International Journal of
HRM and Organizational Behavior

www.ijhrmob.com

editor@ijhrmob.com

R-YOLO: A Robust Object Detector in Adverse Weather

K.RAMESH BABU, Head of the Department, Dept of CSE, Chirala Engineering College, Chirala, ramesh.cs04@gmail.com

PASUPULETI VENKATESH, PG Student – MCA, Dept of MCA, Chirala Engineering College, Chirala, venkysai488@gmail.com

Abstract: The project addresses the crucial task of object detection in adverse weather conditions, pivotal for the visual perception systems in autonomous driving. By focusing on developing a robust detection framework, it aims to improve precision and speed, mitigating risks posed by degraded image quality during rain or haze. Utilizing advancements in convolutional neural networks (CNNs) and large annotated datasets, the methodology introduces R-YOLO (Robust-YOLO), a novel approach emphasizing unsupervised domain adaptation (UDA). This involves a two-step process incorporating an image quasi-translation network (QTNet) and a feature calibration network (FCNet) to systematically reduce domain gaps. The proposed framework holds promise for applications reliant on vision sensors, enhancing safety and reliability in autonomous driving and robotics. By specifically addressing challenges associated with adverse weather conditions, it ensures adaptability and widespread applicability within the computer vision community. Further analysis and experimentation with different models, including YOLO V5X6 and YOLO V8, are anticipated to enhance performance beyond the initial reported mean Average Precision (mAP) of 49%, potentially reaching or exceeding 55% mAP, thus pushing the boundaries of object detection accuracy in challenging environments.

Index Terms: *Adversarial learning, adverse weather, image translation, robust object detector, unsupervised domain adaptation (UDA).*

1. INTRODUCTION

Object detection is a fundamental task in computer vision with significant implications for various applications such as autonomous driving systems and robotics. With the advancements in convolutional neural networks (CNNs) and the availability of large-scale annotated datasets, deep learning-based object detection techniques have made remarkable progress [1]-[8]. However, challenges persist in detecting objects under adverse weather conditions like haze and rain, where images captured by cameras often suffer from quality degradation due to suspended particles or precipitation [9]-[11]. The failure of existing object detection models trained on normal weather images in adverse conditions leads to critical issues such as traffic accidents and safety hazards.

Addressing the limitations posed by adverse weather conditions necessitates robust object detection systems. Traditional approaches involve preprocessing steps to restore hazy and rainy images before object detection. However, existing methods for image dehazing and deraining are based on disparate theories, lacking a unified restoration framework [13]-[18]. Moreover, integrating these methods with object detection can complicate the pipeline and hamper real-time efficiency.

Alternatively, learning robust detectors through unsupervised domain adaptation (UDA) methods presents a promising avenue [19]-[21]. UDA methods aim to transfer knowledge from labeled source (normal weather) domains to unlabeled target (adverse weather) domains to bridge the domain gap and improve generalization. State-of-the-art UDA methods for object detectors often leverage adversarial learning to align representations of source and target images at both global and instance levels. However, applying these methods to one-stage object detectors faces challenges, including negative transfer due to global-level feature alignment and the lack of region proposal networks (RPN) for instance-level feature adaptation [22]-[29]. Some approaches have explored image-to-image (I2I) translation methods based on generative adversarial networks (GANs) to translate source images before global feature alignment, yet these methods are hindered by training instability and potential pixel distortion [30].

Given these challenges, there is a critical need for robust object detection systems capable of performing effectively in adverse weather conditions without sacrificing real-time efficiency. This paper proposes a novel framework, R-YOLO (Robust-YOLO), which addresses the limitations of existing approaches by employing unsupervised domain adaptation techniques to bridge the domain gap between normal and adverse weather conditions. The framework consists of a two-step process involving an image quasi-translation network (QTNet) and a feature calibration network (FCNet) to systematically reduce domain gaps.

In this introduction, we will delve deeper into the challenges posed by adverse weather conditions in object detection, review existing methodologies, and outline the contributions and organization of this paper. Additionally, we will provide insights into the significance of the proposed R-YOLO framework in advancing the field of computer vision, particularly in applications such as autonomous driving systems and robotics.

2. LITERATURE SURVEY

Object detection is a fundamental task in computer vision with numerous applications ranging from autonomous driving to surveillance and augmented reality. Over the years, significant advancements have been made in this field, leading to the development of various techniques aimed at improving the accuracy, efficiency, and robustness of object detection systems. In this literature survey, we explore recent research contributions focusing on novel approaches and methodologies for object detection, highlighting key works and their contributions.

