www.rericjournal.ait.ac.th

Enhancing Object Detection in Adverse Weather for Autonomous Driving with YOLOv9

Prerna Saini*, [1], Anusha Dixit*, and Deepak Kumar Sharma*

Abstract – *The integration of autonomous vehicles (AVs) within the society has been a topic of interest since the 1950s when the trials on the first advanced driver assistance system (ADAS) began. The promise of autonomous driving systems is based on their ability to traverse with human-level perception, especially object detection in driving environments. This is essential for safe and reliable navigation but adverse weather conditions like sandstorms, rain, dense fog, and heavy snowfall hinder the robustness of the perception system. To address the problem of low visibility in adverse weathers like these, our proposed approach implemented feature extraction using content-based information retrieval (CBIR) for contrast enhancement and Enhanced Super Resolution Generative Adversarial Networks (ESRGAN) for image restoration. Further, YOLOv9 - the latest breakthrough in object detection was used. This work was the first initiative taken towards exploring the implementation of adverse weather driving dataset on Yolov9. The methodology was successful in achieving an average precision of 97.17% in detecting objects across all weathers with an overall mean Average Precision (mAP) of 72.87. These encouraging results might prove to be beneficial for safer and more reliable autonomous vehicle operation in diverse weather conditions.*

Keywords – **adverse weather, autonomous driving, image enhancement, object detection, YOLOv9**

1. INTRODUCTION

The sensors in an object detection system form the basis of all data extracted from the driving environment. In current ADAS systems, three types of sensors are employed for data collection, namely Light Detection and Ranging (LiDAR), Radio Detection and Ranging (RADAR) and Cameras. Among these sensors, the multiple cameras mostly present on the windscreen, on the sides and at the rear of the car are the sole source of all visual data input for processing. This data is essential in enabling AVs to identify and locate entities such as vehicles, pedestrians, traffic-lights and signs and other potential obstacles.

The images taken by the camera are divided into training and testing sets according to the preference of training employed. The training set consists of annotated images, i.e. images with bounding boxes drawn around objects that are crucial for detection. The bounding boxes coordinates signify the edges of an object which are noted as labels for the specific image that serve as the ground truth for supervised learning.

You Only Look Once (YOLO) is a fully convolutional neural network that anticipates bounding boxes and class probability confidence scores immediately from entire input image dataset in one screening. YOLO models sport a light-weight architecture and have lower computational costs in comparison to other object recognition models. This allows them to run on devices with moderate graphics processing unit (GPU) capabilities. Another lucrative feature of YOLO models is that unlike faster Region-Based Convolutional Neural Network (R-CNN) that requires separate stages for region proposal and classification, YOLO performs both tasks in a single forward pass. This makes it more efficient for real time applications.

While YOLOv9 is an avant-garde advancement in the field of object recognition, its ability to perform under adverse weather conditions remains largely untested [1]. Study on autonomous driving in adverse weather becomes a matter of importance due to the introduction of noise in the data. In heavy rainfall, the pixel intensity decreases hence, blurring the object outline. Similarly, in dense fog, the overall image contrast and visibility is reduced due to hazy and opaque environments. In the case of sand-storms, the dust clouds result in scattering of light due to which accurate bounding box predictions become difficult. During heavy snow conditions, visibility is drastically reduced and reflection of sunlight introduces glare in the inputs. Other than the qualitative degradation of images, suspended particles stick to the camera lens resulting in occlusions and low range vision.

Thus, image enhancement is crucial for restoring temporal information in the images for better object recognition. Temporal information can help identify moving objects, e.g. vehicles, pedestrians; understanding object interactions, actions, and behaviors, e.g. cars stopping in front of pedestrians. Image enhancement techniques can help restore the subtle details within an image that might have been lost by filling them with relevant information. Some of these methods can be noise reduction, image sharpening, contrast enhancement and increasing pixel resolution. While synthetic datasets offer a controlled environment

*Department of Information Technology, Indira Gandhi Delhi Technical University for Women, Kashmere Gate, Delhi, 110006, India.

[1] Corresponding author;
E-mails: prernas2001@gmail.com; anushadixit03@gmail.com; dk.sharma1982@yahoo.com

for model training, they struggle with real-world driving conditions as they fail to handle variations in light scattering, reflections and occlusions. Hence, the proposed model leverages a real-world adverse weather autonomous driving dataset, Detection in Adverse Weather Nature (DAWN) for enhancing the generalizability and robustness in actual driving conditions. It consists of 1000 images across four different weathers- rain, fog, snow, sandstorm with different traffic flow in varied driving environments (urban, highway, densely populated, sparsely populated) [2].

