not perform optimally under challenging conditions like fog, heavy rain, or low-light environments unless specifically trained on such data.
- YOLOv8 can have difficulty detecting objects that are partially occluded, which is common in real-world scenarios such as crowded scenes or vehicles hidden by other objects

Proposed System

- YOLOv10 is the latest advancement in the YOLO series, with improvements that make it better suited for real-time detection in various environments, including low-visibility or adverse weather conditions. It introduces several innovations in architecture and feature extraction to handle small, occluded, or distorted objects more effectively. YOLOv10 also integrates enhancements that reduce computational cost, making it more efficient without compromising on accuracy.
- The model's architecture allows it to process images faster, enabling its use in time-sensitive applications where rapid decision-making is critical. YOLOv10's increased robustness, especially in complex and challenging environments, makes it a superior choice for vehicle detection, surveillance, and other computer vision tasks.

International Journal of Advanced Research in Education and TechnologY(IJARETY)

4-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203050

Proposed System Advantages

- YOLOv10 enhances object detection capabilities, particularly for small or occluded objects, making it more robust in challenging environments such as fog, nighttime, or busy scenes.
- YOLOv10 is designed with features that enable it to handle difficult scenarios like low visibility or cluttered scenes, making it more reliable in real-world applications.
- YOLOv10 incorporates enhancements in feature extraction and model architecture, allowing it to handle a wide range of detection tasks effectively.

System Architecture

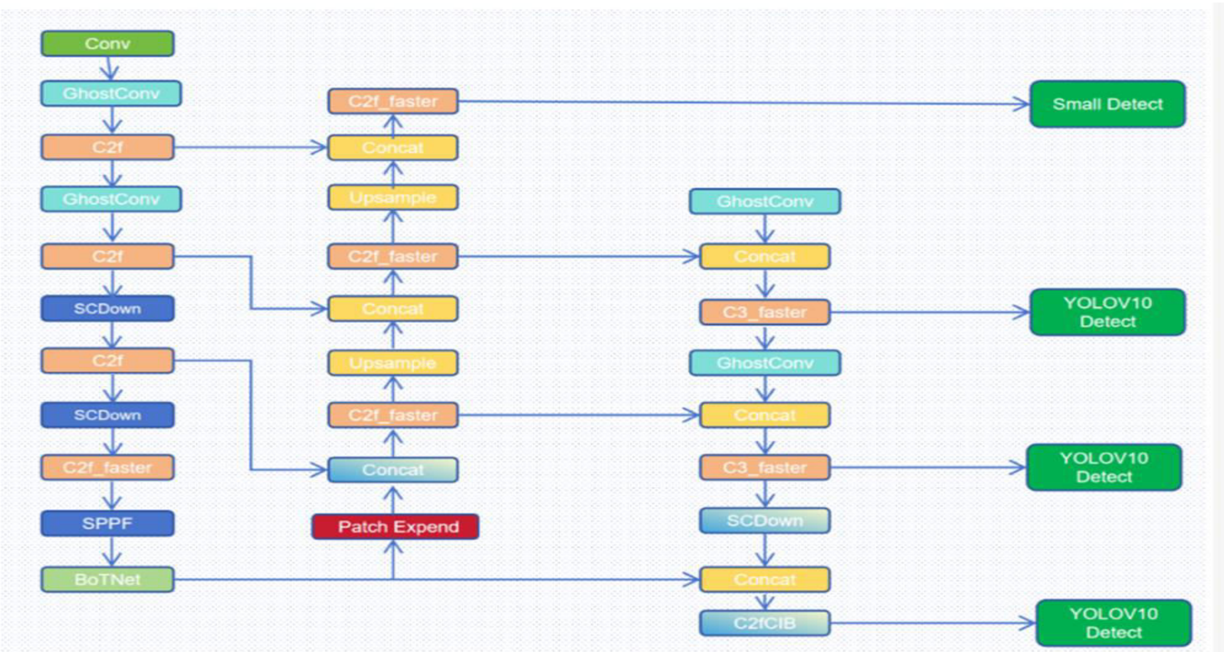


Fig 1 System Architecture

The diagram illustrates an advanced object detection architecture based on an enhanced YOLOv10 framework. It begins with a standard convolution followed by GhostConv and C2f modules, which are designed for efficient and lightweight feature extraction. The network then performs a series of downsampling operations using SCDown blocks to reduce spatial dimensions while enriching feature representations. Deeper in the network, the SPPF (Spatial Pyramid Pooling Fast) module is applied to capture multi-scale context, followed by a BoTNet module that integrates global attention for long-range dependency modeling. In the decoder path, the network upsamples the features using C2f_faster and Upsample modules, then fuses them with encoder features through Concat layers to retain spatial details. A Patch Expand module assists in resolution scaling during this process. On the right branch, GhostConv and C3_faster modules are employed to refine feature maps before passing them to YOLOv10 detection heads at multiple scales. These heads allow the model to detect small, medium, and large objects effectively. The inclusion of the C2fCIB module at the final stage further enhances low-resolution features for better detection. Overall, this architecture efficiently combines lightweight modules like GhostConv with attention and feature enhancement blocks such as BoTNet and C2f_faster, enabling high accuracy and speed for real-time object detection across various object sizes and lighting conditions.

