International Journal of Engineering Research and Science & Technology

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

Research Paper

ROAD OBJECT DETECTION IN FOGGY COMPLEX SCENES BASED ON IMPROVED YOLOV10

K. VIJAY KUMAR [1], MALI AMULYA [2], PAREPALLI KAVYA[3], SHAIK KAMAL [4], EJJIGIRI JESHWANTH [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— 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 detection.

Index Terms – Vehicle detection, human detection, YOLOv10, foggy weather, object detection, image dehazing, DehazeFormer, dark channel prior, lightweight deep learning,

attention mechanism, feature extraction, adverse weather vision, intelligent transportation systems, computer vision.

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. 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 data set such 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 spatial agnostic. 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.

D. DINO: DETR with improved de noising anchor boxes for end-to-end object detection

The present DINO (DETR with Improved deNoising anchOr boxes), a strong end-to-end object detector. DINO improves over previous DETR-like models in performance and efficiency by using a contrastive way for denoising training, a look forward twice scheme for box prediction, and a mixed query selection method for anchor initialization. DINO achieves 49.4AP in 12 epochs and 51.3AP in 24 epochs on COCO with a ResNet-50 backbone and multi-scale features, yielding a significant improvement of +6.0AP and +2.7AP, respectively, compared to DN-DETR, the previous best DETR-like model. DINO scales well in both model size and data size. Without bells and whistles, after pre-training on the Objects365 dataset with a SwinL backbone, DINO obtains the best results on both COCO val2017 (63.2AP) and test-dev (63.3AP) with model size under 1 billion parameters. Compared to other models on the leaderboard, DINO significantly reduces its model size and pre-training data size while achieving better results. The code will be available.

III. PROPOSED METHODOLOGY

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.

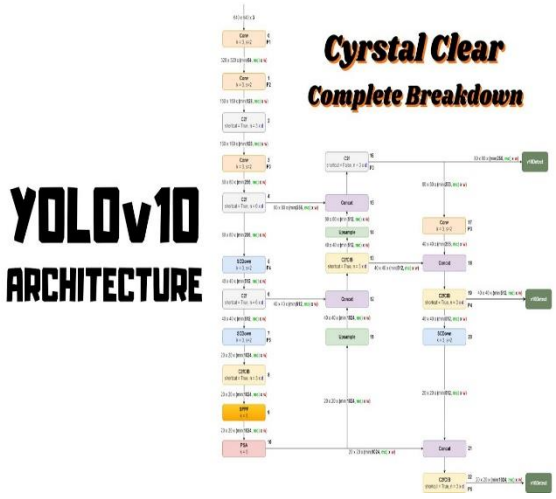


Fig. 1: System Overview

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 Label Img 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.

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.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental results show that the proposed improved YOLOv10 model achieves higher detection accuracy for vehicles and pedestrians in foggy environments than conventional methods.

The integration of DehazeFormer, dark channel preprocessing, and an enhanced attention mechanism significantly improves object visibility and feature extraction under low-visibility conditions.

The lightweight architecture also maintains fast inference speed while reducing computational complexity. Overall, the proposed framework provides an efficient

and reliable solution for real-time road object detection in adverse weather conditions.



Fig. 2: Car Detected



Fig. 3: Vehicles Detected



Fig. 4: Multi Object Detected

V. 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.

[11] M. Wang, W. Yang, L. Wang, D. Chen, F. Wei, H. KeZiErBieKe, and Y. Liao, ''FE-YOLOv5: Feature enhancement network based on YOLOv5 for small object detection,'' J. Vis. Commun. Image Represent., vol. 90, Feb. 2023, Art. no. 103752, doi: 10.1016/j.jvcir.2023.103752

AUTHORS Profile



Mr. K. VIJAY KUMAR Received my B.Tech(CSE) in 2017 from Sriindu college of Engineering and Technology (JNTUH). I completed my master of Technology (MTech), computer science in 2021 from Sreenidhi institute of science and Technology (JNTUH) With 3 years of,dificated experience in academia and Research. I currently serves as a Assistant professor in CSE Department at Gurunanak institutions Technical campus, ibrahimpatnam, Telangana, india.



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



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



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



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