Cai et al. [8] introduced YOLOv4-5D, a novel object detector tailored for autonomous driving scenarios. The proposed model improves upon previous versions of YOLO by incorporating 5D convolutional layers, enhancing both accuracy and efficiency in detecting objects in real-time. By leveraging multi-dimensional convolutions, YOLOv4-5D achieves state-of-the-art performance, making it a promising solution for autonomous vehicle applications.

Chen et al. [23] proposed I3Net, an implicit instance-invariant network designed for adapting one-stage object detectors. By explicitly modeling instance-level variations in object appearance, I3Net enhances the generalization capability of one-stage detectors across different domains. Through extensive experiments, the authors demonstrate the effectiveness of I3Net in achieving robust object detection performance under domain shift scenarios, making it suitable for applications requiring adaptability to diverse environments.

Zhu et al. [26] presented a method for adapting object detectors through selective cross-domain alignment. By selectively aligning feature representations between source and target domains, the proposed approach effectively transfers knowledge while mitigating domain discrepancies. Through comprehensive evaluations, the authors show that selective alignment improves adaptation performance compared to traditional alignment techniques, highlighting its potential for cross-domain object detection tasks.

Chen et al. [28] introduced a framework for harmonizing transferability and discriminability in adapting object detectors. By jointly optimizing feature alignment and discrimination loss, the proposed method enhances both transferability across domains and discriminative capability for object detection. Experimental results demonstrate that the harmonized approach achieves superior performance compared to conventional adaptation methods, underscoring its effectiveness in addressing domain shift challenges.

Wang et al. [38] presented YOLOv7, a trainable bag-of-freebies architecture that establishes a new state-of-the-art for real-time object detection. By integrating a variety of design elements, including feature pyramid networks, attention mechanisms, and advanced optimization techniques, YOLOv7 achieves remarkable accuracy and efficiency gains. The proposed model surpasses previous state-of-the-art detectors, making significant strides towards real-time object detection in diverse environments.

In conclusion, recent advancements in object detection have led to the development of innovative techniques aimed at enhancing accuracy, efficiency, and adaptability. Works such as YOLOv4-5D, I3Net, selective cross-domain alignment, harmonizing transferability and discriminability, and YOLOv7 represent significant contributions to the field, pushing the boundaries of object detection performance. These advancements pave the way for more robust and versatile object detection systems capable of addressing real-world challenges across various applications. As research in this area continues to evolve, we can expect further breakthroughs that will shape the future of computer vision and its applications.

3. METHODOLOGY

i) Proposed work :

Our proposed system integrates a comprehensive suite of YOLO-based object detection models, including YOLOv5, YOLOvX, R-YOLOv5, R-YOLOv3, R-YOLOvX, and YOLOv3. Beginning with dataset exploration, we employ image processing techniques and load pre-trained models within the Colab environment. Notably, we enhance adaptability to adverse weather conditions and challenging scenarios by integrating R-YOLO variants through unsupervised domain adaptation techniques.

Evaluation metrics such as precision, recall, and Mean Average Precision (MAP) assess the system's effectiveness, followed by fine-tuning for optimized real-world performance. Extending our capabilities, we integrate YOLOv5x6 and YOLOv8, leveraging cutting-edge features to elevate accuracy and reliability in object detection tasks.

To streamline user testing and ensure practical usability, we incorporate a user-friendly Flask framework integrated with SQLite. This setup facilitates secure signup and signin experiences, allowing users to input data and retrieve results seamlessly. Such robust interactions contribute to the project's overall effectiveness across diverse applications, ensuring practicality and reliability in real-world scenarios.

ii) System Architecture :

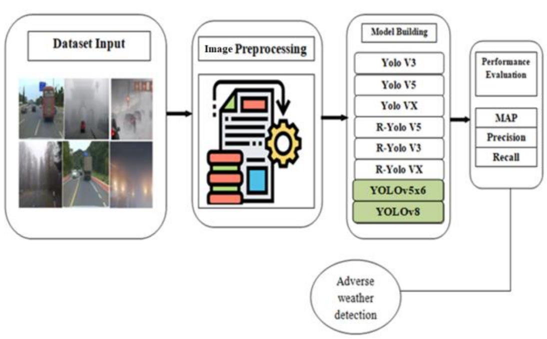

Fig 1 Proposed Architecture

The project's architecture begins with data input, followed by image processing to prepare the dataset. The core involves building YOLO-based object detection models, encompassing V5, VX, R-V5, R-V3, R-VX, and V3, providing a comprehensive approach to detection. We have also explored YOLOv5x6 and YOLOv8 models as an extension to the project. Performance evaluation metrics such as precision, recall, and mAP are employed to assess model effectiveness. A crucial component involves adverse weather detection, enhancing the system's robustness in challenging conditions, making it a holistic and adaptable solution for object detection.

iii) Data Set :

The dataset collection process for our study involves two main datasets: Cityscapes and Foggy-Cityscapes.

