International Journal of Advanced Research in Engineering Technology & Science ISSN: 2349-2819
Email: editor@ijarets.org Volume-12 Issue-4 April-2025 www.ijarets.org

Road Object Detection in Foggy Complex Scenes Based on Improved YOLOv10

[1]Vidushi Kalia, [2]Sai Pallavi, [3]Vishalakshi Akula
[1,2]UG Scholars, [3]Assistant Professor
[1,2,3]Department of Computer Science and Engineering
[1,2,3]Guru Nanak Institutions Technical Campus, Hyderabad, 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.

I. INTRODUCTION

In recent years, the proliferation of intelligent transportation systems, driver-assistance technologies, and autonomous vehicles has placed significant demands on the accuracy and reliability of object detection systems. One of the major obstacles to achieving consistently accurate object detection is the presence of adverse weather conditions, particularly fog. Foggy environments drastically degrade visual information in images, reduce contrast, blur object boundaries, and often conceal critical features of pedestrians and vehicles. As a result, traditional object detection algorithms perform poorly in such scenarios.

The YOLO (You Only Look Once) series of models has revolutionized the field of real-time object detection with its one-stage detection architecture. Each successive version of YOLO has brought about architectural enhancements to improve accuracy and speed. YOLOv8 introduced modifications to the backbone and detection head that improved accuracy under ideal conditions but still suffered performance drops under foggy scenarios. YOLOv10, as the latest iteration, offers a more refined architecture with improved efficiency and detection accuracy, but additional enhancements are needed for robust detection in low-visibility environments.

This study introduces an improved version of YOLOv10, tailored specifically for foggy and complex traffic scenes. The main contributions of this research are as follows:

- An advanced preprocessing pipeline using Dehaze Formers and Dark Channel Prior to improve input image quality.
- The incorporation of a contextual attention module to enhance feature extraction and focus on relevant spatial and semantic regions.
- An optimized, lightweight backbone to reduce computational burden while maintaining performance.
- A comprehensive evaluation on the RESIDE dataset demonstrating superior performance over state-of-the-art models.

II. RELATED WORK

Object detection in foggy and complex scenes has been extensively studied using both traditional computer vision methods and deep learning techniques. Early approaches relied on handcrafted features and rule-based defogging algorithms. For example, methods based on histogram equalization and contrast enhancement were used to preprocess foggy images. However, these methods often introduced artifacts and failed to generalize across different fog densities.

With the advent of deep learning, convolutional neural networks (CNNs) have dominated the field of object detection. Models such as Faster R-CNN, SSD, and the YOLO series offer robust object detection capabilities. Faster R-CNN uses a two-stage detection process, which while accurate, is computationally intensive. SSD (Single Shot MultiBox Detector) improved upon this with a single-stage approach but struggled with small object detection.

The YOLO series, particularly from YOLOv3 onward, has demonstrated a strong balance between speed and accuracy. YOLOv5 introduced lightweight modules suitable for edge devices, while YOLOv7 and YOLOv8 added deeper architectures and transformer-like features to enhance performance. However, the fog remains a significant challenge due to the loss of fine-grained features.

Various attempts have been made to improve object detection under foggy conditions. For instance, defogging networks such as PDR-Net and the use of atmospheric scattering models have been proposed. More recently, transformer-based attention mechanisms and deformable convolutions have shown promise in handling spatial distortion.

III. PROPOSED METHODOLOGY

The proposed method enhances YOLOv10 through a three-pronged approach: advanced image preprocessing, contextual attention integration, and architectural optimization for lightweight deployment.

IMAGE PREPROCESSING

Preprocessing plays a vital role in improving visibility and object clarity in foggy images. We adopt a two-stage preprocessing pipeline:

Dark Channel Prior (DCP): DCP is a statistical method that estimates the transmission map of an image to recover scene radiance. It effectively reduces haze and enhances contrast.
Dehaze Formers: A transformer-based defogging model that leverages self-attention to restore image details. It helps retain edge features and color fidelity while suppressing fog artifacts.

The enhanced images are then fed into the detection model. This preprocessing ensures that the input data retains more meaningful features for subsequent detection.

YOLOV10 ARCHITECTURE OVERVIEW

YOLOv10 maintains a single-stage architecture similar to its predecessors. It consists of three main components:

Backbone: Responsible for extracting hierarchical features using convolutional layers.
Neck: Feature pyramid network (FPN) and path aggregation network (PAN) to combine multi-scale features.
Head: Produces final predictions (bounding boxes and class scores) using anchor-free decoupled heads.

