International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(13s) 2026 pp. 60-75
https://doi.org/10.70917/ijcisim-2026-1803
Article

A Hybrid Technique for Robust Object Detection for Identification of Classifying Objects

Bhawna Narwal[1], Anil Garg[2], Shikha Bhardwaj[3] and Ravinder Singh[4]

[1] Department of Electronics and Communication Engineering, Maharishi Markandeshwar Engineering College, Maharishi Markandeshwar Deemed to be University, Mullana, Ambala, Haryana, India, 133207
bhawnanarwal9@gmail.com
[2] Department of Electronics and Communication Engineering, Maharishi Markandeshwar Engineering College, Maharishi Markandeshwar Deemed to be University, Mullana, Ambala, Haryana, India, 133207
mmuanilgarg@gmail.com
[3] Department of Electronics and Communication Engineering, University Institute of Engineering and Technology, Kurukshetra University, Kurukshetra, Haryana, India, 136119
sbhardwaj2015@kuk.ac.in
[4] Department of Computer Science Engineering, Terminal Ballistics Research Laboratory - Ramgarh, DRDO, Chandigarh, India, 160003
ravinder007@gmail.com

Abstract: Object detection in real-world environments faces significant challenges from blur, noise, occlusion, and varying illumination, which degrade detection accuracy and robustness. This paper proposes an enhancement-assisted Faster R-CNN technique that integrates classical image restoration with deep learning to improve detection under degraded conditions. The method applies Wiener filtering to recover lost structural details, followed by adaptive histogram equalization for contrast restoration. Boundary-box aware augmentation & optimized anchor estimation to improve localization exactness across diverse object scales during training. At inference, Soft-NMS and TTA enhance prediction stability in crowded scenes. The system is built on a ResNet-50 backbone within the Faster R-CNN architecture and is evaluated under three conditions: original, noisy & blurred images. Results demonstrate mAP improvement from 93.57% to 98.41%, a 3.5% increase in detection rate, and enhanced PSNR and SSIM scores. These findings confirm that integrating image restoration with optimized detection strategies significantly improves robustness without architectural modifications.

Keywords: Faster R-CNN, Image Enhancement, Augmentation, Anchor tuning, Soft-NMS (Non-Maximum Suppression), Test-Time Augmentation (TTA)

1. Introduction

An object identification and tracking are crucial tasks that is vital in computer vision through which machine discern and understand visual images. With these techniques, systems can find objects of interest in pictures or videos and keep track of their actions overtime [1]. Today, modern object recognition methods are primarily driven by deep learning, especially through the use of CNN (convolutional neural networks) [1]. Generally, these approaches fall into two categories:

- Single-stage detector - These techniques predict bounding boxes and class probabilities in a single step, which leads to quicker inference times. Example includes YOLO & SSD [2]. Figure 1 illustrates single-stage detector.

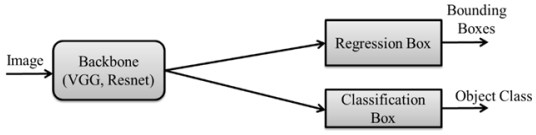

Figure 1. Single-stage Detector [2]

- Two-stage detector - These methods first create recent proposals and then classify them. A well-known example is Faster R-CNN [2]. Figure 2 depicts dual-stage detector.

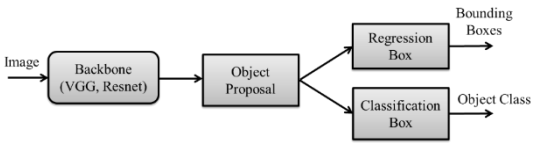

Figure 2. Dual-stage Detector [2]

The problem of object detection in computer vision is still fundamental, but it has applications in autonomous driving, surveillance, medical imaging and remote sensing. Low contrast, illumination variations, & limited annotated data are common issues with real world pictures, despite significant progress made by deep learning-based detector leading to poor detection performance [2] [3].

Vehicle detection is essential for intelligent transportation systems, autonomous driving, and traffic surveillance. Correct detection must be dependable in terms of different lighting, background noise, obstruction, and spatial resolutions [3]. Utilizing region proposal mechanism have demonstrated that two-stage detectors are typically more precise in localization than one-stage detector.

The two-stage detector of faster R-CNN is popular, still with its limited capabilities, due to the fact that it provides a balance between accuracy and computing power. Lack of training data, insufficient scaling of anchor boxes to object size and over aggress NMS suppression may all affect performance [4]. The degradation of performance is not however a bad process.

The robust feature appearance and strong localization of Faster R-CNN make it popular, yet it relatively slows two-stage detector [4]. Why? The reason is that lack of diversity in training prevents generalization, fixed anchor boxes fail to match object scale distribution and non-maximum suppression (NMS) can suppress valid detection in crowded scenes crowded scenes [4] [5]. Figure 3 depicts the simplified version of the building block of standard Faster R-CNN model.

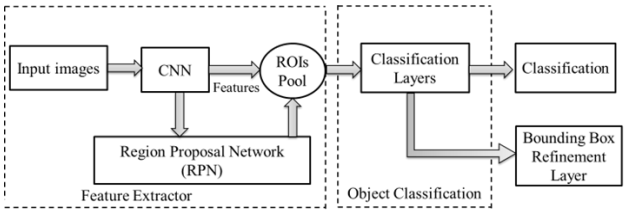

Figure 3. Architecture of Standard Faster R-CNN [5]

This work focus on enhancing the detection model utilizing a hybrid technique based on Faster R-CNN that improves degraded images before detection and employs the concept of boundary-box aware augmentation with optimized anchors to ensure correctness and spatial uniformity across the training phase. Additionally, Soft-NMS and test-time augmentation are useful in fine tuning predictions during inference to improve the reliability and the robustness of detection in adverse environments such as a blur and noise. The key contribution of this paper are as follows: -

- Image Enhancement: apply a pre-detection image restoration pipeline that includes wiener filtering and adaptive histogram equalization (CLAHE) to restore structure/edges of degraded images with contrast improvement before they enter into the training phase. This enables the model to deal with real world conditions such as blur and noise & that are not easily handled by the standard detectors.
- Boundary-box aware augmentation with optimized anchor estimation: during training phase, utilized boundary-box aware augmentation that deals with different types of transformations to ensure label consistency across all augmented images. To generate more accurate region proposals for objects of varying different aspect ratios and scales, utilizing anchors estimation via calculating mean IoU (Intersection over Union).