The purview of this paper can be encapsulated as follows:

1. Achieving all-weather autonomy to broaden the operational scope of autonomous vehicles and to develop a comprehensive pipeline of processed images that is input into YOLOv9 for robustness.
2. Developing a content based feature extraction method for image enhancement to improve visibility in various weather conditions.
3. Enhancing object detection by improving mAP and precision for autonomous vehicles in rain, fog, sandstorms, and snow.
4. Lowering computational costs of pre-processing dataset and detection model by using light-weight YOLOv9 model.

The further sections of the paper are structured as follows: Section 2 presents an overview of the related works in the field of ADAS safety and object detection along with image enhancement techniques. In Section 3, implementation of the proposed methodology is presented and Section 4 discusses its results obtained. Section 5 focuses on the prospects of future work and conclusion drawn.

2. RELATED WORKS

This section discusses the relevant works that happened in the field of autonomous driving, methodologies employed and their direct impact on the present research scenarios. As discussed earlier, hostile weather conditions like heavy rain, fog, sandstorms, snow creates situations where object detection is unreliable, such as water reflections on road and glare due to snowy driving environments which can be detrimental to the efficiency of ADAS.

Research has been conducted on simulating uncertainties in autonomous driving such as adverse weather and road conditions. A safety score was calculated using Response Surface Method (RSM), that kept in account the uncertainties which were compared from the safety threshold. This was a robust method for ensuring safe driving [3]. Thus, the need to increase the effectiveness of the sensors was realized which could work better in harsh weather. Another work showed the effects of weather conditions on several sensors including cameras [4]. In autonomous electric vehicles (EVs) fusion of different sensors like cameras and LiDAR showed promising results in feature recognition of 3D object detection [5].

Despite the many advancements made to increase sensor quality, they had their physical limitations. Thus, for better efficacy, making improvements to the ADAS system depended on the enhancement of image processing methods or on improving the object detection model. A paper mentioned testing different systems on traffic monitoring and tracking tasks in bad weather to analyse their performance. It also highlighted that despite the good performance of YOLOv5x, it struggled in multiple combinations of bad weather. Hence, it became important to introduce weather-aware models for handling different environments [6].

A survey on a custom-based dataset evaluated various deep learning object detection models like R-CNN, Faster R-CNN and YOLO. It also showcased the need to improve their architectures for marking out the arena of their interest in an image [7]. In another work, DarkNet framework was used within tiny YOLOv4, which achieved 80% accuracy in simulation of autonomous EVs [8]. Nevertheless, the various improvements made in deep learning-based object detection models could not completely tackle the noise introduced by adverse weather variables. Thus, image enhancement could be a reliable pre-processing method. Some previous works discussed the various techniques of enhancing images degraded due to weather conditions. The methods used included de-hazing, de-noising, contrast enhancement and altering Red-Green-Blue (RGB) component channels in images [9]. The Reinforced Image-based Object Detection (RIOD) YOLO model aimed at optimizing image quality by estimating illumination and transmission factors. It reported better results when implemented on YOLOv4 and YOLOv5 models for both images and videos [10]. Some deep learning based image processing techniques also showed break-through results in image processing [11]. Another work used a novel unsupervised auto encoder technique for extracting features from images. This stemmed from the idea that while detection models work well in normal weathers, their accuracy reduced drastically in weather conditions that they were not trained for [12].

Other than machine learning models, Generative Adversarial Network (GAN) architectures were also used for enhancing object recognition through two major applications- restoration of lost temporal information due to adverse weather and increasing resolution [13]. Adding to these advancements in the field of automation and detection, the work done in this paper was built leveraging the existing ESRGAN model with proposed CBIR. It handled different weathers causing image degradation by targeting features like pixel intensity, number of edges and their standard deviations to find similar images in the dataset. Images with high temporal similarity were processed in the same manner.

3. PROPOSED METHODOLOGY

The proposed methodology consists of the following sections: Section 3.1 discusses about the dataset used. Section 3.2 mentions the contrast enhancement technique using CBIR and Section 3.3 focuses on the enhancement of image resolutions using ESRGAN. In Section 3.4, the paper discusses the method used for object detection, namely YOLOv9.

3.1 Dataset

The dataset used in this work, was selected keeping in mind the variable factors introduced during adverse weather conditions that could affect the perception ability of cameras. Hence, after careful considerations, DAWN dataset was found to have greater similarities with real-world driving conditions. It consisted of 1000 images demonstrating four adverse weather scenarios of heavy fog, rain, snow and sandstorms extracted from Mendeley repository. The different traffic flows in diverse driving environments (urban, highway, densely populated, sparsely populated) helped in covering a wide range of training and testing scenarios for model generalization.

This dataset consisted of labelled ground truths that were adjusted manually with RoboFlow [14] to suit the requirements of tight bounding boxes around objects of significance in training. This ensured more accurate localization of objects and reduced the chances of including irrelevant background pixels, thereby reducing the number of false positives.

The annotations consisted of the following number of ground truth boxes in the dataset:

Table 1. DAWN dataset distribution.

