



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

R-YOLO: A Robust Object Detector in Adverse Weather

Mandhadi Nikhitha, Heena Fathima Baig , Ayisha Begum, Dr .D. Shravani

B.E., Department of CME, SCETW, OU Hyderabad, Telangana , India

Associate Professor, ADCE, SCETW, OU Hyderabad, Telangana, India

ABSTRACT: Learning a robust object detector in adverse weather with real-time efficiency is of great importance for the visual perception task for autonomous driving systems. We propose a framework called R-YOLO for improving the YOLO object detector's robustness in adverse weather conditions without the need for annotations. The framework consists of an image quasi-translation network (QTNet) and a feature calibration network (FCNet) for adapting normal weather domain to adverse weather domain gradually. The QTNet generates images that inherit annotations in the normal weather domain and interpolate the gap between the two domains. The FCNet proposes two kinds of adversarial learning-based feature calibration modules to align the feature representations in two domains in a local-to-global manner. The R-YOLO does not change the original YOLO structure and is applicable to all the YOLO-series detectors. The experimental results show that the proposed method outperforms other detectors with dehaze/derain as the preprocessing step and other unsupervised domain adaptation (UDA)-based detectors.

KEYWORDS: quasi-translation network (QTNet), feature calibration network (FCNet), YOLO, R-YOLO.

I. INTRODUCTION

The design of a robust object detector with high precision and speed is of great importance for almost all the perception systems equipped with the vision sensor, such as autonomous driving systems and robots. Driven by the rapid development of convolutional neural networks (CNNs) and the emergence of large-scale annotated datasets, the performance of deep-learning-based object detection has been significantly improved. However, under adverse weather conditions (e.g., hazy and rainy), the images captured by the camera often suffer from quality degradation caused by the suspended tiny particles or fast falling raindrops. The existing object detection models trained with normal weather images often fail to detect objects in adverse weather due to the domain gap between the input images in different weather conditions, leading to traffic accidents and potential dangers. Collecting and annotating a large-scale dataset for object detection in adverse weather is time-consuming and labor-intensive, the existing methods try to solve this problem in two ways. The traditional and immediate way is to restore the hazy and rainy images as a preprocessing step before object detection. However, the existing methods for image dehazing and deraining are usually based on a different theory, and there is no unified restoration framework for both the tasks. In addition, combining the models used by the dehazing, deraining, and detection methods is too complicated and thus can hinder the real-time efficiency of object detection.

Furthermore, it can be difficult to combine the models used by the dehazing, deraining, and detection methods, which can reduce the effectiveness of object detection in real time. Without the need for annotations, we suggest using a framework we call R-YOLO to increase the YOLO object detector's resilience in inclement weather. The framework is composed of a feature calibration network (FCNet) and an image quasi-translation network (QTNet) for gradually adapting the normal weather domain to the adverse weather domain.

II. LITERATURE REVIEW

ImageNet:

ImageNet is like a huge library of pictures, aiming to organize them in a smart way to help computers understand them better. It's based on another system called WordNet and tries to match pictures with words. So far, it has millions of pictures sorted into different categories. Creating such a big collection was tough, and they used a method called Amazon Mechanical Turk to gather the data. ImageNet is expected to help researchers in computer vision a lot.



PASCAL Visual Object Classes (VOC) Challenge:

This challenge provides a set of images with labels to test how well computers can recognize objects in pictures. Since it started in 2005, it's been a reliable way for scientists to see how good their object recognition systems are. In this paper, they talk about what the dataset looks like, how they test different methods, and what they've learned over three years of the challenge.

SSD: Single Shot Multibox Detector:

SSD is a new way to find objects in pictures using a special kind of computer program called a neural network. Instead of doing multiple steps to find objects, SSD does it all in one go, making it faster and easier to use. They tested it on different datasets and found it works just as well as other methods but much faster.

YOLO: Unified Real-Time Object Detection:

YOLO is another method for finding objects in pictures quickly. Unlike older methods, YOLO looks at the whole picture at once, which makes it super fast. It might make a few more mistakes in pinpointing exactly where objects are, but it's really good at not mistakenly thinking something is an object when it's not. YOLO works well even with different kinds of pictures, not just regular photos.

CT-Net: An Efficient Network for Low-Altitude Object Detection Based on Convolution and Transformer:

As drones become more common, there's a need to detect them, especially when they're flying low and might be invading privacy or posing a safety risk. This new method, CT-Net, uses a mix of convolution and a special kind of attention called a transformer to find drones accurately. It's also designed to work fast and use less computer power. When tested, it performed really well, even beating other top detectors.