IV. METHODOLOGY

Modules Name:

- **Data Collection**
- **Data Labels Analysis**
- **Annotations**
- **Data Preprocessing**

- **Model Apply**
- **UI Design**
- **Detection**

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.

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. **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.

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.

5. **Model Apply**
In this module, the improved YOLOv10 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.

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.

7. **Detection**
The detection module represents the core functionality of the system. It processes input images or videos through the trained YOLOv10 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.

Implementation
The implementation of Road Object Detection in Foggy Complex Scenes using an Improved YOLOv10 begins with the preparation of a specialized dataset, either by collecting real-world foggy road images or simulating fog using atmospheric scattering models. Data augmentation techniques such as random brightness adjustments, Gaussian blur, rotation, and cropping are applied alongside normalization to improve model robustness. The architecture is customized as shown in the diagram, starting with a backbone composed of Conv, GhostConv, C2f, SCDown, and BoTNet modules, which enhance feature extraction efficiency and representation. This is followed by a neck structure that

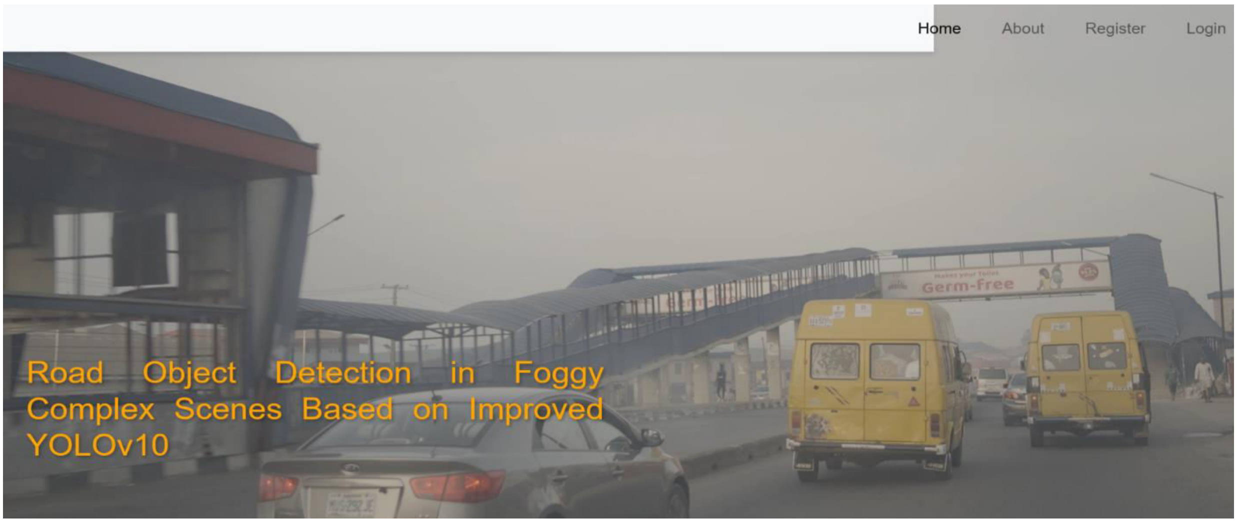
includes a Patch Expand module for upsampling and multi-scale feature fusion through repeated Upsample, Concat, and C2f_faster operations. Attention and residual modules like GhostConv and C3_faster are integrated to improve object localization under fog. The model uses multiple detection heads, each outputting YOLOv10 predictions at different scales to detect small, medium, and large objects. A fog compensation network and enhanced non-maximum suppression (NMS) further refine detection accuracy. During training, transfer learning is leveraged, and the loss function is modified to address class imbalance typical in foggy scenes. Post-training, optimization techniques such as pruning and quantization are employed for real-time deployment on edge devices, while CNN-based dehazing techniques are optionally integrated to boost visibility and detection quality in fog-heavy environments.

V. ALGORITHM USED

Existing Algorithm
YOLOv8 algorithm:
The existing algorithm, YOLOv8, is a modern and efficient object detection model developed by Ultralytics, designed for real-time performance with high accuracy. It is anchor-free, simplifying the detection process by predicting object centers directly without relying on predefined anchor boxes. YOLOv8 features a decoupled head structure that separates classification and localization tasks, resulting in better performance and easier convergence during training. The backbone utilizes C2f modules, which are lightweight and computationally efficient, making the model suitable for deployment on resource-constrained devices. YOLOv8 also integrates advanced data augmentation techniques and a robust loss function, contributing to its strong performance in standard object detection tasks. However, its effectiveness decreases significantly in complex environments like foggy road scenes, where low visibility, contrast loss, and image distortion hinder its ability to detect and localize objects accurately.