Classes	Sand	Fog	Rain	Snow
Car	1751	1024	1349	1306
Bus	57	53	16	23
Truck	199	96	196	87
Pedestrian	188	88	28	138
Motorcycle-Bicycles	68	34	4	2

A dataset unbalanced to this extent needed several augmentation techniques to balance it for preventing any case of under-fitting or overfitting that could impact the precise and accurate decisions of the model. The various augmentation methods used were horizontal flip, rotation of minus 15 and plus 15 degrees and mosaic data augmentation. Additionally, pre-training of the YOLOv9 model was carried out using the MS-COCO dataset comprising of 80 classes. It contained 2,500,000 annotated instances over 80 classes. The pre-trained weights of MS COCO were then used for transfer learning of the model on DAWN dataset. The train-validation-test split ratio used in this was 60:25:10. The Dense-Haze benchmark dataset released by New Trends in Image Restoration and Enhancement (NTIRE) consisted of 55 sets of hazy and their complementary ground truth images of various combinations of scenes employed in transfer learning of ESRGAN [15].

3.2 Contrast Enhancement with Similarity Grouping

Classifying images based on their visual statistical data facilitated the necessary similarity grouping of images based on their spatial and spectral information. This information was further leveraged to apply selective contrast enhancement techniques based on their requirements. CBIR queried the DAWN dataset for fetching visually similar images by comparing their values [16]. The feature extraction processing was done by studying the different color intensities involved in the RGB channel of each pixel; the overall color distribution, standard deviation of pixel intensities and detection of color spaces to identify the boundaries of different objects in a particular image.

These relationships were further extended to determine the various color spaces present within a pixel in an image. Different weather images needed different techniques within their color channels to be employed to suit their needs of a clear, saturated and defined boundary formation of objects. Heavy snow images with more white light and low contrast made it difficult to distinguish background snow from snow covered objects leading to loss of temporal information. Changes were made in the L component of the L*a*b* color space to redistribute the brightness level based on its mean and standard deviation values. The challenges of preserving the original colors and converting the details back to RGB color channels were tackled by defining the lower and upper bounds for contrast stretching and applying histogram equalization, Contrast Limited Adaptive Histogram Equalization (CLAHE) and unsharp masking to define and sharpen the object boundaries.

The yellowish and orange tints present in the sandstorm images which reduced the visibility of present objects was lowered while preserving the colors and enhancing the image side by side. This was implemented by changing the b* channel in the L*a*b* perceptual channel which was responsible for the yellow color. This combined with the unsharp masking made the images visually perceivable. The low-light hazy and misty images obtained from the DAWN dataset of rain and fog weathers were enhanced with the modification of standard and perceptual channels, namely RGB and L*a*b*. The process involved controlling the L* component in the L*a*b* channel while equalizing the RGB channels independently to spread out the pixel intensities. These values were calculated for each pixel based on the cumulative distribution function (CDF).

The different experiments in the color channels helped in finding the optimal balance of contrasts needed for image processing to make the objects more recognizable for object detection. The processed images were free of noise and hence, were suitable for feeding into ESRGAN for resolution enhancement.

3.3 ESRGAN

The images retrieved after contrast enhancement were then further fed into the ESRGAN for increasing the resolution of the image and restoring lost temporal information in them. The GAN architecture consisted of a generator which attempted at restoring image from noisy pixels and a discriminator that tried to tell enhanced image apart from the real image. Thus, as the training progressed, the generator got better at reducing gaps between the ground truth image and generated images, hence producing more realistic and natural textures [17].

As training a computationally expensive GAN architecture could be quite expensive, transfer learning was performed on a smaller 55 image dataset, namely Dense-Haze dataset. The model was thus fine-tuned to the requirements of the DAWN dataset testing. To reduce any existing noise in the dataset and to avoid huge computational expenses due to processing images with very high resolution; the testing dataset was manually down-sampled by a factor of 2. This down-sampled dataset was further fed into the ESRGAN model. The GAN model up-sampled the image by a factor of 2 thus, restoring it back to its original resolutions while also simultaneously enhancing its detail and quality.

The enhanced images were de-hazed and had less noise while also maintaining the object boundaries. This contextually rich and noise-free dataset was then ready to be input as a training dataset for the object detection model.

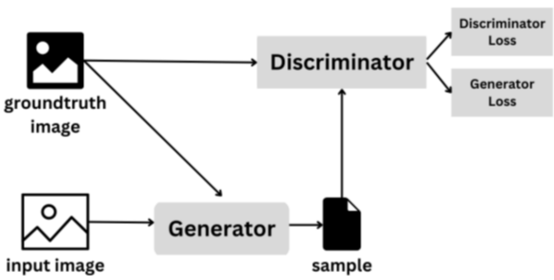


Fig. 1. ESRGAN architecture.