S.N O	TITLE & AUTHORS	METHODOLOGY	PROPOSED SYSTEM	CONS	CONCLUSION
1	<p>TITLE: CT-Net: An Efficient Network for Low-Altitude Object Detection Based on Convolution and Transformer</p> <p>AUTHOR: Tao Ye, Jun Zhan et.al.,</p> <p>LINK: https://ieeexplore.ieee.org/document/9751710 (2022)</p>	<p>The proposed CT-Net method combines an attention-enhanced transformer block (AETB) to enhance feature extraction, a lightweight bottleneck module (LBM) to reduce computational load, and a directional feature fusion structure (DFFS) for multiscale object detection. It achieves high accuracy in low-altitude UAV detection, outperforming YOLOv5.</p>	<p>CT-Net is a novel deep learning approach for low-altitude small-object detection, featuring an attention-enhanced transformer, lightweight modules, and directional feature fusion. It offers superior performance over existing methods, especially when processing small objects.</p>	<p>1. CT-Net may require substantial computational resources due to its transformer-based architecture. 2. High-quality training data for low-altitude small-object detection may be limited.</p>	<p>The CT-Net method, with its attention-enhanced transformer and lightweight modules, demonstrates remarkable performance in low-altitude UAV detection and outperforms existing detectors. The results confirm its feasibility for real-world applications, improving personal privacy and public safety in the face of unauthorized drone flights.</p>



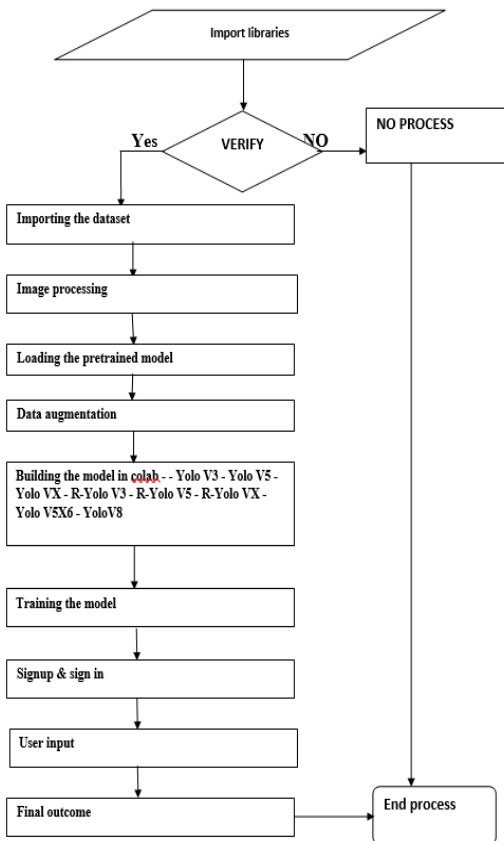
2	<p>TITLE: Three-Channel Infrared Imaging for Object Detection in Haze</p> <p>AUTHOR: Beinan Yu, Yifan Chen et.al.,</p> <p>LINK: https://ieeexplore.ieee.org/document/9745969 (2022)</p>	<p>The proposed system captures shortwave infrared (SWIR) images with two filter types: liquid crystal tunable filters (LCTF) and optical filters. Narrowband SWIR images obtained with LCTF are processed into wideband images. Optimal SWIR bands are determined using the recognition and localization (RL) score and an efficient search algorithm. Three custom optical filters are used for acquiring three-channel SWIR images.</p>	<p>An imaging system captures SWIR images with LCTF and optical filters to enhance object detection in hazy conditions. Optimal SWIR bands are selected based on the RL score, and three custom optical filters are used for acquiring three-channel SWIR images. This system improves detection performance under heavy haze compared to traditional RGB images.</p>	<p>1. SWIR imaging equipment and filters may not be readily available or affordable for widespread use. 2. The system requires precise calibration and maintenance to ensure accurate SWIR image acquisition.</p>	<p>The proposed SWIR imaging system, using LCTF and custom optical filters, improves object detection under heavy haze compared to RGB images. Optimal SWIR bands selection via the RL score and efficient algorithms enhances the system's performance. This technology has great potential for applications in scenarios with challenging visibility conditions.</p>
3	<p>TITLE: Robust-FusionNet: Deep Multimodal Sensor Fusion for 3-D Object Detection Under Severe Weather Conditions</p> <p>AUTHOR: Cheng Zhang; Hai Wang et.al.,</p> <p>LINK: https://ieeexplore.ieee.org/document/9831988 (2022)</p>	<p>The proposed multimodal sensor fusion network addresses issues in object detection under severe weather conditions. It introduces a pointwise aligned data fusion method using K-means++ clustering for accurate data alignment, an implicit feature pyramid network (i-FPN) to suppress exposure changes, and a hybrid attention mechanism (HAM) to enhance adaptability. Experiments are conducted on ONCE and KITTI datasets.</p>	<p>The system combines LiDAR and camera data for object detection in challenging weather conditions. It employs a pointwise aligned data fusion method, implicit feature pyramid network (i-FPN), and hybrid attention mechanism (HAM) to improve alignment, exposure handling, and adaptability. Experimental results on ONCE and KITTI datasets demonstrate enhanced performance.</p>	<p>1. Performance improvements may vary based on the quality and diversity of training data. 2. The system may face challenges when dealing with unforeseen weather conditions or environmental factors.</p>	<p>The proposed multimodal sensor fusion network effectively enhances object detection performance under both clear and severe weather conditions by addressing issues related to data alignment, exposure changes, and adaptability. Experimental results on ONCE and KITTI datasets confirm its robustness and potential for improving environment perception in autonomous driving scenarios.</p>
4	<p>TITLE: YOLOv7:</p>	<p>YOLOv7's success is attributed to its</p>	<p>YOLOv7-E6 is a real-time object</p>	<p>1. YOLOv7 demands</p>	<p>YOLOv7-E6 stands out as a top-</p>