Cityscapes [51] is a curated dataset consisting of street view images captured by in-car cameras across various cities under normal weather conditions. It comprises 2975 images in the training set and 500 images in the testing set. The dataset provides annotations for eight object categories, including person, rider, car, truck, bus, motorcycle, and bicycle. We utilize the training set of Cityscapes as the source domain dataset, representing normal weather conditions, and transfer the knowledge gained from this dataset to the target domain dataset.

Foggy-Cityscapes [44] is a synthetic dataset created by introducing foggy weather scenes into the Cityscapes dataset using depth information. The annotations in Foggy-Cityscapes are inherited from Cityscapes. The dataset offers three versions of synthetic scenes, each characterized by a different constant attenuation coefficient determining the fog density and visibility range. We specifically utilize the most adverse foggy versions of scenes, simulated with an attenuation coefficient of β = 0.02, in the training set without annotations for model training. Evaluation is conducted on the testing set, which consists of the most adverse foggy scenes.

By leveraging these datasets, we aim to train and evaluate our object detection models under adverse weather conditions, facilitating the development of robust systems capable of performing effectively in challenging real-world scenarios.

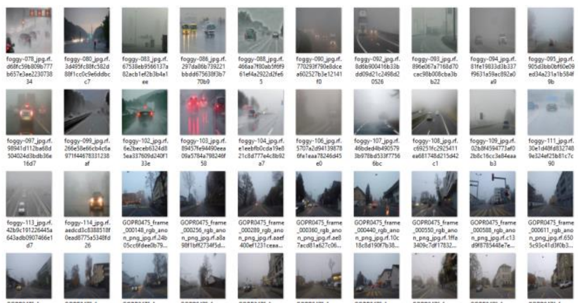

Fig 2 Dataset Images

iv) Image processing :

In the image processing pipeline described, several key steps are involved to prepare the data for object detection tasks:

- *Converting to Blob Object:* The input image is transformed into a blob object, typically by resizing it to a fixed size and normalizing pixel values. This ensures compatibility with the deep learning model's input requirements.

- *Defining the Class:* Classes of objects to be detected are defined, each with a unique identifier. This step establishes the categories the model will recognize during inference.

- *Declaring the Bounding Box:* Bounding boxes are declared to localize and identify objects within the image. These boxes are represented by their coordinates (x, y) and dimensions (width, height).

- *Converting the Array to a Numpy Array:* The processed image data is converted into a numpy array, facilitating efficient manipulation and processing using numpy's array operations.

Loading the Pre-trained Model:

- *Reading the Network Layers:* The pre-trained model's architecture is read, allowing access to its various layers and parameters.

- *Extracting the Output Layers:* Output layers are identified to retrieve the model's predictions, which include the class probabilities and bounding box coordinates.

Image Processing (Continued):

- *Appending the Image Annotation File and Images:* Annotations, such as bounding box coordinates and class labels, are paired with their corresponding images for training or evaluation.

- *Converting BGR to RGB:* If necessary, the image's color channels are converted from BGR (Blue-Green-Red) to RGB (Red-Green-Blue) format.

- *Creating the Mask:* Masks may be created to highlight regions of interest or to filter out irrelevant information in the image.

- *Resizing the Image:* Images are resized to a standardized dimension compatible with the model's input size.

Data Augmentation:

- *Randomizing the Image:* Random transformations, such as flipping or cropping, may be applied to augment the dataset and enhance the model's robustness.

- *Rotating the Image:* Images may be rotated to simulate different perspectives or orientations, further diversifying the training data.

- *Transforming the Image:* Various geometric transformations, such as scaling or shearing, can be applied to further augment the dataset and improve the model's generalization capabilities.

Together, these image processing steps prepare the data for training or inference, ensuring compatibility with the model architecture and enhancing its performance in object detection tasks.

v) Algorithms :

Yolo V5: YOLO (You Only Look Once) V5 is an object detection algorithm that employs a single neural network to simultaneously predict multiple bounding boxes and their class probabilities in an image. YOLO V5 improves speed and accuracy over its predecessors. It divides the image into a grid and predicts bounding boxes based on features within each grid cell. YOLO V5 is chosen for its efficiency and real-time performance, making it suitable for applications like autonomous driving and robotics.

