RIOD:Reinforced Image-based Object Detection for Unruly Weather Conditions

P. P. Pavitha
Department of Electronics & Communication, Presidency University, India
pavithapp@gmail.com (corresponding author)

K. Bhanu Rekha
Department of Electronics &Communication, Presidency University, India
bhanurekha@presidencyuniversity.in

S. Safinaz
Department of Electronics &Communication, Presidency University, India
safinazs@presidencyuniversity.in

Received: 30 November 2023 | Revised: 26 December 2023 | Accepted: 29 December 2023

 DOI: https://doi.org/10.48084/etasr.6689

ABSTRACT

Deep Neural Network (DNN) object detectors have proved their efficiency in the detection and classification of objects in normal weather. However, these models suffer a lot during bad weather conditions (foggy, rain, haze, night, etc.). This study presents a new scheme to reduce the aforementioned issue by attenuating the noise in the input image before feeding it to any kind of neural network-based object detector. In this study, the image optimization function transforms subpar-quality images due to bad weather into pictures with the optimal possible quality by estimating the proper illumination and transmission function. These optimized images showed improved object detection rates in the YOLOv4 and YOLOv5 models. This improvement in object detection was also noticed in the case of video input. This scheme was tested with images/videos from various weather conditions, and the results showed an encouraging improvement in detection rates.

Keywords-self-driving vehicle; YOLOv4; YOLOv5; image pre-processing; deep learning; object detection

I. INTRODUCTION

Object recognition with bounding boxes mentioning its class has numerous implementations in robotics, self-driving vehicles, drones, advanced driver assistant systems, etc. Current algorithms are good at handling camera inputs and identifying multiclass objects in both image and video. Recent studies have shown that these detectors severely underperform in challenging weather conditions, such as snow, rain, fog, etc. This performance decrease is ascribable to the subpar quality image input from the cameras due to reduced illumination and increased noise. In the case of autonomous vehicles, input images or video sequences can have sizeable variations in quality due to the type of camera used, its position, as well as the environmental conditions. Contemporary object detectors based on neural networks are trained and tested in general with appropriately illuminated images. The performance of these object detectors suffers greatly in challenging weather conditions due to low-quality image input. Unfortunately, this is the real-time scenario in most automated applications. In self-driving applications, training a Convolutional Neural Network (CNN)-based model with all possible environmental conditions and scenarios is nearly impossible. Figure 1 shows a few samples of bad situations that can be encountered while driving.

Global research on addressing this issue has been streamlined into three major strategies, which are sensor fusion, image preprocessing, and domain adaptation. The conventional approach of depending only on the input from the camera has been shown to be inadequate in bad weather conditions. Sensor fusion is another strategy to improve object detection. In [1], the fusion of a camera, lidar, and radar was used to complement each other for object detection in the most possible scenarios of driving. The major setback of this strategy is the scarcely available datasets using this combination of sensors. Furthermore, incorporating multiple sensors into an automobile is not cost-effective and can spoil its aesthetics. The change in environmental conditions can be considered a domain shift. A few models that rely on domain adaptation have been proposed for object detection at critical points. Such models demand enormous training datasets, and the annotation of such data is an intricate process. In addition, the use of synthetic datasets has emerged to be a failure due to the decrease in performance when tested in real-time.

Image preprocessing is another way to enhance the attenuated camera images before they are fed into CNN models. In [1-3], it was shown that existing preprocessing strategies developed, trained, and tested using more or less synthetic datasets tend to be unreliable in real-world conditions, particularly for moving objects. In this study, the issue of object detection and classification in bad weather conditions was addressed by incorporating an image preprocessing unit into the object detectors, called Reinforced Image Object Detector (RIOD). The image preprocessing unit works by eliminating noise from the camera output and producing a clearer image, which is then fed to the object detector. As predicting the input domain or situation of the scene is almost impossible, it would be appropriate to remove the noise from the attenuated images and, therefore, improve the performance of the object detectors. The scope of this study can be summarized as follows:

- Designing an image reinforcement filter and using it before the object detectors. Based on certain variables, this filter can transform subpar images from the camera into images with proper illumination and better quality.
- Trainning the "You Only Look Once" (YOLO) v4 and v5 models with annotated datasets consisting of images under varying conditions i.e., fog, night, day, rain, haze, etc.
- Testing by filtering and reinforcing the data from the sensor. This reinforced image/video is then fed to the YOLOv4 and YOLOv5 models for object detection and classification.
- Perform several iterations, separately for images and video inputs, with YOLOv4 and YOLOv5 models to investigate the results of the proposed method in object detection and classification with the camera as the sole input sensor.