3.4 YOLOv9 Object Detection

The YOLO family of models, unlike other object detection algorithms, takes a single shot at an image rather than analyzing images in parts. The YOLO models offer a range of applications especially in real-time implementation such as traffic monitoring, self-driving cars, and inventory management. The several YOLO variants offer a trade-off between speed and accuracy. The higher variants have a more complex architecture, hence providing more efficient results while the lower variants with simpler models provide higher speed of training and inference. YOLOv9 surpassed its predecessors in terms of robustness.

YOLOv9 addressed the problem of information bottleneck where crucial information got lost in the earlier stages of training which led to loss of gradient. Deeper models had higher probabilities of information getting lost due to a greater number of layers. This could be tackled by introducing a higher number of parameters but it also added to the computational complexity of the model. Hence, the YOLOv9 model employed the use of reversible functions to recover the gradient loss, thus maintaining data consistency.

Architecture: The architecture of YOLOv9 can be broadly divided into 2 main modules:

1. Generalized Efficient Layer Aggregation Network (GELAN): It is a light-weight network architecture designed to optimize inference speed, parameters and computational complexity.
2. Programmable Gradient Information (PGI): This module tackles the previously mentioned issue of information bottleneck. PGI introduces a new auxiliary branch in the architecture which ensures that correct gradients are used to update layer weights in the network. Hence it assists in preserving important information for efficient object detection.

YOLOv9 has four variants i.e., v9-Small, v9-Medium, v9-Compact, v9-Extended with each variant having more parameters than the previous one. The proposed methodology employed YOLOv9-C as the primary model. MS-COCO dataset having 80 object classes was used for pre-training the model. Transfer learning was then performed using the pre-trained YOLOv9 weights to generate an adjusted weights file which was fine tuned to suit the proposed model. The model was further validated and tested to generate results.

To fine tune the model, hyperparameter tuning was done on the model. Hyperparameters involved in the training were, learning rate: lr0=0.01, weight_decay=0.0005, box loss threshold: box=7.5, class loss threshold: cls=0.5, distribution focal loss threshold: dfl=1.5, Intersection over Union threshold (IoU): iou_threshold=0.5, mosaic_aug=1.0, mixup=0.15, epochs = 100, batch_size = 16, confidence threshold: conf = 0.4.

The computational prowess of YOLOv9 is unlike its preceding YOLO variants. Compared with YOLOv7, YOLOv9-C has 42% less parameters and 22% less calculations, but achieves the same average precision (53%).

Figure 2, illustrates the implementation of 3 phases of the proposed methodology and its data flow through the pipeline, in the order, contrast enhancement with similarity grouping, ESRGAN and YOLOv9 object detection.

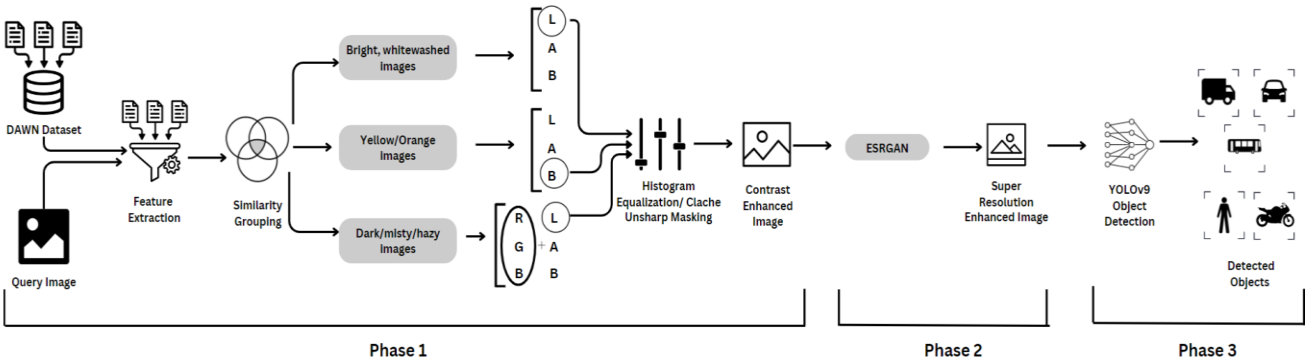


Fig. 2. Proposed model phases (a) contrast enhancement with similarity grouping, (b) ESRGAN, (c) YOLOv9 object detection.