	<p>Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors</p> <p>AUTHOR: Chien-Yao Wang, Alexey Bochkovski</p> <p>LINK: https://arxiv.org/abs/2207.02696 (2022)</p>	<p>architecture improvements, including PANet, SAM, and PANet-SS. It utilizes a self-paced curriculum learning strategy and object-aware feature selection to boost performance.</p> <p>YOLOv7 is trained on MS COCO from scratch, with no pre-trained weights or additional datasets.</p>	<p>detector that outperforms rivals with superior accuracy and speed. It achieves a 55.9% AP on a GPU V100 at 56 FPS, surpassing transformer-based and convolutional-based detectors by significant margins.</p>	<p>powerful GPUs for optimal performance.</p> <p>2. Its training solely on MS COCO may limit performance in specific domains.</p>	<p>performing object detector, offering exceptional speed and accuracy without relying on pre-trained weights or additional datasets. While it excels in many aspects, it may require significant computational resources and poses challenges in terms of domain adaptation and interpretability.</p>
5	<p>TITLE: YOLOv5-Fog: A Multiobjective Visual Detection Algorithm for Fog Driving Scenes Based on Improved YOLOv5</p> <p>AUTHOR: Hai Wang; Yansong Xu et.al.,</p> <p>LINK: https://ieeexplore.ieee.org/document/9851677 (2022)</p>	<p>This work addresses the challenge of automatic driving perception in foggy conditions. A synthetic fog dataset is created from virtual scenes and depth information. The improved YOLOv5-based detection network features a modified ResNeXt backbone and a feature enhancement module (FEM) with an attention mechanism to boost accuracy and speed in fog scenes.</p>	<p>The proposed fog multitarget detection network, based on improved YOLOv5, outperforms the original model. It achieves 77.8% accuracy on the RTTS public dataset and operates at a speed of 31 frames per second, a 14-frame improvement over YOLOv5.</p>	<p>1. The effectiveness of the system should be further tested in diverse real-world fog scenarios.</p> <p>2. The improved model may require substantial computational resources.</p>	<p>The proposed fog multitarget detection network, built upon YOLOv5, demonstrates significant improvements in both detection accuracy and speed for foggy driving conditions. While promising, further real-world testing and considerations for computational resources and adaptability are needed to enhance its practical applicability.</p>

III. METHODOLOGY

In previous work they introduced a method for removing rain streaks from a single image. They proposing a semi-automatic method for generating a high-quality clean image dataset from real rain images and a novel SPatial Attentive Network (SPANet) for removing rain streaks in a local-to-global manner. Their method is evaluated against state-of-the-art deraining methods and performs favorably. The contributions of their work include a large-scale dataset of rain/rain-free image pairs and a new approach for accurately identifying and removing real rain streaks from single images. Some methods use the image-to-image (I2I) translation methods based on generative adversarial networks (GANs), e.g., CycleGAN, to translate the source images before global feature alignment for alleviating the negative transfer.



Proposed System:

We propose a framework called R-YOLO for improving the robustness of the YOLO object detector in adverse weather conditions without the need for annotations. The framework consists of an image quasi-translation network (QTNet) and a feature calibration network (FCNet) for adapting normal weather domain to adverse weather domain gradually. The QTNet generates images that inherit annotations in the normal weather domain and interpolate the gap between the two domains. The FCNet proposes two kinds of adversarial learning-based feature calibration modules to align the feature representations in two domains in a local-to-global manner. The R-YOLO does not change the original YOLO structure and is applicable to all the YOLO-series detectors.