Fig. 1. Images representing the harsh driving scenarios under bad weather conditions. These images are samples collected from the DAWN dataset [4].

II. RELATED WORKS

Object detectors that rely on the deep CNN concept have produced excellent results in detecting objects with bounding boxes and labels. However, these models cannot be employed directly in certain applications, such as autonomous vehicles. In this case, image quality and visibility are highly dependent on environmental conditions. Object detectors can be broadly classified into two groups: single-stage and two-stage detectors. Single-stage detectors are faster in operation than two-stage detectors. Single-Shot multi-box Detector (SSD) [5] and YOLO [6] are two common single-stage object detectors.

Many researchers work on domain adaptation to improve object detection and classification in harsh climate conditions. In [7], the existing faster Region-based CNN (R-CNN) model was complemented with domain adaptive components (image and instance level shift), showing a slight performance improvement under all climate conditions. Although there is a slight performance improvement, R-CNN is slow in processing compared to the YOLO detectors [8]. In [9], domain-adaptive components were integrated into the YOLOv4 backbone at three different levels. This model exhibited a certain level of improvement in detecting objects compared to the original YOLOv4. In [10], a plug-and-play module, called cross-fusion, was added to the YOLOv5 model to combine the features of multiple convolutional layers, but its improved performance came at the cost of increased computational complexity. In [11-12], a multimodal sensor fusion technique was proposed, which presented an enhanced object detection rate for autonomous vehicles. However, the integration of four different sensors and the simultaneous processing of their outputs in real time proves to be cumbersome and expensive compared to other methods.

III. REINFORCED IMAGE OBJECT DETECTOR

YOLO is one of the fastest object-detecting models and has multiple versions. This study used YOLOv4 and YOLOv5 as baseline models, as they exhibit enhanced object detection performance when provided with good-quality images as input [13]. YOLOv4 [14], is a fast-operating object detection system that can be compared to a human body, with CSPdarknet53 being the backbone of the system that extracts multiple features. These features are collected by the layers called Spatial Pyramid Pooling (SPP) and Path Aggregation Network (PAN), which are the neck of the arrangement. This is further fed to the head of the system which is the original YOLOv3. After surveying these features, the head annotates the bounding boxes as well as the probabilities of the class of each box.

YOLOv5 is a faster and more lightweight object detection system that is more suitable for real-time applications. This system relies on a similar three-layer architecture comprising the backbone, neck, and head. The YOLOv5 architecture uses a focus structure with CSPDarknet53 as its backbone, which increases its speed and adds features like data augmentation and auto-learning bounding boxes. The weight files of YOLOv5 are significantly smaller than those of YOLOv4, making YOLOv5 faster and easier to handle. In the proposed RIOD-YOLO, the image/video is preprocessed with the help of image enhancement filter banks before feeding to the object detectors, to ensure that the performance of these detectors is kept stable irrespective of the driving scenario, namely fog, rain, night, day, etc.

A. Filter Design for Image-Enhancement

The output of the sensor, namely the camera output, is directed to the image preprocessing filter. This filter cleans up

the noisy image that is captured under undesirable weather conditions. These refined images are further tested for object detection using the YOLOv4 and YOLOv5 models. In this study, YOLOv4 was trained and tested with a custom dataset, collected specifically for autonomous vehicle applications, and YOLOv5 was tested with trained weights. During bad weather conditions, the images captured by the camera suffer a lot due to low visibility. These subpar images can be upgraded with the help of boundary conditions and regularizations [15]. The filter, designed as in [15], can perform scene transmission, regularization, and optimization to reduce noise and increase overall quality of the image. This filter can be used as a separate module and work as a ready-to-use image enhancer for any CNN model. The brightness of any pixel captured by the camera under any climatic condition can be derived as [15-17]:

$$R(x) = L(x)T(x) + A[1 - T(x)] \quad (1)$$

where $R(x)$ represents the captured pixel value, $L(x)$ is the luminosity of the scene, $T(x)$ is the scene transmission that describes the light that has reached the camera without getting scattered, and A is the atmospheric light. The objective is to recover $L(x)$ from (1), which requires an estimation of the transmission function $T(x)$ and the atmospheric light A. From (1), the scene can be recovered as:

$$L(x) = \frac{R(x)-A}{T(x)} + A \quad (2)$$

Thus, every x, $L(x)$ can be determined from (2) using the transmission function $T(x)$, quantified based on the boundary constraints specified in [15], and by estimating the atmospheric light A. Let us consider that $L(x)$ is bounded between two vectors $C0$ and $C1$ that have valid radiance values for every image. This in turn leaves a boundary constraint on $T(x)$. Using the concept of extrapolation, a boundary constraint $Lb(x)$ is worked out for every pixel x, and the lower bound $Tb(x)$ is determined by rearranging (2) as:

$$\frac{1}{T(x)} = \frac{\|L(x)-A\|}{\|R(x)-A\|} \quad (3)$$

The lower bound $Tb(x)$ ensures that the extrapolated values lie within the radiance vector values, therefore:

$$Tb(x) \leq T(x) \leq 1 \quad (4)$$

This implies that the extrapolated values for each pixel x must be within $[Tb(x), 1]$. $Tb(x)$ is further worked out as [15]:

$$T_b(x) = min\{max_{c\in\{r,g,b\}}(\frac{A_c - R_c(x)}{A_c - c_{c0}}, \frac{A_c - R_c(x)}{A_c - c_{c1}}), 1\} \quad (5)$$

This includes determining the smallest of the two terms, which are based on the color channels (RGB) and factors such as atmospheric light A, pixel intensity $R(x)$, and radiance bound C_{c0} and C_{c1}. Scene transmission $T(x)$ is substantially regularized to enhance the quality of dehazing. This approach employs weighted constraints between adjacent pixels:

$$W(x, y)[T(x) - T(y)] \quad (6)$$

where x and y are two adjacent pixels. The $w(x, y)$ controls the influence of constraints and ensures that similar neighboring pixels have similar transmission values. A bank of higher-order differential filters is used to calculate the weighting function in discrete form. The atmospheric light is estimated by applying a minimum filter to each color channel of the input image and then selecting the maximum value from each channel.

B. RIOD Architecture

Rather than complicating the design of Deep Neural Networks (DNN) [9-12, 18-19], a separate filter was developed to be placed between the camera and the CNN module. This filter mechanism enriches the image captured by the camera and feeds it to CNN models, as shown in Figure 2. This filter arrangement works efficiently for both image and video input. As mentioned above, this filtering mechanism reinforces the input image to the CNN model and, in turn, increases the object detection rate, particularly in adverse climate conditions.

IV. EXPERIMENTAL SETUP AND RESULTS

The YOLOv4 and YOLOv5 models were trained and tested with a custom dataset, consisting of 5232 images collected from a variety of datasets including Dawn, Udacity [20], BDD100k [21], and the Indian driving dataset [22]. The dataset includes images under various climate conditions, such as rain, fog, haze, night, and day. The models were tested using the custom dataset to detect 9 classes of objects, namely car, van, truck, traffic light, person, motorcycle, bus, and bicycle. The tests were carried out on standalone YOLO and RIOD-YOLO.