4. EXPERIMENTS AND RESULTS

4.1 Experimental Setup

The stated methodology was executed using Python 3 as the main programming language. The deep learning networks were trained using TensorFlow and PyTorch. Libraries like OpenCV and Pillow were employed in image pre-processing. The further system requirements used in training and testing the model were as follows: Graphical Processing Unit (GPU) with CUDA cores support using Google Colab (Intel Xeon CPU@2.20 GHz), 16 GB RAM, a Tesla K80 accelerator and 12GB GDDR5 VRAM. The mosaic data augmentation technique was closed for the last 15 epochs to enhance the generalizability of the model and mAP evaluation metric was chosen as the main metric for object detection

4.2 Evaluation Metrics

As the field of object detection keeps developing new benchmark models, it became extremely important to be able to quantify the results achieved from them. This documentation was essential in many ways, such as, determining the health of the dataset used; accuracy of parameters used for training the model; whether the model generalized to the dataset and most importantly if the model performed accurate object detection. The importance of defining the correct metrics as a method for evaluating the YOLO model in the scenario of autonomous driving was even more crucial. In autonomous transport systems, guaranteeing the security of passengers as well as pedestrians and other vehicles on the road was of utmost importance. Hence, the proposed model was closely monitored and evaluated against the three most important evaluation metrics for an object detection system i.e., precision, accuracy, and mAP.

Precision is the ratio between the number of positive samples accurately classified to the total number of samples classified as positive, irrespective of whether they are correctly or incorrectly identified. A higher precision reflects the ability of the model to make a higher number of truly positive predictions as compared to false positive predictions. This implies a higher probability of a model to correctly classify detected objects with their respective ground truths.

Hence a higher precision is considered as an important factor for labeling a model as reliable.

Accuracy is the measure of how well a model performs across all classes in an object detection model. It is an evaluation metric of high importance especially for a model where all classes are of equal importance. It is the ratio of the number of correct predictions to the total number of predictions made. Accuracy might be deceptive when the dataset is heavily imbalanced leading to wrong perception of model's efficiency.

Average precision is the area under the precision-recall curve. The mean of average precisions across all classes in a particular model is defined as the mAP score of the model. In YOLO variants, mAP score is the main performance metric. It is calculated over a particular IoU threshold. This threshold is an indicator of the minimum overlap over ground truth and predicted bounding boxes for a prediction to be considered a success.

4.3 Computational Metrics

Besides the metrics that evaluate the model's performance as an efficient object detector, computational metrics form a central point of assessing how well the model utilizes the available resources during training and testing. Some of the major computational metrics important for evaluating the model are time taken for training, inference time, number of parameters and throughput generated.

Training time is the total time taken for a model to train on the input dataset. While higher training time might also be an indicator of better resources like high performance GPU availability or a relatively larger dataset, but when tested on similar systems and same dataset, a model with less training time indicates more efficient allocation of computational resources whereas inference time is defined as the least time taken by a model to make inferences about a single data point. Smaller inference time indicates that faster predictions are made. This is especially essential in resource constrained devices. Training time is typically measured in minutes or hours and inference time is calculated in milliseconds or seconds per image or sample.

When training a model, it involves several trainable parameters represented in its architecture. These parameters are learned during training. While a higher number of parameters could imply better model capacity, it also implies that it requires more data and higher computational resources. It also means that the model could potentially be suffering from overfitting.

Throughput metric combines batch size and inference time to represent the model's processing speed for multiple data points. It is expressed as frames per second (fps). Higher throughput implies a more robust object detection model.

4.4 Results and Discussions

This work was a first initiative towards analyzing the performance of adverse weather driving DAWN dataset in assessing the performance of YOLOv9 and our proposed methodology. Hence, the DAWN dataset was first trained and tested on the original YOLOv9-C framework without any pre-processing techniques to understand the base results which were achieved. The mAP score of 65.26 achieved during the initial phase indicated the possible chances and potential of improvements which could have enhanced the results, if some image pre-processing methods were additionally used.

By enhancing the resolutions of dataset images, it was noticed that they had sharper edges and the textures appeared more natural. The output images also had increased perceptual quality which would directly reflect as better results for object detection. On training the YOLOv9 model with the resolution enhanced images, an overall accuracy of over 80% was recorded.