- Dual inference refinement and robustness validation: Soft-NMS and test-time augmentation is utilized at inference to minimize suppression of nearby objects and improve prediction consistency.

The remaining portion of the paper is arranged in this manner: section 2 summarizes the pertinent literature, section 3 elucidates the model of the proposed methodology, section 4 interpretate the results, and finally section 5 brings the discussion to the conclusion and outlines potential directions for future research.

A. Motivation

Object recognition and tracking have practical stakes across a wide range of fields that are autonomous vehicles, surveillance systems, healthcare, agriculture, & military applications all depend on models that can identify and follow objects reliably under rea-world conditions [5] [6]. Deep learning-based technologies have rapidly growing innovation consists of Faster R-CNN, YOLO, and SSD etc., but these models still struggle with many issues that arise in everyday situations.

Dense scenes, partial occlusions, poor lighting, & objects at very different scales all cause problems [6]. In traffic monitoring, overlapping cars or pedestrians get missed or confused. In drone footage, fast motion and shifting backgrounds degrade tracking in ways that are hard to catch until the results are already wrong [6] [7].

Secondly, detection processes usually rely on hand-crafted anchor boxes, which may not be suitable for the target dataset and may not provide precise suggestions, leading to more false positives and reduces localization correctness [7]. Additionally, hard NMS typically remove overlapping items, which can be particularly problematic in crowded areas such as wildlife monitoring or crowded analysis.

Beyond detection, tracking requires frame-to-frame consistency that detection metrics don't capture [8]. One missed detection can throw off an object's trajectory, and once identity switches start, they tend to accumulate rather self-correct. Moreover, practical implementation requires efficient and lightweight solutions that can operate in real-time on platforms with limited resources, such as edge hardware, aerial vehicles, and internet of things detectors [8].

2. Related WORK

Zhou et al. [9] (2024) introduced an improved anchorless object identification system derived from FCOS to tackle issues in dynamic traffic conditions namely with tiny objects, scaled volatility, and integrate backgrounds. Their approach incorporates a dynamic convolutional module inside the backbone who diminish inter-class feature similarity, facilitating adaptive extraction of features for various road traffic participants. A dual attention method that integrates channel as well as spatial attention is implemented to mitigate background noise and highlight distinguishing characteristics, while a multi-scale combination of features module improves deep representation of characteristics for small and far away objects. Evaluation on the Cityscapes dataset shows that the proposed DF-FCOS method improves mAP by 2.3% compared to baseline FCOS and surpasses multiple states-of-the-art detection methods, especially under crowded and occluded traffic conditions. This work demonstrates the value of integrating dynamic feature extraction with attention mechanisms for achieving reliable traffic object detection.

Wang et al. [10] (2023) suggested an anchor-free compact finding object network to solve the substantial computational expenses and restricted small-item identification capacity of existing anchor-free approaches. Their methodology substitutes the standard FCOS backbone via MobileNetV3 to decided decrease model variables & inference duration while incorporating a little object augmentation influenced by SSH to augment the representation of features for diminutive targets. Furthermore, a self-training label-assigning technique and centre-correcting mechanism are developed to increase accurate localization by picking ideal feature points along with modifying predicted box boundary centers. Evaluation conducted using Pascal VOC & MS COCO datasets reveal that the suggested method obtains competitive correctness with considerable improvements in tiny-object discrimination while preserving the lightweight properties appreciate for real-time applications.

Mittal [11] (2024) reviewed lightweight deep learning detectors designed for edge devices, focusing on the trade-off between detection precision and computational cost efficiency. The survey covers dual-stage, one-stage, no anchor, & lightweight detection models, in conjunction with widely utilized lightweight core architectures including Mobilenet, Shufflenet, & Efficient Network. It compares cutting-edge lightweight detectors on prevalent standards like MS COCO along with Pascal VOC and examine learning and inference difficulties on limited-resource edge systems. The study shows that a lot of development has been accomplished, but getting high precision for tiny objects while keeping minimal latency along with minimal usage of power is still a research problem for current edge use cases.

Mumuni et al. [12] (2022) surveyed data augmentation techniques in computer vision to address the persistent problem of limited and unrepresentative training data in deep learning.

In addition to advanced methods such as area-level augmentation, attribute-space augmentation, algorithm-based GAN analysis, and neural rendering, and 3D visual data generation, the authors present a new classification of augmentation techniques such as data transformation, information synthesis, and meta-learning-driven approaches and present them. Qualitative and quantitative assessments in different vision tasks, advantages, disadvantages, and application scenarios, as well as major challenges and suggestions for future work to ensure reliable and effective model predictions.

Ren et al. [13] (2016), introduced a Faster R-CNN, which is a dual-stage object identification framework. This method combines a RPN (Region-based Proposal Network) and a speedy R-CNN detector into one unified architecture. RPN works with the detector by sharing convolution features, allowing it to create high-quality region proposals through the use of multi-scale anchor boxes and it effectively removes the computational problem of conventional proposal method like selective search. The outcomes were performed on datasets like Pascal VOC & MS COCO, it shows that Faster R-CNN obtains state-of-the-art-identifying accuracy although operating in nearby real-time, providing a fundamental framework for contemporary region-based object identifier.

YOLOv4, which is a single-stage object identification system designed to achieve the best balance between accurate identification and actual-time speed is proposed by Bochkovskiy et al. [14] (2020). This model incorporates CSPDarkNet-53 core, SPP with PAN necks, and comprehensive training and architectural improvements known as BoF (Bag of Freebies) as well as BoS (Bag of Specials), consisting montage data enrichment, CIoU loss, enhanced normalization approaches, and mish activation. Experimental findings from the MS COCO dataset reveal that YOLOv4 achieves an impressive 43.5% average precision (AP) at real-time inference speed. This performance not only surpasses many leading detectors but also a solid foundation for effective real-time object detection systems.

Han et al. [15] (2024) suggested an improved method for detecting small objects in industrial quality control through the optimization of anchor boxes in the YOLOv10 framework, which is used to detect non-metallic foreign objects in tobacco fibres. The anchor box sizes were customized to fit smaller contaminants scales and aspect ratios, with k-means clustering applied to a custom tobacco dataset to overcome the shortcomings of the default anchors used on generic dataset. Experiments results showed significant improvements in detection performance, with significant gains in precision, recall, & mAP at significantly lower inference speeds, which are near real-time, proving the advantage of anchor box optimization in detecting small objects in complex industrial settings.