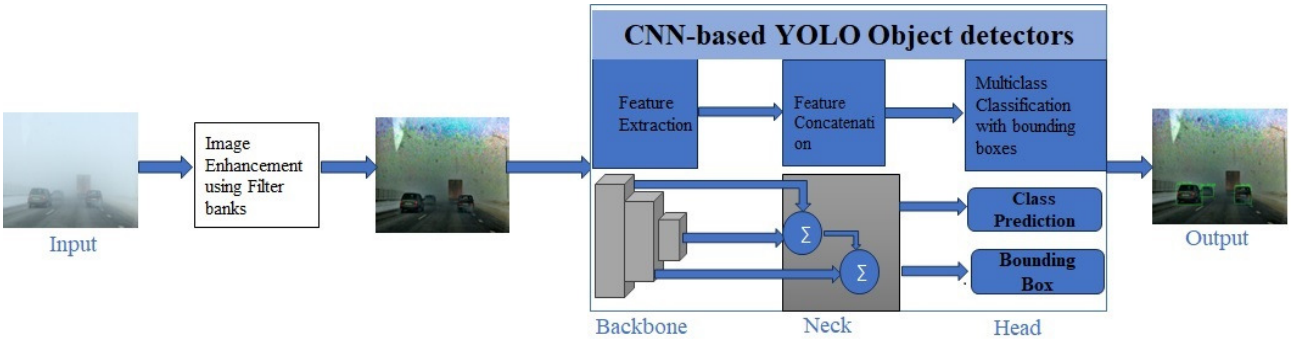

Fig. 2. RIOD architecture: A filter is placed between the camera and the CNN models, enhancing the image before feeding into the YOLO-based object detector. The output from the YOLO-based detector shows the multi-box classification of objects. The chosen images are from the DAWN dataset [4].

Fig. 3. Reinforced filter output for different climatic conditions such as fog, rain, and snow. All images are from the DAWN dataset [4].

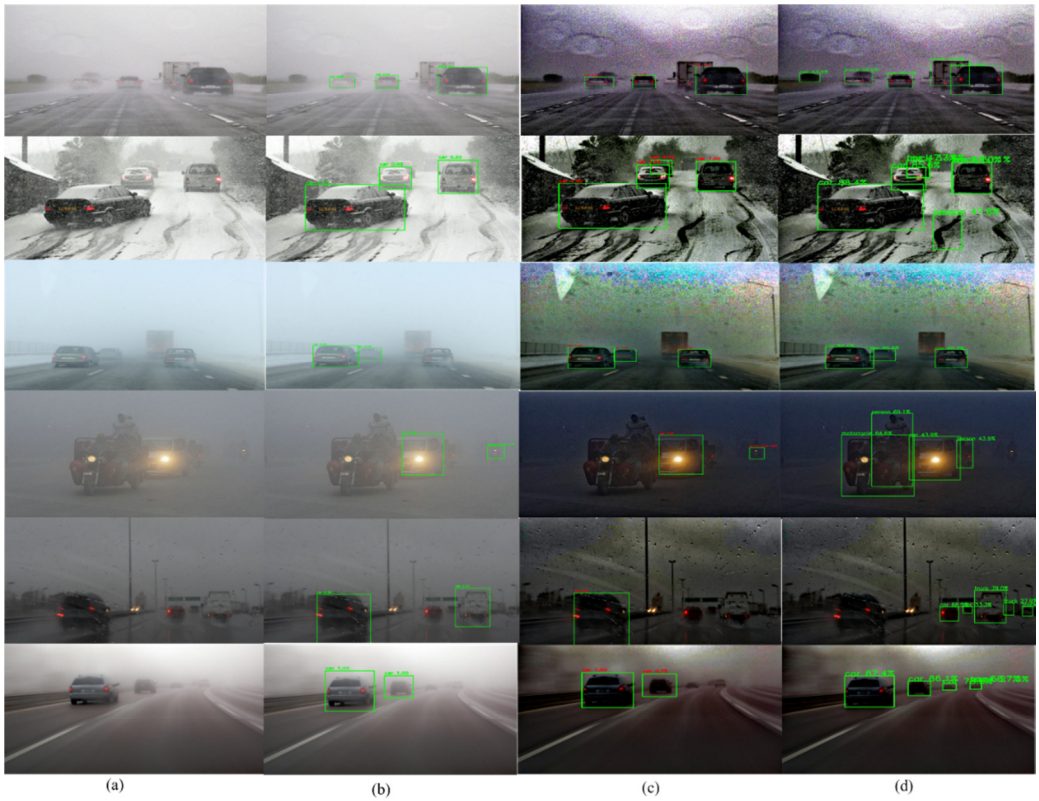

Fig. 4. Object detection under various climatic conditions: (a) input image, (b) object detection using customized YOLOv4, (c) RIOD using YOLOv4, and (d) RIOD using YOLOv5. All images are from the DAWN dataset [4].