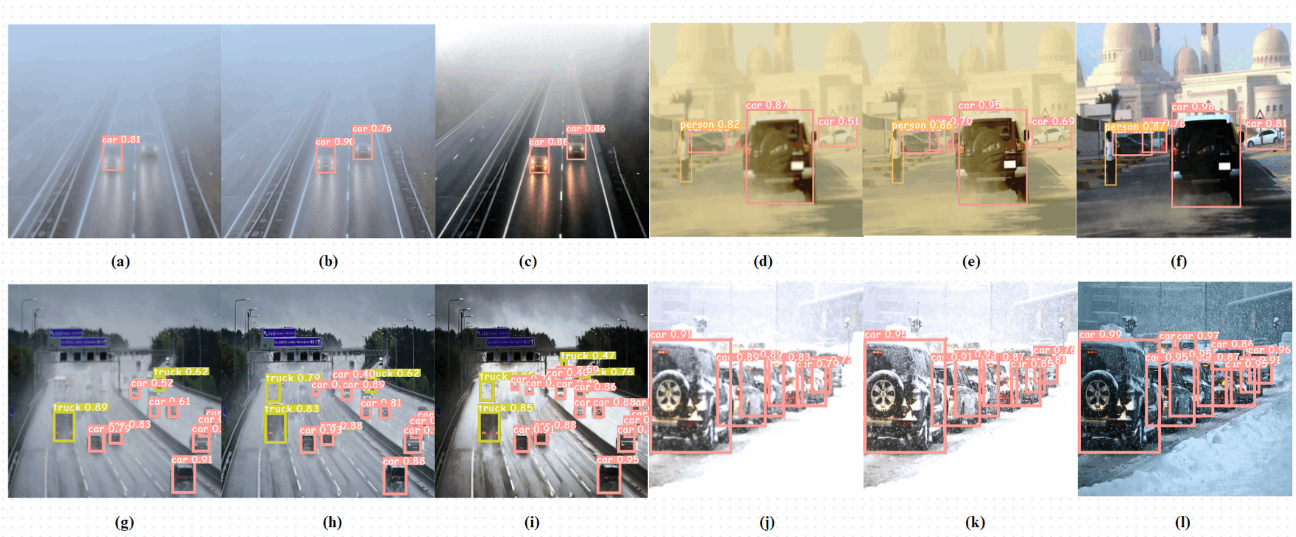


Fig. 3. Comparison of object detection using (a) only YOLOv9, (b) YOLOv9 + ESRGAN, (c) proposed methodology in dense fog; (d) only YOLOv9, (e) YOLOv9 + ESRGAN, (f) proposed methodology in sandstorm; (g) only YOLOv9, (h) YOLOv9 + ESRGAN, (i) proposed methodology in heavy rain; (j) only YOLOv9, (k) YOLOv9 + ESRGAN, (l) proposed methodology in heavy snow.

As can be referenced from the Figure 3 (a), (d), (g), (j) the raw images obtained from DAWN dataset had noise and blurriness. Due to the haze and dust particles present in the environment, visibility was reduced to such levels that the perception of objects which were farther away from the camera was hardly distinguishable. This resulted in low number of detections as compared to the ground truth and low confidence scores for the objects that were successfully detected.

After passing them through ESRGAN and then performing object detection the images present in Figure 3 (b), (e), (h), (j) looked almost like the images in Figure 3 (a), (d), (g), (j), but had better defined object boundaries and higher confidence scores with more precision. However, while preserving the original texture of the input images, there was an increase in the level of smoothing. Moreover, the resolution increment was not sufficient to tackle the issues due to haze and noise present in the color channels of the image.

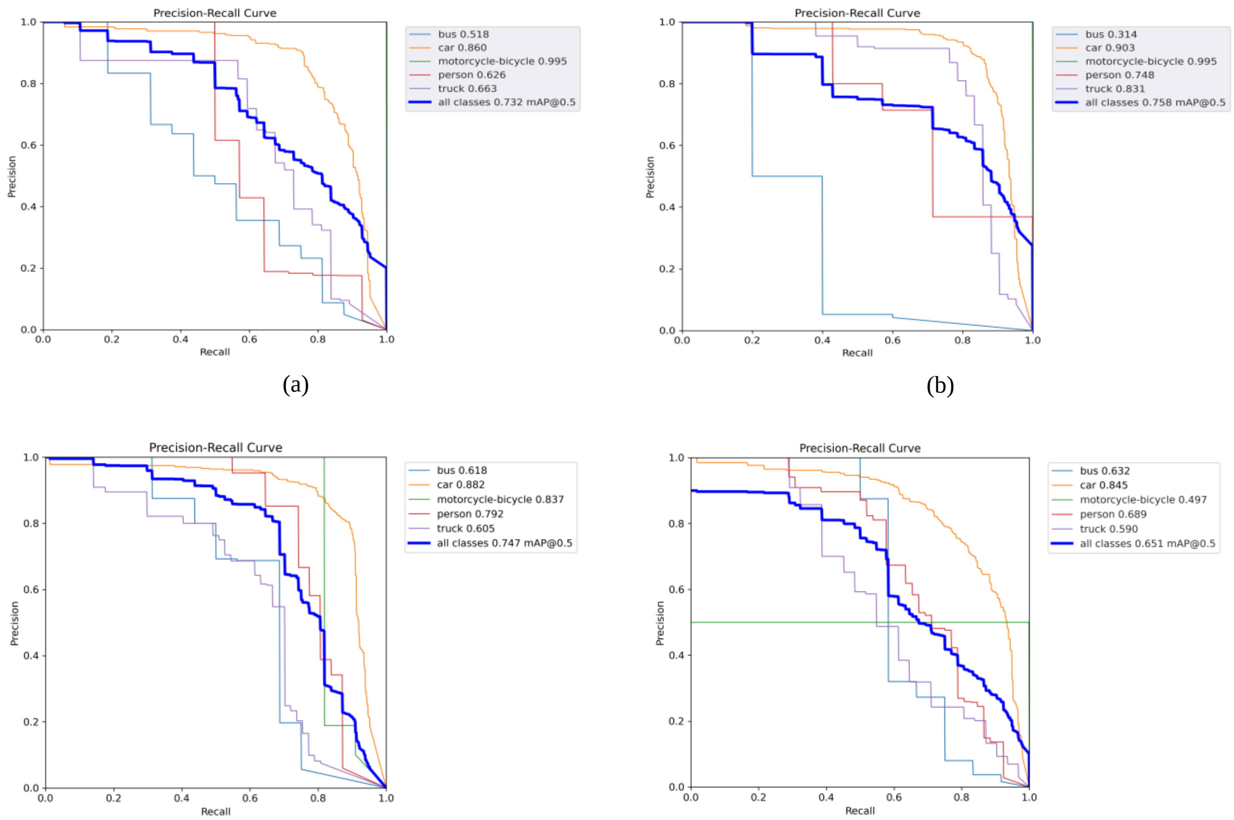
To accomplish even better results, similar images were contrast enhanced. Instead of simply altering all the images in a dataset with the same contrasting methods, images with similar range of information were grouped based on certain characteristics like color and intensity or spatial similarity. They were then processed with techniques employed in our proposed methodology for generation of saturated and color enhanced images as can be referenced in Figure 3 (c), (f), (i), (l). These results, when fed into YOLOv9 for object detection, recorded impressive results as compared to original results. While the accuracy was like that of ESRGAN enhanced images, the precision and mAP values were significantly better.