Hosain et al. [16] (2024) surveyed object detection using modern deep leaning architectures, starting with early feature-engineering approaches. The review organizes detectors into one-stage, two-stage, and transformer-based categories, and covers application domains such as autonomous driving, surveillance, healthcare, and agriculture, as well as datasets, evaluation metrics, and common frameworks. On the challenges side, the paper cites real-time performance, scale variation, obstruction, tiny object detection, and interpretability as areas where present methods still fall short.

Lim et al. [17] (2024) addressed class imbalance in vehicle detection by using an adaptive oversampling strategy designed to better represent minority vehicle classes during training. They tested this method on several one-stage and dual-stage detectors on the Malaysian Traffic Dataset (MMUVD), where YOLOv8n came out ahead in both accuracy and training time. The oversampling approach yielded significant benefits: mAP50 increased from 0.686 to 0.950 and mAP50-95 increased from 0.439 to 0.717, and performance became more consistent across vehicle classes. The results show that correcting data imbalances at the source is a practical and very cost-effective way to improve detection reliability in traffic applications.

Alam et al. [18] (2023) modified Faster R-CNN to handle better scale variation, occlusion, and tiny vehicle detection in real traffic conditions. Their version switches to a new VGG-16 backbone, adds soft NMS, and improves the region proposal mechanism to preserve detail on tiny objects and reduce missed detections. When evaluated on a custom dataset and the KITTI benchmark, the modified model improved mAP and ran faster than the standard Faster R-CNN a fair trade-off, given that the changes remained within the existing architecture.

Zakaria et al. [19] (2022) modified S2A-Net by inserting an ILD (Instance-Level Denoising) module before the feature alignment stage. The idea is fairly straightforward: small, tightly packed, and rotated objects in aerial images are noisy by nature, so the module tries to suppress background interference and separate object features before the detector has to work with them. They also replaced the standard regression loss with a KLD-based alternative, which handles orientation and scale variation more gracefully than traditional approaches. On DOTA v1.0, the modified

model outperformed the original S2A-Net on mean Average Precision, with the clearest gains on small & clustered which is exactly where the baseline struggled.

3. Proposed Methodology

It has been demonstrated that the standard Faster R-CNN model includes a ResNet-50 backbone for effective feature extraction. During inference, it depends on default anchor estimation and hard non-maximum suppression (NMS), without applying any test-time augmentation [20]. The proposed model enhances the standard Faster R-CNN model by making the following improvements:

1. Improving feature quality by incorporating image enhancement.
2. Using boundary-box conscious augmentation to improve the diversity of training.
3. Optimizing anchor-aware detection to enhance proposal generation.
4. Using Soft-NMS instead of Hard-NMS to handle overlapping objects more effectively.
5. To increase the robustness of inference, use test-time augmentation.

An amended approach for spotting objects employing the Faster R-CNN (Region-based Convolutional Neural Network) system is presented in this section. Applied image enhancement techniques, random augmentations during training, anchor detection via mean Intersection over Union (IoU) and advanced post-processing methods, including Soft Non-Maximum Suppression (Soft-NMS) and Test-Time Augmentation (TTA), to improve detection accuracy and robustness. Experiments are conducted on a small vehicle dataset fragmented into training & testing sets, via comparisons between the standard Faster R-CNN and the proposed technique. This approach highlights the efficacy of combining data-level and inference-level enhancements for object detection tasks. The proposed technique follows an eight-stage pipeline that combines classical image processing with deep learning-based detection. The complete workflow is illustrated in Figure 4 and each stage is described below.

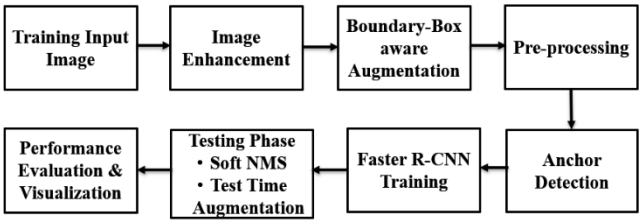

Figure 4. The Design Architecture of the Proposed Technique

❖ Training Input image Acquisition

A small vehicle dataset is used, where each image comes with ground-truth bounding box annotations in (x, y, width, height) format. To remove ordering bias, utilize random shuffled dataset and is equally divided into training and testing subsets. The images are loaded as RGB format in order to retain the complete color information that can be used in the enhancement step [20].

❖ Image Enhancement

Real-world images often suffer from blur and noise caused by camera motion, focus issues, or atmospheric conditions [21]. To restore image quality before detection, image enhancement pipeline is applied.

❖ Gaussian Noise Modeling

Gaussian noise introduces random distortion to pixel values within an image that can be modelled as:

$$I_n(x, y) = I(x, y) + N(x, y) \quad (1)$$

Where, *In* (x, y) is the noisy image, *I (x, y)* is the original image, *N (x, y)* ~ N (0, σ) is the Gaussian noise with 0 mean & variance σ2. This noise increases the difference between the original image and the noisy image, thereby leading to higher values of the MSE, and reducing the PSNR and SSIM values, which is a sign of deterioration on the image quality [21].

❖ **Wiener Filtering**

To recover edge sharpness from the blurred image, wiener filtering is also applied:

$$I_{wiener} = \frac{H^*}{|H^2|+K} I_b \quad (2)$$

Where, H is the frequency-domain representation of the PSF, H^* is its complex conjugate, and $K = 0.001$ is the noise-to-signal ratio.

Wiener filtering minimizes the mean square error (MSE) between the restored and original image [21] [22].

❖ **LAB-Space Contrast Enhancement**

After deblurring, the image is converted from RGB to the CIE LAB color space. Adaptive Histogram Equalization (CLAHE) is applied only to the luminance channel (L) to improve local contrast without altering color balance:

$$L_{enhanced} = adapthisteq(L) \quad (3)$$

The enhanced image is then converted back to RGB and clipped to [0, 1] to prevent any out-of-range values [22].

Enhancements to the contrast-based image are added as a preprocessing step, with the aim of improving discriminability among features under low-contrast conditions and different illumination levels. In this enhancement, image intensity distributions are adjusted to emphasize the structural details and object boundaries while conserving semantic content [23]. The use of enhanced images for training is exclusive to this method, resulting in improved strength without any test-time bias. The improved set is then random-sized and divided into training and testing subsets to uphold consistency in statistical terms. Poor visual quality is addressed by this step, which often results in a decrease of object recognition performance [23]. Fig. 5 depicts the comparison between blur vs noisy vs enhanced images.