REQUIREMENTS:

The following are the hardware and software requirements that have used to implement the proposed system

HARDWARE REQUIREMENTS

- 1) Operating System : Windows Only
- 2) Processor : i5 and above
- 3) Ram : 4gb and above
- 4) Hard Disk : 50 GB

SOFTWARE REQUIREMENTS

- 1) Visual Studio Community Version
- 2) Nodejs (Version 12.3.1)
- 3) Python IDEL (Python 3.7)

Advantages:

1. Our work (R-YOLO) does not require any annotations for adapting the normal weather domain to the adverse weather domain, making it more efficient and less prone to human bias.
2. Our model utilizes an image quasi-translation network (QTNet) to generate images that gradually adapt the normal weather domain to the adverse weather domain. This allows for a more flexible and continuous adaptation to various adverse conditions, potentially leading to better robustness.
3. However, we proposes a feature calibration network (FCNet) that employs adversarial learning-based feature calibration modules to directly align the feature representations in two domains in a local-to-global manner.
4. Our work (R-YOLO) is specifically designed to improve the robustness of YOLO object detectors in adverse weather conditions without changing the original YOLO structure. This makes R-YOLO more accessible and adaptable for different YOLO-series detectors.

Disadvantages:

1. The problem is that the GAN-based image translation methods suffer from the notorious problem of instability training, and the generated images may contain pixel distortion as the semantic consistency cannot be guaranteed.
2. The existing work relies on a semi-automatic method for generating a clean image dataset from real rain images. This process might involve manual annotation, which can be time-consuming and may introduce subjective errors.
3. The existing work creates a large-scale dataset of rain/rain-free image pairs. However, the size of this dataset may still be limited, and it might not cover all possible adverse weather conditions comprehensively.
4. The existing work used image-to-image (I2I) translation methods based on generative adversarial networks (GANs) like CycleGAN. This additional step could introduce artifacts or inconsistencies during the translation process, affecting the deraining performance.

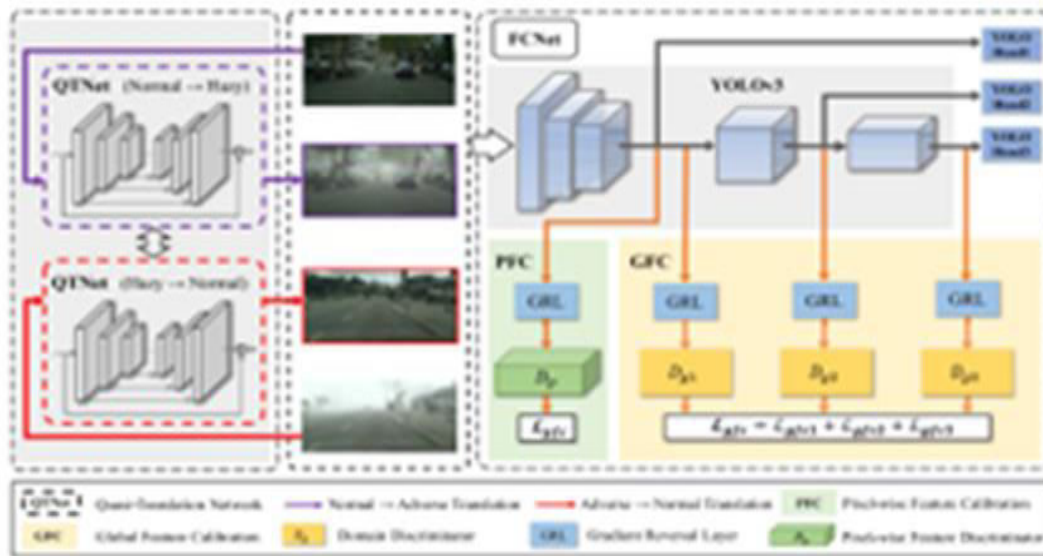


Fig.1: System architecture

MODULES:

To implement this project we have designed following modules.