A. Setup 1

The original YOLOv4 model with CSPdarknet53 as its backbone was initialized with a pre-trained weight file. The weights were updated by training the model using a custom dataset in batches of 64 images with a learning rate of 0.0001. This updated custom weight file was then used to test the model. Table I shows the test results. The reinforcement filter was mapped out based on the notion of extrapolation. First, the input images were passed through the reinforcement filter before being sent to the YOLO models for testing. Figure 3 compares the input and the corresponding output of the reinforcement filter for every image. These reinforced images are then fed into the customized YOLOv4 model for testing. Figure 4 shows the improvement in object detection rate after using the reinforcement filter in various climatic conditions.

B. Setup 2

The reinforcement filter functions as a plug-and-play module that can work with different versions of YOLO. In this iteration, the filter was attached to the YOLOv5 model, which was tested with both a pre-trained weight file (YOLOv5s) and a customized weight file. Figure 4 shows the improvement in object detection rate after attaching the reinforcement filter.

C. Evaluation Parameters

The performance was measured and quantitatively evaluated using average Precision (*P*) and Mean Average Precision (*MAP*). Their computation procedure was:

$$P = \frac{TP}{TP+FP} \quad (7)$$

$$R = \frac{TP}{TP+FN} \quad (8)$$

where *TP* represents true positive, *FP* also false positive, and *FN* also false negative. The Average Precision (*AP*) is given by:

$$AP = \int_0^1 P(R)\, dR \quad (9)$$

The Mean Average Precision (*MAP*) is given by:

$$MAP = \frac{1}{N}\sum_{x \in N} AP_x \quad (10)$$

where *N* is the number of object classes. The models were tested with the custom dataset, consisting of randomly selected images under a variety of climatic conditions, and Table I shows the MAP values.

TABLE I. COMPARISON OF DIFFERENT CNN METHODS

Method	Dataset	Classes						MAP
		C	B	V	TL	P	MC	
Domain adaptive faster R-CNN [7]	Cityscapes	41	35	22	-	25	20	29
YOLOv4[6]	Cityscapes	47	30	-	-	32	17	31
(MS-DAYOLOV4) [9]	Cityscapes	56	36	-	-	39	29	40
Robost YOLOv3	Cityscapes	58	51	-	-	36	26	43
Robost YOLOvX	Cityscapes	61	57	-	-	45	42	51
YOLOv4	BDD100K	73	-	-	47	42	-	54
YOLOv4	INIT	74	-	-	48	45	-	56
RIOD-YOLOv4	**custom dataset**	**65**	**41**	**22**	**16**	**26**	**70**	**40**
RIOD-YOLOv5	**custom dataset**	**78**	**77**	**60**	**43**	**60**	**75**	**65**

column classes C: car, B: bus, V: van, TL: traffic light, P: person, MC: motorcycle

V. CONCLUSION

The proposed RIOD-YOLO aimed to improve object detection efficiency, particularly under adverse climatic conditions. In this method, the output from the sensor is fed to the image enhancement filter, which comprehensively improves the image quality before being fed to the object detectors, improving the rate of object detection. A customized dataset was used to evaluate the effectiveness of the proposed method, and the results clearly showed an improvement in the rate of object detection, especially under unfavorable climatic conditions. When compared to the baseline YOLO models, the proposed RIOD-YOLO produced promising results in every aspect. The scope of RIOD-YOLO is not limited to just autonomous driving, as it can also be used in surveillance systems that use image/video input to improve visibility, especially in adverse weather conditions.

REFERENCES

[1] P. P. Pavitha, K. B. Rekha, and S. Safinaz, "Perception system in Autonomous Vehicle: A study on contemporary and forthcoming technologies for object detection in autonomous vehicles," in *2021 International Conference on Forensics, Analytics, Big Data, Security (FABS)*, Bengaluru, India, Sep. 2021, vol. 1, pp. 1–6, https://doi.org/10.1109/FABS52071.2021.9702569.

[2] M. Hnewa and H. Radha, "Object Detection Under Rainy Conditions for Autonomous Vehicles: A Review of State-of-the-Art and Emerging Techniques," *IEEE Signal Processing Magazine*, vol. 38, no. 1, pp. 53–67, Jan. 2021, https://doi.org/10.1109/MSP.2020.2984801.

[3] W. Ritter, M. Bijelic, T. Gruber, M. Kutila, and H. Holzhüter, "DENSE: Environment Perception in Bad Weather—First Results," in *Electronic Components and Systems for Automotive Applications*, 2019, pp. 143–159, https://doi.org/10.1007/978-3-030-14156-1_12.