Figure 5. Blur vs Noisy vs Enhanced Images

❖ **Training–Testing Data Split**

The dataset is categorized with equal distribution of data in the training and testing sets, making the evaluation impartial and proportional to fairness. Before partitioning, random permutation is utilized to eliminate order bias [24]. These datasets have equal splits that separate them, allowing for direct comparison under same experimental conditions. Bounding-box labels are organized in the box label datastore format to ensure integration with the Faster R-CNN training workflow.

❖ **Boundary-box-aware Augmentation**

Standard augmentation methods apply transformations to images different from their annotations, causing the bounding box coordinated after the transformation to be misaligned with the actual object's positions. This label corruption directly reduces the recognition accuracy by providing incorrect supervision signals during training [24] [25]. To address this, the augmentation strategy ensures that every transformation performed on the image is applied simultaneously and consistently to its corresponding bounding box coordinated, maintaining complete alignment between the image content and its label throughout the augmentation process.

Augmentation schemes include random horizontal flipping, scaling, cropping, geometric transformations, & photometric intensity adjustments. For each transformation, the bounding box coordinates are updated using the same transformations parameters applied to the image, ensuring that the boxes remain tightly aligned with the object regions at all times [25]. All augmentations are performed in real-time during training via a transform-based data pipeline, meaning the model sees a different augmented version of each image at each epoch without any increase in storage requirements.

After each augmentation step, a validation check is performed on the box that is created. Any box that extends beyond the image boundary is clipped back to the valid image region, and any box whose area falls below the minimum size is removed to eliminate bad annotations that would introduce noise into the training process. This combined strategy effectively increases the diversity of the training data, improves the model's ability to generalize across different object scales, positionings, & lighting conditions, and maintains annotation integrity throughout. Figure 6 shows the input and output images of boundary-box-aware augmentation.

Figure 6. Input and Output images of Boundary-Box-Aware Augmentation

❖ **Pre-processing**

Before feeding the images into the network, all images are resized to 224*224*3 pixels using bilinear interpolation to match the expected input dimensions of the ResNet-50 backbone. This setting is used for both training and testing pictures. To maintain the spatial resolution, the boundaries are scaled accordingly [26]. Pre-processing transforms all pictures to the similar dimensions and applies normalization to ensure consistent numerical ranges, which promotes stable optimization and fast convergence during network training. Additionally, this standardization step ensures that the processed data aligns with the input requirements of the anchor estimation module and the Faster R-CNN training pipeline [26]. Figure 7 shows the resized image to 224*224*3.

Figure 7. Resize Image (224×224×3)

❖ **Anchor detection**

Correct anchor box configuration is essential for successful operation of the Region Proposal Network (RPN). Why? This is because average intersection over union (IoU) is used to evaluate different anchor configurations, with the aim of optimizing their selection [27]. For training purposes, the number of anchors that produce the highest average IoU with the ground-truth bounding boxes is selected. More anchor-aware optimization is responsible for increasing the quality of proposals, reducing localization error and increasing recall, especially for objects with different scales and aspect percentages [27].

Intersection-over-Union (*IoU*) is calculated by

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (4)$$

A predicted bounding box is considered to be correct if IoU > □, where □ is the IoU threshold usually 0.5.

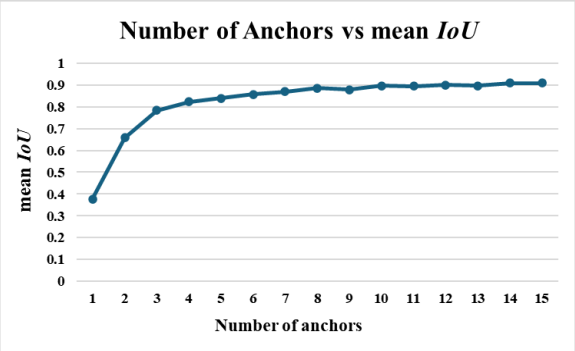

Figure 8. Anchor-box Detection via mean IoU

Figure 8 illustrates that how the average Intersection over Union (IoU) changes with the number of anchor boxes used during the anchor estimation process. Anchor configurations ranging from 1 to 15 were evaluated to find the best number of anchors that effectively represented the scale and aspect ratio distribution of the objects in the dataset.

Experimental results show that when the anchor count is increased from 1 to 3, the average IoU increases significantly, which confirms that using too few anchors is insufficient for modelling the diverse range of object geometries and scales in the dataset [27] [28]. This swift improvement demonstrates the need for different anchors to create high-quality region proposals. After reaching threes anchors, the average IoU continues to increase, but at a slower pace, eventually stabilizing around 8 to 10 anchors.

Using more anchors increases the average IoU to around 0.90 to 0.91, indicating that adding more anchors beyond this point will yield diminishing returns in terms of localization & accuracy that is the quality of region proposals will not improve significantly, & will acquire additional computational stress [28]. Using this analysis, select the anchor configuration having maximum mean IoU to train the proposed model.

The process of optimizing anchor selection helps to align anchor boxes more accurately with the actual object boundaries, leading to better localization exactness and complete detection performance.

❖ **Faster R-CNN Training**

- The object detector is built on the Faster R-CNN model with the following configuration:
- Backbone: It uses ResNet-50 [28]. Take features from the activation_40_relu layer
- Input size: The input size is 224*224*3
- Optimizer: The model utilizes Stochastic Gradient Descent with Momentum, which is also known as SDGM [28]

Some key training settings are as follows:

- Primary learning rate: 0.0001
- Training epochs: The model is trained for 85 epochs
- Mini-batch size: The mini-batch size is 1
- IoU range: It is between 0.6 and 1
- Negative IoU range: It is between 0 and 0.3

Faster R-CNN architecture has four components that work together: It has a feature extractor that uses the ResNet-50 backbone. The region proposal network (RPN) generates object regions. Then the ROI pooling layer takes these proposals and extracts features of a fixed size. Lastly, the detection head classifies objects & adjusts the bounding box [28].

The Faster R-CNN model is trained at once and it uses a GPU for faster processing. The training is done in MATLAB [28]. The Faster R-CNN architecture and the ResNet-50 backbone play a role in the detection model. The model depends on the RPN to generate candidate object regions [29].