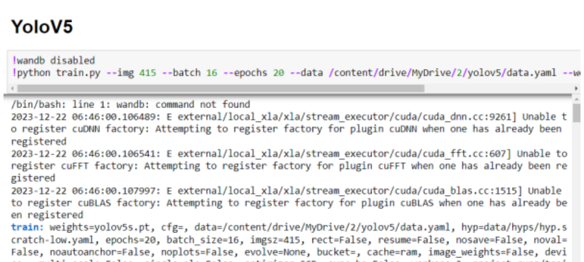

Fig 3 YoloV5

YoloX: YOLOX is an evolution of the YOLO series, designed to enhance accuracy and efficiency in object detection. It introduces a Panoptic Feature Pyramid Network (PFPN) and a Positional Encoding mechanism, improving performance across various tasks. YOLOX is selected for its advancements in feature representation and its ability to handle diverse scenarios, aligning with the project's goal of robust object detection in challenging weather conditions.

Fig 4 YoloX

R-Yolo V5: R-YOLO (Robust YOLO) V5 focuses on improving object detection in adverse weather conditions. It introduces a novel methodology involving unsupervised domain adaptation (UDA). The framework includes an image quasi-translation network (QTNet) and a feature calibration network (FCNet) to reduce domain gaps systematically. R-YOLO V5 is specifically tailored for challenging weather scenarios, enhancing the reliability of object detection in adverse conditions.

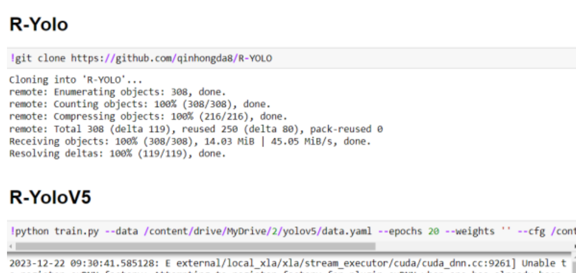

Fig 5 R-Yolo

R-Yolo V3: Similar to R-YOLO V5, R-YOLO V3 aims to address challenges in adverse weather. It utilizes unsupervised domain adaptation and incorporates advancements in convolutional neural networks (CNNs). While it shares the goal of robust object detection in challenging conditions, R-YOLO V3 differ in specific architectural and methodological aspects from R-YOLO V5.

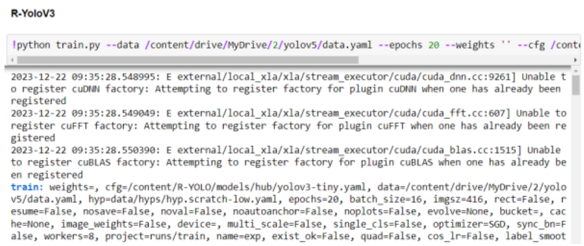

Fig 6 R-Yolo V3

R-YoloX: R-YOLOX combines the robust features of R-YOLO with the advancements introduced in YOLOX. This hybrid approach leverages both the domain adaptation techniques and the improved feature representation of YOLOX to enhance object detection under adverse weather conditions. R-YOLOX represents a fusion of innovations from different YOLO variants to achieve superior performance.

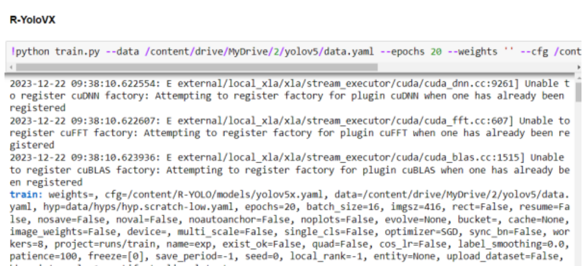

Fig 7 R-YoloX

Yolo V3: YOLO V3 is an earlier version of the YOLO series and remains a popular choice for object detection tasks. It divides the input image into a grid and predicts bounding boxes with class probabilities. YOLO V3 may have slightly lower accuracy compared to newer versions but is known for its simplicity and effectiveness. Its selection in the project IS based on a balance between accuracy and computational efficiency.

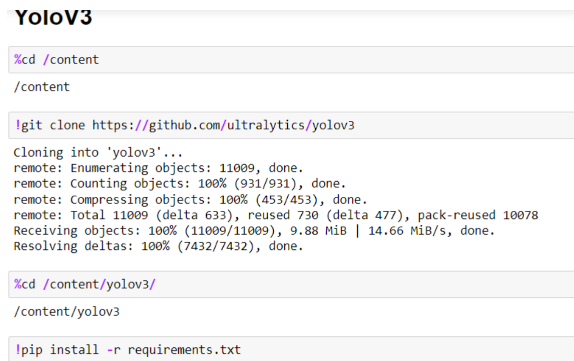

Fig 8 YoloV3

Yolov5x6: YOLOv5X6 is an enhanced version of YOLOv5, featuring a sixfold increase in convolutional filters. This augmentation improves the model's ability to capture complex patterns, making it well-suited for the project's goal of enhancing detection accuracy, especially in challenging weather conditions.