The mAP values records for all four weathers as shown in Figure 4 (a), (b), (c), (d): fog, rain, snow, and sandstorms were recorded to be 0.747 for sandstorms, 0.732 for foggy conditions, 0.758 for rain and 0.651 for snowy weather. Furthermore, the class wise mAP of all five classes was found out to be 0.8725 for cars, 0.6722 for trucks, 0.5205 for bus, 0.7137 for pedestrians, 0.831 for motorcycles and bicycles.

Table 2. Results comparison.

Metric	YOLOv9	YOLOv9 + ESRGAN	Proposed Model
Accuracy	73.45	80.03	80.01
Precision	94.01	96.27	97.17
mAP	65.26	70.55	72.87

The object detection model used also had an edge over previous models as it had a lesser number of parameters, making it comparably robust. The YOLOv9-C version used in the paper had only 25.3M parameters as compared to YOLOv7 that had 36.9M parameters. The YOLOv9-C model had 102.1G FLOPs compared to YOLOv7 which had 104.7G FLOPs. Thus, it can be concluded that the model had a potential to become a more light-weight and computationally affordable alternative to its predecessors. The average time taken for training the model on YOLOv9 was recorded to be 0.330 hours whereas the same took 0.538 hours for training on the YOLOv7 model.



(a) (b) (c) (d)

Fig. 4. mAP results for (a) foggy weather, (b) heavy rain, (c) sandstorm, (d) heavy snow.

5. CONCLUSION AND FUTURE SCOPE

The research explored the potential of combining feature similarity grouping with ESRGAN to improve object recognition in difficult weather conditions for self-driving vehicles. It was noticed that a combination of targeted contrast enhancement and upscaling reveals more crucial details abstracted in the image due to noise. This was especially critical for adverse weather conditions where safety and reliability were of utmost importance. The common issues faced in object detection in bad weather conditions included light scattering, haze and low range vision. While the credibility and efficacy of deep learning-based algorithms for image enhancement have gone through rigorous testing, they might introduce a bias if the dataset is not diverse leading to poor performance in real-world scenarios. Hence the method of non-deep learning technique was implemented. It is also worth mentioning that a robust state-of-the-art object recognition algorithm, YOLOv9 was used in the proposed model. The approach of implementing an adverse weather dataset on YOLOv9 was a novelty. While YOLOv9 is still in its primitive stage, this work showed promising results in the field of object detection while leveraging its reversible function to reduce gradient loss. Hence, this paper aimed to reach human like perception levels in self-driving vehicles while navigating through challenging driving environments.

While significant contributions were made, there remains room for further research. The computational costs saved in the operation of this methodology can also serve as the basis of autonomous driving systems with resource constraints like EVs where efficient battery consumption is an important function of operation. The paper further initiates more research to be done in the field of conventional image enhancement techniques with methods like similarity-based contrast enhancement as used in this work. This can ensure ADAS systems to be better equipped to deal with unforeseen driving conditions. This work can be further expanded to video enhancement technologies. It can serve as the turning point in the field of object detection as videos store more temporal information regarding driving scenes than images. Also, testing on real-world video clips can be a better way of analyzing the model's efficacy in real-world driving scenarios. Research on image enhancement methods that are unsupervised is another topic that needs more future work. This can enable models to effectively learn from data that is unlabeled or has very few labels which can make them more practical even for real-world usage. Much work also needs to be done on developing methods that can better explain the reasoning behind decisions taken by the ADAS. Trust is an important factor when it comes to autonomous vehicles hence work must be initiated in the direction of building white-box methods to build consumer trust.