Proposed Algorithm
YOLOv10:
The proposed algorithm, Improved YOLOv10, is specifically designed to overcome the challenges presented by foggy and visually degraded road scenes. Building upon the foundation of YOLOv8, YOLOv10 introduces several critical enhancements tailored for such conditions. The architecture incorporates GhostConv layers for efficient feature extraction, SCDown modules for spatial-channel information compression, and C2f_faster/C3_faster blocks that accelerate computation while maintaining accuracy. A notable addition is BoTNet, a self-attention mechanism that enables the model to capture long-range dependencies and focus on critical features even in obscured areas. Furthermore, YOLOv10 integrates a fog compensation network and optional CNN-based dehazing, improving image clarity before detection. Multi-scale feature fusion and enhanced Non-Maximum Suppression (NMS) boost detection accuracy, especially for small or overlapping objects. The loss function is adapted to address class imbalance common in road scenes, and transfer learning from YOLOv8 weights ensures faster training. These enhancements make YOLOv10 a robust and fog-aware detection framework, significantly more effective than YOLOv8 in challenging weather conditions.

VI. RESULTS AND DISCUSSION

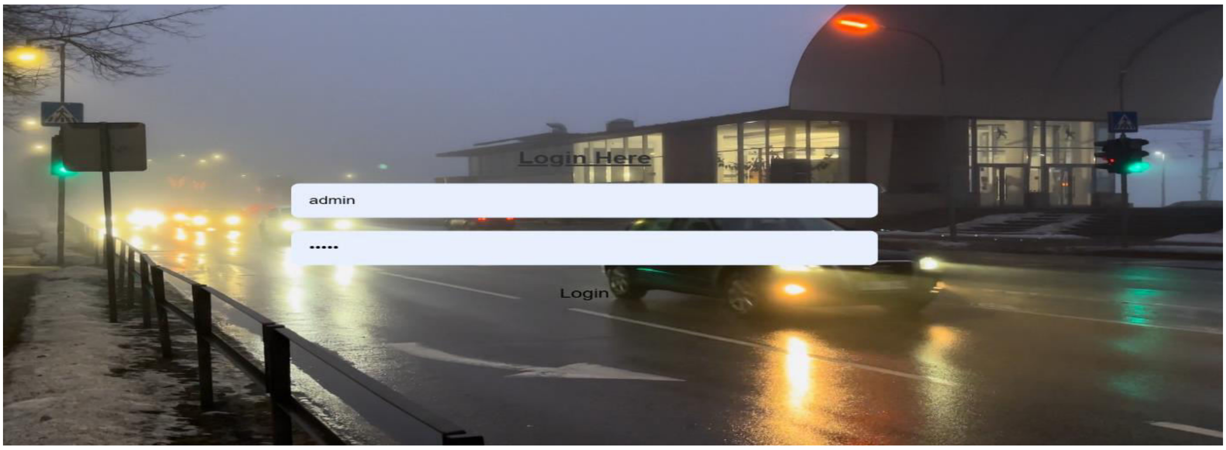


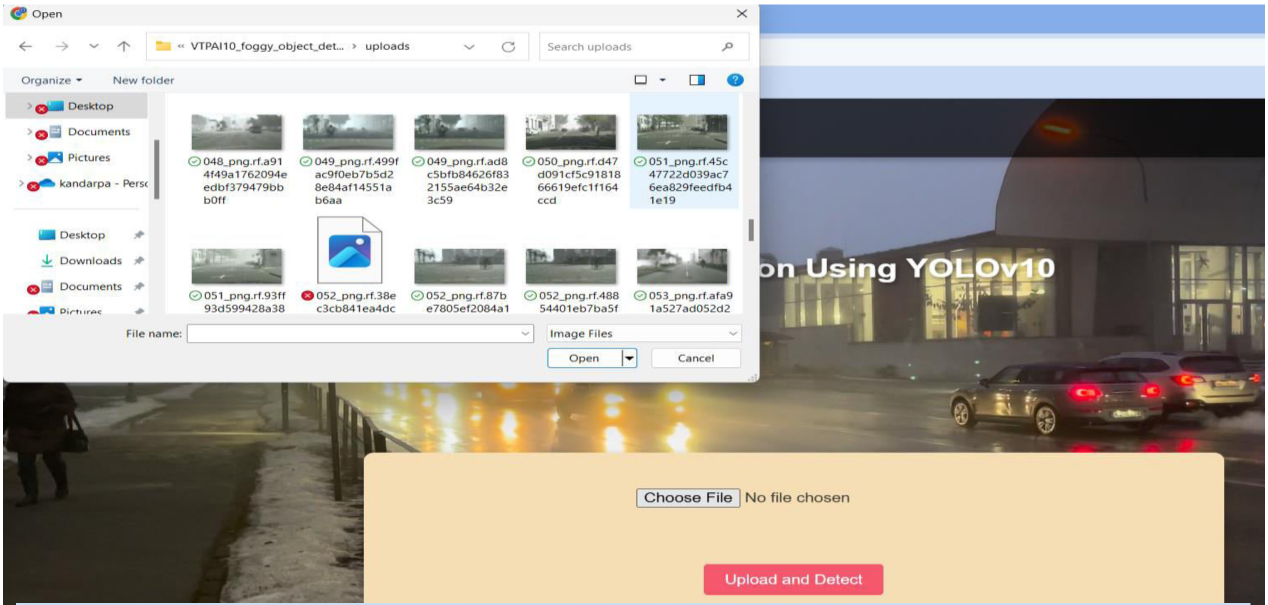
International Journal of Advanced Research in Education and TechnologY(IJARETY)