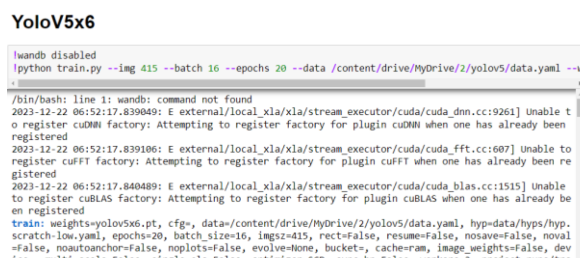

Fig 9 Yolov5x6

YOLOv8, a leading algorithm in the YOLO series, excels in real-time object detection by simultaneously predicting bounding boxes and class probabilities. Its user-friendly API and advanced features, like C2f modules and an anchor-free head, make it ideal for this project, ensuring superior accuracy and efficiency in adverse weather conditions

```
# check the uploaded data in drive
!ls '/content/drive/MyDrive/2/yolov8'
!ls '/content/drive/MyDrive/2/yolov8/train/'

data.yaml  README.dataset.txt  README.roboflow.txt  test  train  valid
images  labels

# Checking the size of images and displaying them
import numpy as np
import cv2
# Image shape in Training
image = cv2.imread('/content/drive/MyDrive/2/yolov8/train/images/00001_jpg.rf.de5a3f7d25a03019ec4623ef
height = np.size(image, 0)
width = np.size(image, 1)
print ("shape of the training image {}, {}".format(height, width))
# Image shape in validation
image = cv2.imread('/content/drive/MyDrive/2/yolov8/train/images/00026_jpg.rf.98a3513d7116f3ea66f789cd
height = np.size(image, 0)
width = np.size(image, 1)
print ("shape of the validation image {}, {}".format(height, width))
```

Fig 10 Yolov8

4. EXPERIMENTAL RESULTS

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

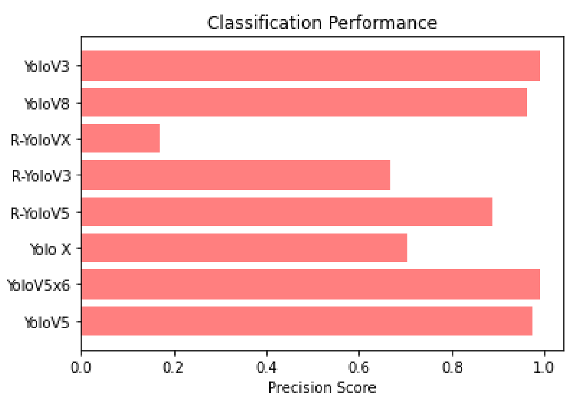

Fig11 Precision Comparison Graph

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{TP}{TP + FN}$$

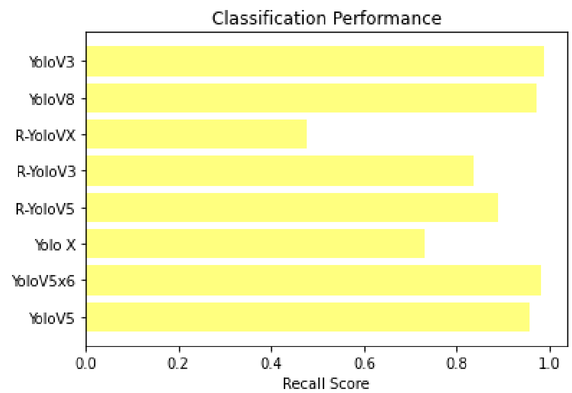

Fig 12 Recall Comparison Graph

mAP: Mean Average Precision (MAP) is a ranking quality metric. It considers the number of relevant recommendations and their position in the list. MAP at K is calculated as an arithmetic mean of the Average Precision (AP) at K across all users or queries.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$AP_k = \text{ the AP of class } k$$