[4] M. A. Kenk and M. Hassaballah, "DAWN: Vehicle Detection in Adverse Weather Nature Dataset." Mar. 06, 2020, https://doi.org/10.17632/766ygrbt8y.3.

[5] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," in *Computer Vision – ECCV 2016*, Amsterdam, Netherlands, 2016, pp. 21–37, https://doi.org/10.1007/978-3-319-46448-0_2.

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 779–788, https://doi.org/10.1109/CVPR.2016.91.

[7] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, "Domain Adaptive Faster R-CNN for Object Detection in the Wild." arXiv, Mar. 08, 2018, https://doi.org/10.48550/arXiv.1803.03243.

[8] J. Fayyad, M. A. Jaradat, D. Gruyer, and H. Najjaran, "Deep Learning Sensor Fusion for Autonomous Vehicle Perception and Localization: A Review," *Sensors*, vol. 20, no. 15, Jan. 2020, Art. no. 4220, https://doi.org/10.3390/s20154220.

[9] M. Hnewa and H. Radha, "Multiscale Domain Adaptive Yolo For Cross-Domain Object Detection," in *2021 IEEE International Conference on Image Processing (ICIP)*, Anchorage, AK, USA, Sep. 2021, pp. 3323–3327, https://doi.org/10.1109/ICIP42928.2021.9506039.

[10] Q. Ding *et al.*, "CF-YOLO: Cross Fusion YOLO for Object Detection in Adverse Weather with a High-quality Real Snow Dataset." arXiv, Jun. 03, 2022, https://doi.org/10.48550/arXiv.2206.01381.

[11] M. Bijelic *et al.*, "Seeing Through Fog Without Seeing Fog: Deep Multimodal Sensor Fusion in Unseen Adverse Weather," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 11679–11689, https://doi.org/10.1109/CVPR42600.2020.01170.

[12] H. G. Doan and N. T. Nguyen, "Fusion Machine Learning Strategies for Multi-modal Sensor-based Hand Gesture Recognition," *Engineering, Technology & Applied Science Research*, vol. 12, no. 3, pp. 8628–8633, Jun. 2022, https://doi.org/10.48084/etasr.4913.

[13] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680–1716, Dec. 2023, https://doi.org/10.3390/make5040083.

[14] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection." arXiv, Apr. 22, 2020, https://doi.org/10.48550/arXiv.2004.10934.

[15] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient Image Dehazing with Boundary Constraint and Contextual Regularization," in *2013 IEEE International Conference on Computer Vision*, Sydney, Australia, Sep. 2013, pp. 617–624, https://doi.org/10.1109/ICCV.2013.82.

[16] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 713–724, Jun. 2003, https://doi.org/10.1109/TPAMI.2003.1201821.

[17] K. He, J. Sun, and X. Tang, "Single Image Haze Removal Using Dark Channel Prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, Sep. 2011, https://doi.org/10.1109/TPAMI.2010.168.

[18] W. Ali, G. Wang, K. Ullah, M. Salman, and S. Ali, "Substation Danger Sign Detection and Recognition using Convolutional Neural Networks," *Engineering, Technology & Applied Science Research*, vol. 13, no. 1, pp. 10051–10059, Feb. 2023, https://doi.org/10.48084/etasr.5476.

[19] V. Saikrishnan and M. Karthikeyan, "Mayfly Optimization with Deep Learning-based Robust Object Detection and Classification on Surveillance Videos," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11747–11752, Oct. 2023, https://doi.org/10.48084/etasr.6231.

[20] Y. Hou, Z. Ma, C. Liu, and C. C. Loy, "Learning to Steer by Mimicking Features from Heterogeneous Auxiliary Networks." arXiv, Nov. 06, 2018, https://doi.org/10.48550/arXiv.1811.02759.

[21] F. Yu *et al.*, "BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning." arXiv, Apr. 08, 2020, https://doi.org/10.48550/arXiv.1805.04687.

[22] G. Varma, A. Subramanian, A. Namboodiri, M. Chandraker, and C. V. Jawahar, "IDD: A Dataset for Exploring Problems of Autonomous Navigation in Unconstrained Environments." arXiv, Nov. 26, 2018, https://doi.org/10.48550/arXiv.1811.10200.