❖ **Testing Phase**

There are two strategies i.e Soft-NMS and TTA used in testing phase for improving the detection accuracy beyond hard NMS that is used in standard model. These are utilized by the proposed technique to enhance detection robustness during inference:

Soft-NMS (Non-Maximum Suppression): Confidence scores in Soft-NMS are decayed based on the magnitude of overlap, rather than being completely eliminated [29]. Instead of completely removing overlapping detections, Soft-NMS reduces their confidence scores using a Gaussian decay function based on their overlap with the highest-scoring box:

$$s_i = s_i . e^{-\frac{IoU^2}{\sigma}} \quad (5)$$

Where *si* is the confidence score of detection *i*, *IoU* is its overlap with the top-scoring box, and *σ* controls the decay rate. This prevents objects that are close together or partially occluded from being incorrectly suppressed, which improves recall in crowded traffic scenes [29].

Test-Time Augmentation (TTA): Inference is applied to both horizontally flipped test images and those that were originally formed. The combination of multiple passes' predictions and refinement through Soft-NMS results in the output of a final detection. Without retraining the model, this approach enhances the models' ability to adapt to viewpoint shifts [29]. This ensemble technique reduces prediction variance and thus, improving localization accuracy.

❖ **Performance Evaluation**

Assessing object detection systems is about finding out how well a model can spot and recognize objects in an image [30]. The system is tested under three conditions: original image, noisy image & blurred image. The following metrics are used:

- **Peak signal to Noise ratio (*PSNR*):** this measures the image restoration quality in decibels [30]

$$PSNR = 10log_{10}(\frac{MAX^2}{MSE}) \quad (6)$$

- **Mean Squared Error (*MSE*):** this measures pixel-level reconstruction error [30]

$$MSE = (\frac{1}{MN})\sum(I - \hat{I})^2 \quad (7)$$

- **Structural Similarity Index Measure (*SSIM*):** this measures the perceptual similarity in terms of luminance, contrast & structure between the original and restored images [30]
- **mean Average Precision (*mAP*):** it is the primary detection metric, averaged across all object classes [31]. Suppose there are N object classes in the dataset:

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad (8)$$

Where AP_i means the Average Precision for the i-th group

Table 1. Outcome Matrix for Object Finding Guesses [31]

Prediction vs Ground Truth	Object Present	Object Absent
Predicted Object	*TP* (True Positive)	*FP* (False Positive)
No Prediction	*FN* (False Negative)	*TN* (True Negative) (not used in detection)

Precision & Recall metrics come from Table 1 outcomes for each finding.

- **Precision** — *Precision* tells us how accurate those detections are—when precision is high, it means the detector rarely makes mistakes in recognizing objects [31].

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

- **Recall** — *Recall* measures how thoroughly the model finds all the actual objects present. A high recall specifies that the model successfully detects most of the actual objects out there [31] [32].

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

- **Total detection rate (*Detection %*) —** the overall rate of correct detections [32].

Total recognition area is calculated by the entire recognition percentile:

$$Detection\ (\%) = \frac{\sum TP}{\sum GT} * 100 \quad (11)$$

where $\sum TP$ is the total correctly detected object and $\sum GT$ is the total ground-truth object

- **Robustness drop:** This measures how much blur there is, which indicates a decrease in performance [32] [33].

$$Drop\ (\%) = \frac{mAP_{orig} - mAP_{blur}}{mAP_{orig}} * 100 \quad (12)$$

4. Experimental results and analysis

The experiments were implemented using MATLAB R2025b, 4Gb memory and 64-bit windows. A small vehicle dataset is used for these experiments and some sample pictures are shown in Figure 9 [3]. This dataset contains 295 images taken from real-world driving situations, that shows city streets, intersections, highways, and open roadside areas. This dataset is divided into training and testing subgroups in a 50:50 ratio. This is done in MATLAB with GPU acceleration [34]. The model uses a ResNet-50 network, trained over 85 epochs with a batch size of one.

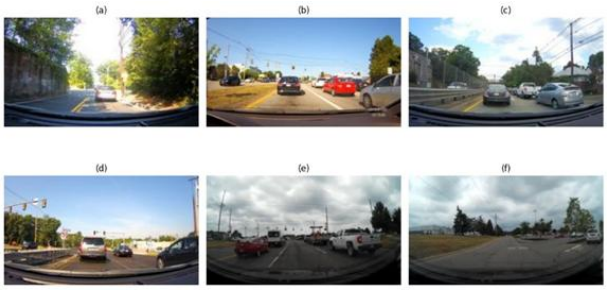

Figure 9. Some sample imageries from the Dataset [3]

A. Evaluation Metrics

This section examines how well the proposed method performs against the standard Faster R-CNN model. Standard object detection metrics are utilized for evaluating performance, which includes mean Average Precision (mAP), average recall, average precision, and the overall detection percentage [34] [35]. To keep things fair, both models are trained and tested under the same experimental conditions.

The performance of each model is measured using common metrics that look at how accurately it locates objects, how complete its detections are, and how trustworthy the predicted bounding boxes are [35].

B. Results and Analysis

1) Image Quality Evaluation

The image quality is quantitatively measured using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Mean Squared Error (MSE), and Shannon Entropy for evaluating the picture enhancement of the proposed pre-processing pipeline [36]. These metrics were calculated under two techniques: (i) Wiener-filter that is used for deblurring the image (ii) LAB-based CLAHE that is used for picture enhancement. Table 2 shows the quantitative results of PSNR, SSIM, MSE, Entropy under the degradation and enhancement situations.

Table 2. Image Quality Metrics

Condition	PSNR (dB)	SSIM	MSE	Entropy
Enhance Image	25.6132	0.8774	0.002746	7.8009
Blurred Image	20.5570	0.8500	0.008796	7.7983
Noisy Image	14.6775	0.5406	0.034061	7.7077

Comparing the image quality in the three conditions using quantitative analysis proves that the presence of gaussian noise is the most harmful form of degradation as it reduces the image quality by 0.5406 and 14.68 dB in SSIM and PSNR respectively, while the use of gaussian blur has a medium degradation effect as the SSIM and PSNR are 0.8500 and 20.56 dB respectively. The proposed enhancement pipeline has the best performance in all metrics, including PSNR of 25.61 dB, SSIM of 0.8774, MSE of 0.002746, & entropy of 7.8009 and hence, the preprocessing stage effectively recovers the structural fidelity and enhances the picture information content, which provides the detection backbone with better feature representation than any worst-case scenario.