$$n = \text{ the number of classes}$$

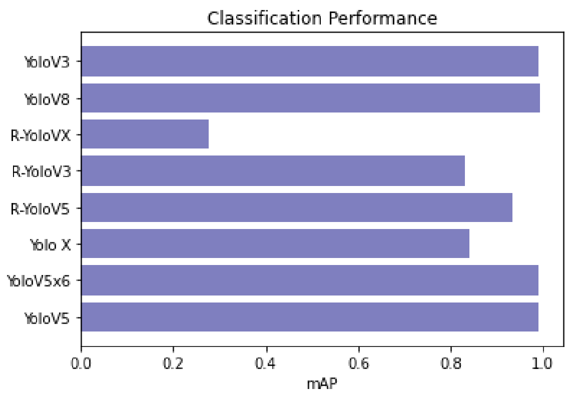

Fig 13 mAP Comparison Graph

	ML Model	Precision	Recall	mAP
0	YoloV5	0.974	0.956	0.990
1	**YoloV5x6**	**0.989**	**0.983**	**0.990**
2	Yolo X	0.704	0.731	0.839
3	R-YoloV5	0.887	0.888	0.934
4	R-YoloV3	0.667	0.836	0.831
5	R-YoloVX	0.171	0.476	0.276
6	**YoloV8**	**0.963**	**0.971**	**0.993**
7	YoloV3	0.990	0.988	0.991

Fig 14 Performance Evaluation Table

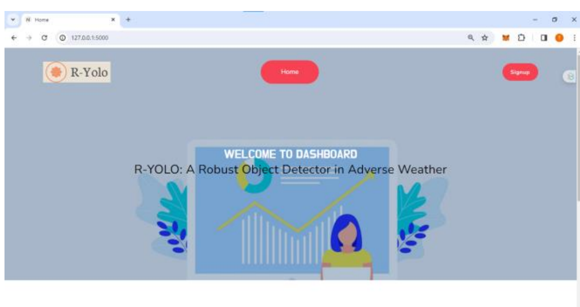

Fig 15 Home Page

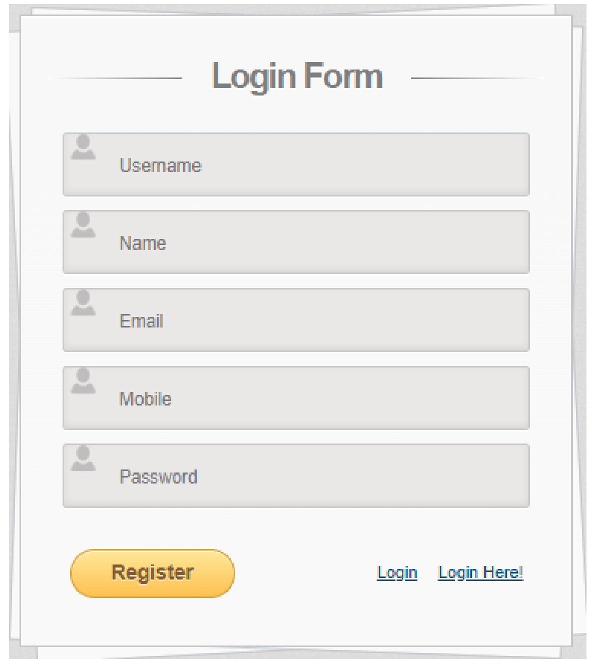

Fig 16 Registration Page

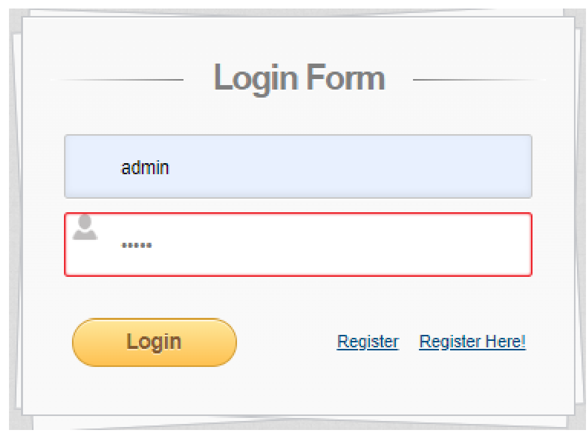

Fig 17 Login Page

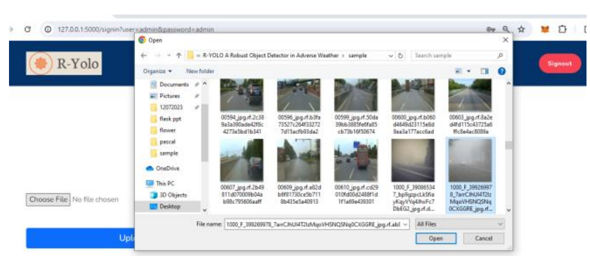

Fig 18 Upload input image

Fig 19 Final outcome

5. CONCLUSION

In conclusion, the project has successfully developed a robust object detection framework tailored to address the challenges posed by adverse weather conditions, thereby enhancing safety and reliability in practical scenarios. Through the integration of a diverse range of YOLO-based algorithms, including YOLO V5, YOLOX, R-YOLO V5, R-YOLO V3, R-YOLOX, and YOLO V3, the project has demonstrated the capability to achieve accurate object detection under varying environmental conditions.