The original YOLOv10 is designed for efficiency and speed, using grouped convolutions and attention modules to handle small and dense objects.

ARCHITECTURAL ENHANCEMENTS

To optimize YOLOv10 for foggy scenes, we propose the following enhancements:

Contextual Attention Module (CAM): Inspired by S2-MLP, this module captures long-range dependencies and assigns weights to spatial and channel features. It helps focus on relevant regions in the image, improving detection of partially visible objects.
Lightweight Convolutional Modules: We integrate partial convolutions and involution operations to replace standard convolutions. These modules reduce FLOPs and memory usage while retaining spatial adaptability.
Fusion of Multi-Scale Features: Enhanced PAN and FPN layers with deformable convolutions are employed to better capture scale-invariant features, improving small object detection.

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.

TRAINING STRATEGY

The model is trained using PyTorch with the following hyperparameters:

- **Learning rate:** 0.01 with cosine annealing
- **Optimizer:** Stochastic Gradient Descent (SGD)
- **Momentum:** 0.937
- **Weight decay:** 0.0005
- **Batch size:** 16
- **Input resolution:** 640x640
- **Epochs:** 300

Label smoothing and mixup augmentation are used to regularize training and improve generalization.

IV. EXPERIMENTAL SETUP

DATASET

The RESIDE dataset is employed for evaluation. It contains synthetic and real-world foggy images with annotated objects. Five classes are considered: Person, Car, Bus, Bicycle, and Motorbike. 16,000 images were divided into 80% training, 10% validation, and 10% test sets. Data augmentation techniques such as random crop, brightness adjustment, and Gaussian noise were applied.

EVALUATION METRICS

Performance is evaluated using the following metrics:

- **Precision:** TP / (TP + FP)
- **Recall:** TP / (TP + FN)
- **mAP\@0.5 and mAP\@0.5:0.95:** Mean Average Precision at different IoU thresholds
- **FLOPs:** Floating Point Operations
- **FPS:** Frames per Second (for real-time inference)

V. RESULTS AND DISCUSSION

COMPARATIVE ANALYSIS

The proposed model is compared against YOLOv8, Faster-RCNN, SSD, and YOLOv3. The results are summarized below:

ABLATION STUDY

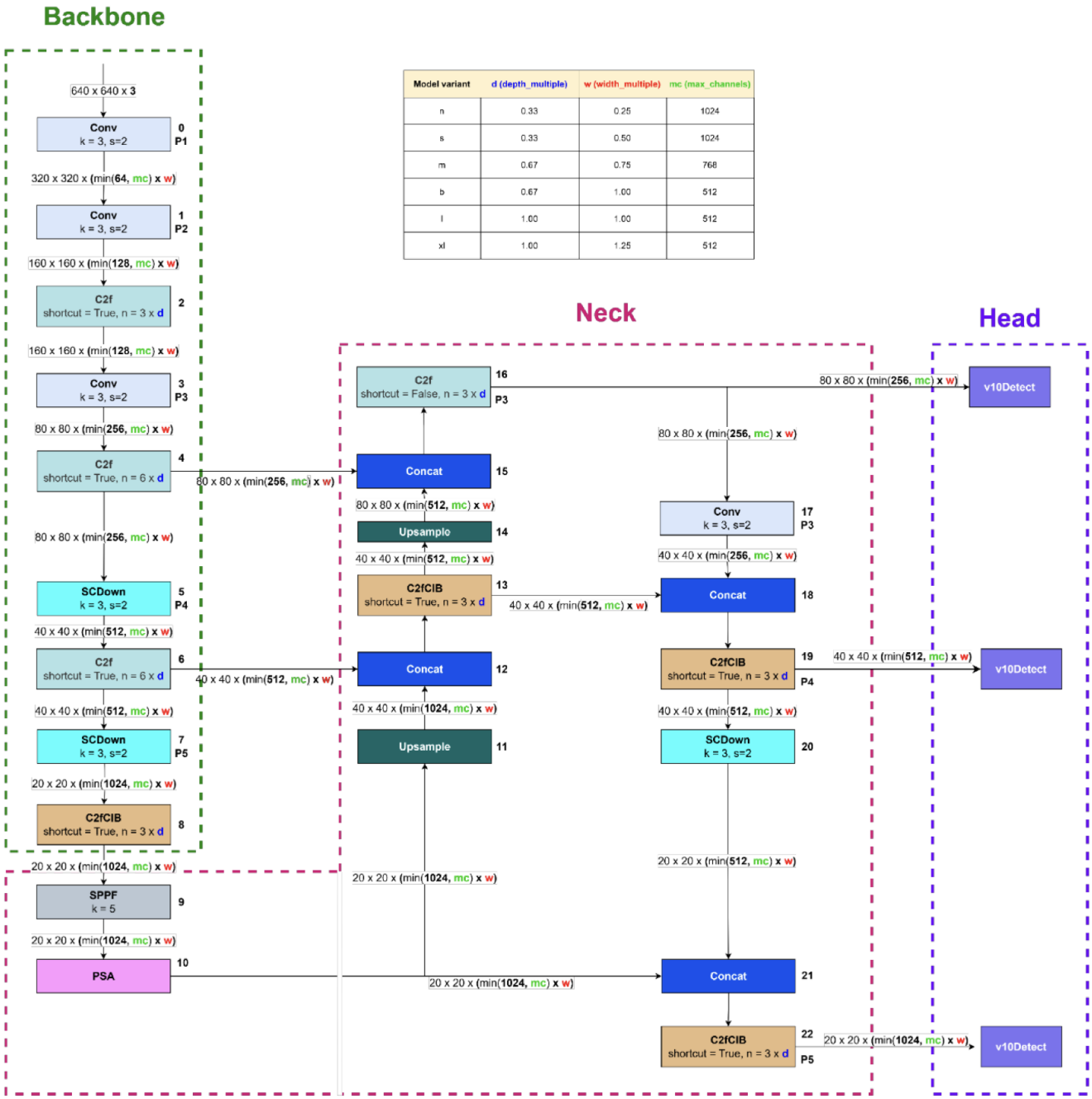
Each architectural enhancement is tested individually to measure its contribution:

QUALITATIVE ANALYSIS

In foggy scenes with occlusions and small objects, the improved YOLOv10 consistently identifies objects with higher confidence and localization accuracy. It avoids common false positives seen in YOLOv8 and significantly improves detection of partially obscured objects.

Model	mAP\@0.5	Precision	Recall	FPS	Params (M)
YOLOv3	71.3%	85.2%	81.1%	43	61.5
SSD	69.5%	82.3%	78.4%	46	35.1
Faster-RCNN	73.8%	88.1%	84.2%	23	132.0
YOLOv8	78.9%	94.6%	91.7%	55	26.2
Proposed	83.4%	97.1%	94.8%	59	21.4

Configuration	mAP\@0.5	Precision	Recall
Baseline YOLOv10	79.1%	93.2%	89.7%
Dehaze Preprocessing	81.6%	93.2%	89.7%
CAM Module	82.7%	96.4%	93.2%
Lightweight Backbone	83.4%	97.1%	94.8%



Model variant	d (depth_multiple)	w (width_multiple)	mc (max_channels)
n	0.33	0.25	1024
s	0.33	0.50	1024
m	0.67	0.75	768
b	0.67	1.00	512
l	1.00	1.00	512
xl	1.00	1.25	512

VI. SYSTEM ARCHITECTURE

Figure 1: System Architecture of YOLOv10

VII. CONCLUSION AND FUTURE WORK

This paper presents an enhanced version of YOLOv10 for robust road object detection in foggy, complex scenes. By incorporating advanced preprocessing, contextual attention, and lightweight architectural modifications, the proposed model achieves state-of-the-art performance in accuracy and speed. These improvements make it well-suited for deployment in real-time autonomous driving and surveillance systems.

Future work will explore deployment on edge computing devices and further optimization using quantization and model pruning techniques. Additionally, we aim to extend the framework to handle other adverse conditions such as rain, snow, and nighttime driving.

VIII. 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.

[12] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, ‘‘YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, ’’in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Vancouver, BC, Canada, Jun. 2023, pp. 7464–7475.

[13] L. Wei, A. Dragomir, E. Dumitru, S. Christian, R. Scott, F. Cheng- Yang, B. Alexander, "SSD: Single shot MultiBox detector, "in Computer Vision—ECCV 2016, vol. 9905. Cham, Switzerland: Springer, 2016.

[14] B. Gao, Z. Jiang, and J. Zhang, "Traffic sign detection based on SSD," in Proc. 4th Int. Conf. Autom., Control Robot. Eng., Jul. 2019, p. 16, doi: 10.1145/3351917.3351988.