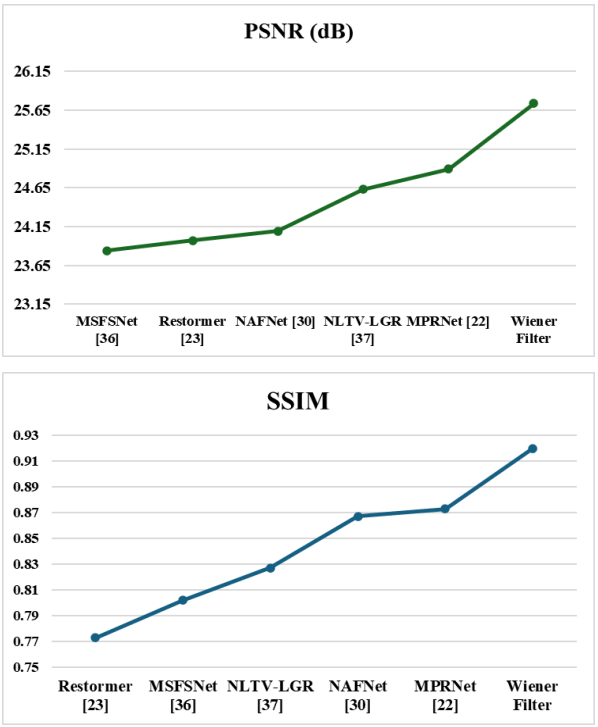

Figure 10. Comparative Analysis of PSNR and SSIM with State-of-the-Art Techniques

A systematic comparison of image restoration techniques measured by PSNR & SSIM over the period 2022 to 2025 is illustrated in Figure 10. Existing techniques including MPRNet [22], Restormer [23], NAFNet [30], MSFNet [36], and NLTV-LGR [37] deblurring methods, reported PSNR values ranging from 23.84 to 24.90 dB and SSIM values ranging from 0.773 to 0.873 [37]. Wiener-based deblurring filter achieves an PSNR of 25.61dB & SSIM of 0.877. Although the resulting restoration quality image is higher, because this technique is fast to calculate, requires zero GPU assistance, and processing time per image is less than 100ms, which matches the requirements of real-time object detection. The balance between quality and speed seems to be right as the later detection part compensates for this with robust feature extraction, boundary-box improvements and better anchor guessing [37].

2) Detection Performance Comparison

Experimental evaluations in three test conditions that are original, noisy & blurred images consistently demonstrate that the proposed model outperforms the standard Faster R-CNN model in both absolute recognition accuracy and degradation robustness [38].

On the original images, the proposed method achieves a mAP of 0.9814, which is 5.14% higher than standard model and improves the overall detection percentage by 3.5% as shown in Figure 11. Under noisy and blurring

degradation, the proposed model exhibits a measurable low robustness degradation, which confirms that the integration of frequency-domain deblurring, perceptual contrast enhancement, Soft-NMS & TTA produces a detection pipeline that is resilient to the picture degradation most commonly encountered in real-world vehicle monitoring applications [38]. These results establish the proposed model as a practical and effective approach for robust object detection under poor imaging conditions. Table 3, 4, 5 demonstrate the detection performance of mAP, recall & precision under three different situations that are original, noisy, & blur images.

Table 3. Detection Performance under Original Images

Method	**mAP (%)**	**Recall**	**Precision**
Standard Faster R-CNN	0.9357	0.4898	0.9876
Proposed Work	0.9814	0.5118	0.9959

Table 4. Detection Performance under Noisy Images

Method	**mAP (%)**	**Recall**	**Precision**
Standard Faster R-CNN	0.8355	0.4444	0.9705
Proposed Work	0.9251	0.4874	0.9858

Table 5. Detection Performance under Blur Images

Method	**mAP (%)**	**Recall**	**Precision**
Standard Faster R-CNN	0.7755	0.4017	0.9861
Proposed Work	0.8214	0.4240	0.9942

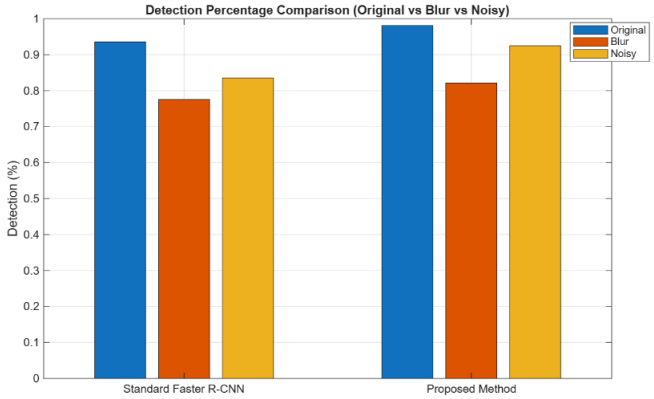

Figure 11. Detection Percentage Comparison between Standard Faster R-CNN and Proposed Method (Original vs Blur vs Noisy)

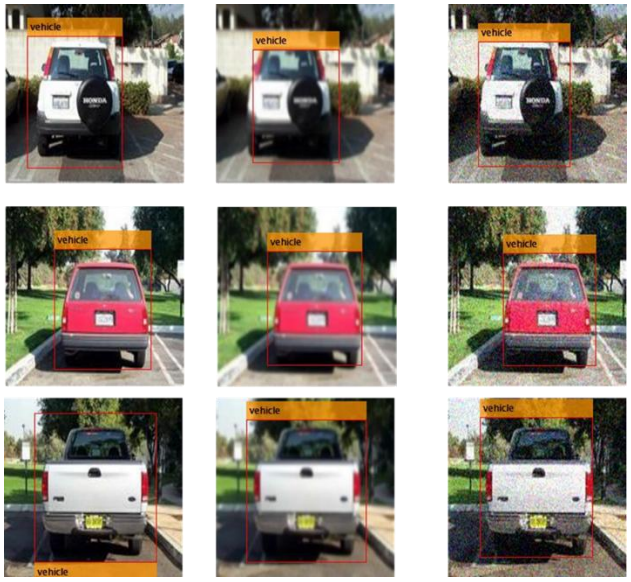

Figure 12. Object Detection results under three different image conditions (a) Original (b) Blur (c) Noisy images

Figure 12 illustrated the identified and classified object with boundary-box under three different imaging conditions: (a) gaussian blur-degraded image with reduced edge sharpness, (b) gaussian noise-degraded image with low contrast & grainy texture, (c) original image with clear appearance. The technique successfully detects and localizes objects with consistent bounding-box accuracy and obtains high confidence scores across all three imaging situations.