1. Data exploration: using this module we will load data into system
2. Image processing: Using the module we will process of transforming an image into a digital form and performing certain operations to get some useful information from it.
3. Data augmentation: using this module used to address both the requirements, the diversity of the training data, and the amount of data
4. Model generation: Building the model in colab - Yolo V3 - Yolo V5 - Yolo VX - R-Yolo V3 - R-Yolo V5 - R-Yolo VX - Yolo V5X6 - YoloV8.
5. User signup & login: Using this module will get registration and login
6. User input: Using this module will give input for prediction

7. Prediction: final predicted displayed

Extension work:

In the base paper, the author mentioned to use different techniques for analysis the datasets for detection with different models in which r-yolov5, mAP got 49%. However, we can further enhance the performance by exploring other techniques such as Yolov5x6 and YoloV8, from which yolo v5x6 may get 55% mAP or above, With the above As an extension we can build the front end using flask framework for user testing with authentication.

IV. IMPLEMENTATION

ALGORITHMS:

Yolo V3:

YOLOv3 (You Only Look Once version 3) is an advanced tool for spotting objects quickly and accurately in both pictures and videos. It uses a special kind of neural network to do this job all at once, without needing to go back and forth multiple times. Basically, it looks at an image or video and instantly figures out where things are, like cars or people, and how likely they are to be what it thinks they are. It works fast because of its smart design, which includes different layers that help it understand the size and shape of objects in the image. This makes it super useful in things like self-driving cars, security cameras, and robots. For example, in a self-driving car, YOLOv3 can quickly identify pedestrians, other vehicles, traffic lights, and signs, helping the car navigate safely. Compared to older methods like R-CNN, which are slower because they need to go over the image multiple times to find objects, YOLOv3 is much quicker because it does everything in a single go. This speed makes it perfect for situations where quick responses are crucial, like avoiding accidents on the road.

YoloV8 :

YOLOv8, the newest addition to the YOLO family of object detection models, emerged in 2021 courtesy of Alexey Bochkovskiy and his team. This iteration, following in the footsteps of YOLOv7 and YOLOv6, introduces several upgrades in computer vision technology. Notably, YOLOv8 boasts enhanced precision and swiftness compared to its predecessors. These improvements stem from a larger, more complex neural network architecture trained on extensive datasets. Moreover, YOLOv8 adopts the "swish activation" method to expedite network convergence during training. Another pivotal enhancement lies in its innovative backbone architecture, CSPDarknet, engineered to streamline information flow with fewer convolutional layers without compromising accuracy. Additionally, YOLOv8 integrates features like data augmentation, learning rate schedules, and refined training strategies to further bolster its performance. Overall, YOLOv8 stands as a cutting-edge object detection algorithm, elevating both accuracy and speed and earning it widespread recognition across diverse applications in computer vision.

YoloV5 :

YOLOv5 utilizes a Convolutional Neural Network (CNN) backbone to extract features from images. These features are combined within the model and passed to the head. The head of the model then analyzes these combined features to determine the category of an image. YOLO stands for You Only Look Once, and we're using Version 5, which was introduced by Ultralytics in June 2020 and is currently the most advanced object detection algorithm. It's a unique CNN that accurately identifies objects in real-time. This method employs a single neural network to process the entire image, breaking it down into sections to predict bounding boxes and probabilities for each part. These bounding boxes are weighted based on their expected probability. The approach "looks once" at the image because it makes predictions after just one forward pass through the neural network. It then provides the detected items after applying non-maximum suppression, ensuring that each object is identified only once.

V. RESULTS AND DISCUSSIONS

During our project on robust object detection in adverse weather conditions using YOLO algorithms, we encountered various results and discussions. One significant finding was the algorithm's ability to maintain a reasonable level of accuracy even in challenging weather environments, such as heavy rain or fog. However, we also observed instances where the detection performance decreased under extreme weather conditions, particularly with objects partially obscured by snow or rain droplets. Through discussions, we explored potential strategies to improve the model's robustness in adverse weather, including data augmentation techniques specific to weather conditions and fine-tuning model parameters to enhance detection capabilities. Additionally, we discussed the importance of considering environmental factors in the training process to ensure the algorithm's effectiveness in real-world scenarios. These

discussions led to insights on potential areas for further research and development to advance the robustness of object detection algorithms in adverse weather using YOLO.

VI. CONCLUSION

- The project achieved robust object detection, specifically targeting adverse weather conditions, enhancing safety and reliability in practical applications like autonomous driving. Implemented YOLO-based algorithms (V5, YOLOX, R-YOLO variants) for accurate object detection in diverse scenarios, ensuring adaptability and performance. We Explored V5x6 and V8 models to enhance robustness and accuracy, contributing to a comprehensive and adaptable object detection system in challenging conditions.
- Integrated a user-friendly Flask front-end for efficient testing. Users input images, and the model's outcomes are displayed, enhancing practical usability and accessibility.
- Applications relying on vision sensors, such as autonomous driving and robotics, benefit from the project's robust object detection, improving safety and reliability in real-world scenarios.

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.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details