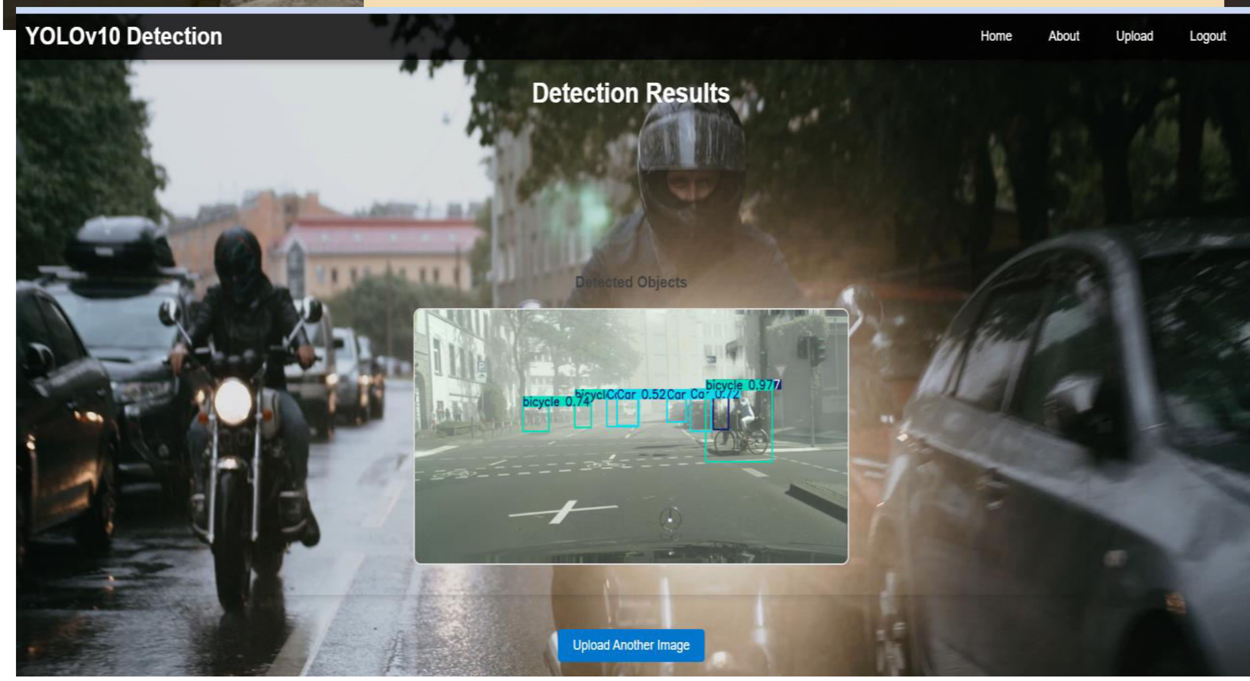
| www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203050







VII. CONCLUSION

In this study, we proposed an efficient and lightweight YOLOv10 network model tailored for traffic target detection in foggy weather. By integrating advanced modules such as DCN, Involution, and FasterNex, we successfully reduced model parameters and size while enhancing detection accuracy and performance. A novel S5attention module was introduced to improve feature fusion, and an additional small target detection layer significantly boosted the accuracy of detecting small objects and refined boundary box regression. The optimized YOLOv10 model effectively addresses the challenges posed by foggy road conditions, offering improved detection performance and computational efficiency for real-world applications

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.

IJARETY

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 Impact Factor: 8.152

www.ijarety.in editor.ijarety@gmail.com