3) Robustness Analysis

The proposed model particularly enhances the robustness of recognition against blur and noise degradation. This model reduces the average precision (mAP) drop and it achieves from 17.12% to 16.30% and 10.70% to 5.74% in the blur and the noise condition. This significant reduction in noise-related degradation actually reflects the improved stability of the detection process [39]. These results clearly show that it achieves better mAP robustness drop without changing the basic architecture of Faster R-CNN.

C. State-of-the-art Comparison

The performance of the proposed system has been compared with other methods to authenticate the system. Fig. 13 provides a systematic comparison of the propsed system and the methods of detection recently published during the period of 2022 to 2025. Existing techniques, including optimized versions of Faster R-CNN, transformer-based models, & YOLO-based detectors shows mAP values ranging from 0.67 to 0.922 [39]. In contrast, the standard Faster R-CNN with ResNet-50 baseline used in this study achieves an even higher mAP of 93.57%, while the proposed model maintains a solid mAP of 98.14%. These results emphasize that this approach not only maintain competitive mAP but also improves detection accuracy under different situations through the use of image restoration, boundary-box aware augmentation, and optimized post-processing although preserving the original network architecture.

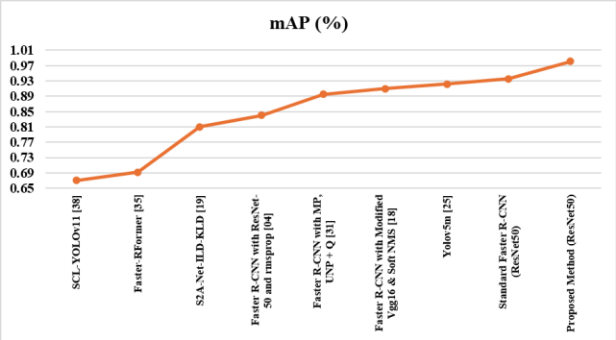

Figure 13. Comparison Analysis of Proposed Method (mAP) with State-of-the-Art Techniques

As shown in Figure 14, the latest detection techniques show precision values ranging from 0.86 to 0.97. In particular, optimized Faster R-CNN and transformer-based models can reach an accuracy of up to 0.97 [40]. In contrast, the standard Faster R-CNN with ResNet-50 baseline used in this study achieves an even higher precision of 0.9876, while the proposed model maintains a solid precision of 0.9959. This shows that there is a better balance between detection accuracy and robustness. These results emphasize that this approach not only maintain competitive precision but also improves detection stability under different situations.

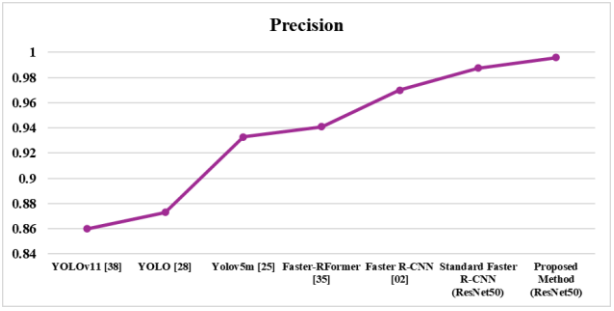

Figure 14. Comparison of Proposed Technique (Precision) with State-of-the-Art Techniques

5. Conclusion

This study introduced an enhancement-assisted Faster R-CNN technique that integrates wiener filtering, adaptive histogram equalization, boundary-box aware augmentation, optimized anchor estimation, soft-nms, and test-time Augmentation to address object detection under degraded imaging conditions. This model is implemented in MATLAB and evaluated on a small vehicle dataset across original, blurred, and noisy scenarios. Results demonstrate that image restoration consistently improves PSNR and SSIM, while the complete model increases mAP from 93.57% to 98.41% and detection rate by 3.5%, with minimal robustness degradation under blur compared to standard Faster R-CNN—all without architectural modifications. These findings validate that combining classical image processing with deep learning effectively enhances detection resilience in practical applications. Future work will explore end-to-end trainable restoration networks, extension to video sequences, nighttime detection, and adverse weather scenarios to further assess the technique's generalizability across diverse real-world conditions.

Declaration of competing interest

All authors declare that they have no conflicts of interest.

References

1. M. A. Elhosseini et al., “A hybrid YOLOv10—Faster R-CNN framework for mobility-aid detection and traffic optimization in disability-inclusive smart cities”, Alexandria Engineering Journal, Elsevier, Vol. 129, 2025, pp. 1279–1298.
2. V. K. Sharma, R. N. Mir, “Saliency guided Faster-RCNN (SGFr-RCNN) model for object detection and recognition”, Journal of King Saud University – Computer and Information Sciences, Elsevier, 2022.
3. P. Sun et al., “Sparse R-CNN: An End-to-End Framework for Object Detection”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 45, no. 12, 2023, pp. 15650-15664.