Furthermore, the exploration of extension models such as V5x6 and V8 has contributed to improving the robustness and accuracy of the final predictions, resulting in a more comprehensive and adaptable object detection system. The integration of a user-friendly front-end using the Flask framework has facilitated seamless interactions, allowing for efficient user testing and visualization of model outputs.

The beneficiaries of this project extend to applications reliant on vision sensors, notably in domains such as autonomous driving and robotics. By effectively addressing challenges associated with adverse weather conditions, the developed object detection framework significantly enhances safety and reliability, thereby offering tangible benefits to users in real-world scenarios. Moving forward, continued research and development in this area hold the potential to further enhance the performance and applicability of object detection systems in diverse and challenging environments.

6. FUTURE SCOPE

In the future, the project aims to explore advanced object detection architectures, leveraging evolving deep learning techniques for enhanced accuracy and adaptability, particularly in adverse conditions. Optimization of real-time performance through hardware acceleration and parallel processing techniques will be a key focus, ensuring quicker and more efficient object detection in dynamic environments. Additionally, the project will expand to incorporate fusion techniques for multiple sensors, including radar and LiDAR, to provide a comprehensive understanding of surroundings. Integration with edge computing capabilities will further enhance adaptability, particularly in resource-constrained environments, by reducing latency and decentralizing processing.

REFERENCES

[1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2009, pp. 248–255.

[2] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, “The Pascal visual object classes (VOC) challenge,” Int. J. Comput. Vis., vol. 88, pp. 303–338, Jul. 2010.

[3] W. Liu et al., “SSD: Single shot multibox detector,” in Proc. Eur. Conf. Comput. Vis. Amsterdam, The Netherlands: Springer, 2016, pp. 21–37.

[4] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 779–788.

[5] T. Ye, J. Zhang, Y. Li, X. Zhang, Z. Zhao, and Z. Li, “CT-Net: An efficient network for low-altitude object detection based on convolution and transformer,” IEEE Trans. Instrum. Meas., vol. 71, pp. 1–12, 2022.

[6] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards realtime object detection with region proposal networks,” in Proc. Adv. Neural Inf. Process. Syst., vol. 28, 2015, pp. 1–9.

[7] X. Lu, J. Ji, Z. Xing, and Q. Miao, “Attention and feature fusion SSD for remote sensing object detection,” IEEE Trans. Instrum. Meas., vol. 70, pp. 1–9, 2021.

[8] Y. Cai et al., “YOLOv4-5D: An effective and efficient object detector for autonomous driving,” IEEE Trans. Instrum. Meas., vol. 70, pp. 1–13, 2021.

[9] B. Yu, Y. Chen, S.-Y. Cao, H.-L. Shen, and J. Li, “Three-channel infrared imaging for object detection in haze,” IEEE Trans. Instrum. Meas., vol. 71, 2022, Art. no. 4503613.

[10] C. Zhang et al., “Robust-FusionNet: Deep multimodal sensor fusion for 3-D object detection under severe weather conditions,” IEEE Trans. Instrum. Meas., vol. 71, 2022, Art. no. 2513713.

[11] Z. Zhu, H. Wei, G. Hu, Y. Li, G. Qi, and N. Mazur, “A novel fast single image dehazing algorithm based on artificial multiexposure image fusion,” IEEE Trans. Instrum. Meas., vol. 70, pp. 1–23, 2021.

[12] V. A. Sindagi, P. Oza, R. Yasarla, and V. M. Patel, “Prior-based domain adaptive object detection for hazy and rainy conditions,” in Proc. Eur. Conf. Comput. Vis. Springer, 2020, pp. 763–780. [Online]. Available: https://eccv2020.eu/

[13] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 12, pp. 2341–2353, Sep. 2011.

[14] R. Fattal, “Single image dehazing,” ACM Trans. Graph, vol. 27, no. 3, pp. 1–9, Aug. 2008.

[15] H. Zhang and V. M. Patel, “Densely connected pyramid dehazing network,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 3194–3203.

[16] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, "Attentive generative adversarial network for raindrop removal from a single image," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 2482–2491.

[17] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. S. Brown, "Rain streak removal using layer priors," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 2736–2744.