By addressing the various challenges and future studies mentioned above, it can be concluded that image enhancement object detection models have the potential to bring significant improvements in the fields of self-driving vehicles that operate in dynamic and difficult weather conditions.

REFERENCES

[1] C. Yao Wang, I-H. Yeh, H.-Yuan M. Liao, 2024. "YOLOv9: Learning What You Want to Learn Using Programable Gradient Information", Available open access Online at Cornell University, New York, USA, 29 Feb: 1-18. DOI: https://arxiv.org/pdf/2402.13616.

[2] Kenk M.A. and Hasaballah Mahmoud, 2020 “DAWN: Vehicle Detection in Adverse Weather Nature Dataset,” Available open access Online at Cornell University, New York, USA, 12 Aug: 1–6. DOI: https://arxiv.org/abs/2008.05402.

[3] Chen F., and Chen S., 2011. “Reliability-based assessment of vehicle safety in adverse driving conditions,” *Transp. Res. Emerg. Technol.(19)*, Feb. 2011: 156–168.

[4] Appiah Emmanuel Orwusu and Mensah Solomon, 2024, “Object detection in adverse weather condition for autonomous vehicles,” *Multimedia Tools and Appl.* 83 (9): 28235–28261.

[5] D. Dai, Z. Chen, P. Bao, and J. Wang, 2021. “A review of 3d object detection for autonomous driving of electric vehicles,” *World Electr. Veh. J.*, 12 (3).

[6] Patel Vatsa S., Agrawal K., and Nguyen Tam V., 2024. “A Comprehensive Analysis of Object Detectors in Adverse Weather Conditions,” in *58th Annual Conference on Information Science and System (CISS)*, 13-15 March: 1–6.

[7] Abbas Syed Mazhar and Dr. Singh Shailendra N., 2018. “Region-based Object Detection and Classification using Faster R-CNN,” in *4th Conference on Computational Intelligence & Communication Technology*, IEEE, 9-10 Feb: 1–6

[8] I. V. Fanthony, Z. Husin, H. Hikmarika, S. Dwijayanti, and B. Y. Suprapto, 2021. “YOLO Algorithm-Based Surrounding Object Identification on Autonomous Electric Vehicle,” *Int. Conf. Electr. Eng. Comput. Sci. Informatics*, October 2021: 151–156.

[9] Liu W., *et al.*, 2022 “Image-Adaptive YOLO for Object Detection in Adverse Weather Conditions,” in *Proceedings of the AAAI Conf. on Artificial Intelligence, volume 36,* 28 Jun: 1792–1800.

[10] Pavithra P. P., R.K. Bhanu, and Safinaz S., 2024. “RIOD:Reinforced Image-based Object Detection for Unruly Weather Conditions,” *Engineering Technology Application Science Research*, 14 (1), Feb 2024: 13052–13057.

[11] Bachute M.R. and Subhedar J.M., 2021. “Autonomous Driving Architectures: Insights of Machine Learning and Deep Learning Algorithms,” *Machine Learning with Appl.*, volume 6, Sep 2021: 100-164.

[12] Nguyen V. D., *et al.* 2020. “Robust Vehicle Detection Under Adverse Weather Conditions Using Auto-encoder Feature,” 2020 *International Journal of Machine Learning and Computing*, 10 (4), Jul 2020: 549–555.

[13] Fabio Cuzzolin, Andrew Bradley, Valentina Mușat, I. Fursa, P. Newman, 2021 “Multi-weather city: Adverse weather stacking for autonomous driving,” in *2021 International Conf. on Computer Vision Workshops (ICCVW)*: 2906–2915.

[14] Roboflow. Available online: https://roboflow.com/

[15] Ancuti C. O., *et al.* 2019. "Dense haze: A benchmark for image dehazing with dense-haze and haze-free images", in *IEEE International Conference on Image Processing (ICIP)*, Sep. 2019: 22-25.

[16] Sarath C. Yenigalla, Rao K. Srinivasa, and Ngangbam P. S., 2023. “Implementation of Content-Based Image Retrieval Using Artificial Neural Networks,” in *Engineering Proceedings*, 13 March, 2023.

[17] X. Wang, K. Yu, *et al.* 2018. “ESRGAN: Enhanced super-resolution generative adversarial networks,” in *Proceedings of the Europen Conf. on CV Workshop*, Germany: 63–79.