4. G. D. Deepak, S. K. Bhat, “Optimization of deep learning-based faster R-CNN network for vehicle detection”, Scientific Reports, 2025, pp. 15:38937. Weber, M. & Perona, P. Caltech Cars 1999 (1.0) [Data set]. CaltechDATA. https://doi.org/10.22002/D1.20084.
5. K. Vijiyakumar, V. Govindasamy, V. Akila, “An effective object detection and tracking using automated image annotation with inception based faster R-CNN model”, International Journal of Cognitive Computing in Engineering, Vol. 5, 2024, pp. 343–356.
6. V. K. Vatsavayi, N. Andavarapu, “Identification and classification of wild animals from video sequences using hybrid deep residual convolutional neural network”, Multimedia Tools and Applications, Springer, 2022.
7. K. Tarkasis, K. Kaparis, A. C. Georgiou, “Enhancing Trustworthiness in Real Time Single Object Detection”, Information Systems Frontiers, Springer, 2025.
8. Z. Cai, N. Vasconcelos, “Cascade R-CNN: Delving into high quality object detection”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 43, no. 5, 2021.
9. T. Zhou, W. Liu, H. Li, “A feature enhancement FCOS algorithm for dynamic traffic object detection”, Connection Science, Taylor & Francis, Vol. 36, no. 1, 2024, pp. 2321345.
10. W. Wang, Y. Gou, “An Anchor-Free Lightweight Object Detection Network”, IEEE Access, Vol. 11, 2023.
11. P. Mittal, “A comprehensive survey of deep learning-based lightweight object detection models for edge devices”, Artificial Intelligence Review, Springer, 2024.
12. A. Mumuni, F. Mumuni, “Data augmentation: A comprehensive survey of modern approaches”, Array, Elsevier, Vol. 16, 2022, pp. 100258.
13. S. Ren, K. He, R. Girshick, J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016.
14. A. Bochkovskiy, C. Y. Wang, et al., “YOLOv4: Optimal speed and accuracy of object detection”, IEEE Access, 2020.
15. S. Han, W. Liu, et al., “Improving Small Object Detection in Tobacco Strands Using Optimized Anchor Boxes”, IEEE Access, Vol. 11, 2024.
16. M. T. Hosain, A. Zaman, M. R. Abir, et al., “Synchronizing Object Detection: Applications, Advancements and Existing Challenges”, IEEE Access, Vol. 11, 2024.
17. C. H. Lim, T. Connie, et al., “Visual-based vehicle detection with adaptive oversampling”, International Journal of Information Technology, Springer, Vol. 16, 2024, pp. 4767–4777.
18. M. K. Alam, A. Ahmed, R. Salih et al., “Faster RCNN-based robust vehicle detection algorithm for identifying and classifying vehicles”, Journal of Real-Time Image Processing, Springer, Vol. 20, no. 93, 2023.
19. Y. Zakaria et al., “Improving Small and Cluttered Object Detection by Incorporating Instance Level Denoising into S2A-Net”, IEEE Access, Vol. 11, 2022.
20. A. Chaudhuri, “Smart traffic management of vehicles using a faster R-CNN-based deep learning method”, Scientific Reports, Vol. 14, 2024, pp. 10357.
21. K. Hu, F. Lu, M. Lu, et al., “A Marine Object Detection Algorithm Based on SSD and Feature Enhancement”, Complexity, Wiley, 2020, pp, 5476142.
22. S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M. H. Yang, “Multi-Stage Progressive Image Restoration”, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 14816-14826.
23. S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M. H. Yang, “Restormer: Efficient Transformer for High-Resolution Image Restoration”, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 5718-5729.
24. L. Alzubaidi, J. Zhang, et al., “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions”, Journal of Big Data, Springer, 2021, pp. 8:53.
25. A. Samy Talaat, S. Sappagh, “Enhanced aerial vehicle system techniques for detection and tracking in fog, sandstorm, and snow conditions”, The Journal of Supercomputing, Springer, Vol. 79, 2023, pp. 15868–15893.
26. T. Sharma, A. Chehri, I. Fofana, et al., “Deep Learning-Based Object Detection and Classification for Autonomous Vehicles in Different Weather Scenarios of Quebec, Canada, IEEE Access, 2024.
27. L. Liu, W. Ouyang, X. Wang, et al., “Deep learning for generic object detection: A survey”, International Journal of Computer Vision, Springer, 2021.
28. S. P. Yadav, M. Jindal, P. Rani, V. H. C. de Albuquerque, C. dos Santos Nascimento, and M. Kumar, “An improved deep learning-based optimal object detection system from images”, Multimed Tools Appl, Springer, Vol. 83, no. 10, 2024pp. 30045–30072.
29. N. Bodla, B. Singh, R. Chellappa and L. S. Davis, “Soft-NMS — Improving Object Detection with One Line of Code”, IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 5562-5570.
30. L. Chen, X. Chu, X. Zhang, J. Sun, “Simple Baselines for Image Restoration”, Lecture Notes in Computer Science-ECCV, Springer, Vol. 13667, 2022.
31. A. Magdy, M. S. Moustafa, H. M. Ebied, M. F. Tolba, “Lightweight faster R-CNN for object detection in optical remote sensing images”, Scientific Reports, Springer, Vol. 15, 2025, pp. 16163.
32. P. Tsirtsakis, G. Zacharis, et al., “Deep learning for object recognition: A comprehensive review of models and algorithms”, International Journal of Cognitive Computing in Engineering, Elsevier, Vol. 6, 2025, pp. 298–312.
33. N. H. Abdulghafoor, H. N. Abdullah, “A novel real-time multiple objects detection and tracking framework for different challenges”, Alexandria Engineering Journal, Elsevier, Vol. 61, 2022, pp. 9637–9647.

34. N. V. Rao, D. V. Prasad, M. Sugumaran, “Real-time video object detection and classification using hybrid texture feature extraction”, International Journal of Computers and Applications, Taylor & Francis Group, 2018.
35. X. Kong, X. Li , X. Zhu , Z. Guo , L. Zeng, “Detection model based on improved Faster-RCNN in apple orchard environment”, Intelligent Systems with Application, Elsevier, Vol. 21, 2024, pp. 200325.
36. H. Gao, D. Dang, “Exploring Richer and More Accurate Information via Frequency Selection for Image Restoration”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 35, no. 3, 2025, pp. 2689-2700.
37. J. Sun, H. Shen, Q. Yuan, L. Zhang, “Super-Resolution for Remote Sensing Imagery via the Coupling of a Variational Model and Deep Learning”, IEEE Transactions on Geoscience and Remote Sensing, Vol. 63, 2025.
38. S. Zhuo, H. Bai, L. Jiang, X. Zhou, X. Duan, Y. Ma, “SCL-YOLOv11: A Lightweight Object Detection Network for Low-Illumination Environments”, IEEE Access, Vol. 13, 2025, pp. 47653-47662.
39. J. Bai, S. Li, L. Huang, H. Chen, “Robust Detection and Tracking Method for Moving Object Based on Radar and Camera Data Fusion”, IEEE Sensors Journal, Vol. 21, no. 9, 2021.
40. Y. Xiong, Y. Ge, P. Johan, “An improved obstacle separation method using deep learning for object detection and tracking in a hybrid visual control loop for fruit picking in clusters”, Computers and Electronics in Agriculture, Elsevier, Vol. 191, 2021, pp. 106508.