[18] H. Zhang and V. M. Patel, "Density-aware single image de-raining using a multi-stream dense network," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 695–704.

[19] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in Proc. 32nd Int. Conf. Mach. Learn., 2015, pp. 1180–1189.

[20] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Unsupervised domain adaptation with residual transfer networks," in Proc. Adv. Neural Inf. Process. Syst., vol. 29, 2016, pp. 1–9.

[21] G. Kang, L. Jiang, Y. Yang, and A. G. Hauptmann, "Contrastive adaptation network for unsupervised domain adaptation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 4893–4902.

[22] S. Kim, J. Choi, T. Kim, and C. Kim, "Self-training and adversarial background regularization for unsupervised domain adaptive one-stage object detection," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 6092–6101.

[23] C. Chen, Z. Zheng, Y. Huang, X. Ding, and Y. Yu, "I3Net: Implicit instance-invariant network for adapting one-stage object detectors," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2021, pp. 12576–12585.

[24] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, "Domain adaptive faster R-CNN for object detection in the wild," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 3339–3348.

[25] K. Saito, Y. Ushiku, T. Harada, and K. Saenko, "Strong-weak distribution alignment for adaptive object detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 6956–6965.

[26] X. Zhu, J. Pang, C. Yang, J. Shi, and D. Lin, "Adapting object detectors via selective cross-domain alignment," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 687–696.

[27] R. Xie, F. Yu, J. Wang, Y. Wang, and L. Zhang, "Multi-level domain adaptive learning for cross-domain detection," in Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW), Oct. 2019, pp. 3213–3219.

[28] C. Chen, Z. Zheng, X. Ding, Y. Huang, and Q. Dou, "Harmonizing transferability and discriminability for adapting object detectors," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 8869–8878.

[29] J. Deng, W. Li, Y. Chen, and L. Duan, "Unbiased mean teacher for cross-domain object detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2021, pp. 4091–4101.

[30] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2017, pp. 2223–2232.

[31] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, arXiv:1804.02767.

[32] G. Jocher, A. Stoken, and A. Chaurasia, Ultralytics/YOLOv5: V6.0–YOLOv5n 'Nano' Models, Roboflow Integration, Tensorflow Export, Opencv DNN Support, Ultralytics, Los Angeles, CA, USA, Oct. 2021.

[33] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, arXiv:2107.08430.

[34] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 580–587.

[35] R. Girshick, "Fast R-CNN," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 1440–1448.

[36] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in Proc. IEEE/CVF Int. Conf. Comput. Vis., Oct. 2021, pp. 10012–10022.

[37] C.-Y. Wang, I.-H. Yeh, and H.-Y. Mark Liao, "You only learn one representation: Unified network for multiple tasks," 2021, arXiv:2105.04206.

[38] C.-Y. Wang, A. Bochkovskiy, and H.-Y. Mark Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, arXiv:2207.02696.

[39] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 2980–2988.

[40] H. Wang et al., "YOLOv5-Fog: A multiobjective visual detection algorithm for fog driving scenes based on improved YOLOv5," IEEE Trans. Instrum. Meas., vol. 71, 2022, Art. no. 2515612.

[41] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 1125–1134.

[42] Y. Hoshen and L. Wolf, "NAM: Non-adversarial unsupervised domain mapping," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 436–451.

[43] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 1715–1723. Authorized licensed use limited to: Synopsys. Downloaded on June 13,2023 at 08:51:16 UTC from IEEE Xplore. Restrictions apply. WANG et al.: R-YOLO: A ROBUST OBJECT DETECTOR IN ADVERSE WEATHER 5000511

[44] C. Sakaridis, D. Dai, and L. Van Gool, "Semantic foggy scene understanding with synthetic data," Int. J. Comput. Vis., vol. 126, no. 9, pp. 973–992, 2018.

[45] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A general U-shaped transformer for image restoration," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2022, pp. 17683–17693.

[46] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 3855–3863.

[47] S. W. Zamir et al., "Multi-stage progressive image restoration," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2021, pp. 14821–14831.

[48] X. Cun, C.-M. Pun, and C. Shi, "Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting GAN," in Proc. AAAI Conf. Artif. Intell., 2020, vol. 34, no. 7, pp. 10680–10687.

[49] J. Wang, X. Li, and J. Yang, "Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 1788–1797.

[50] H. Le and D. Samaras, "Shadow removal via shadow image decomposition," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 8578–8587.

[51] M. Cordts et al., "The cityscapes dataset for semantic urban scene understanding," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2016, pp. 3213–3223.