Impact of Raindrops on Camera-Based Detection in Software-Defined Vehicles

Yichen Luo, Daoxuan Xu, Gang Zhou, Yifan Sun, Sidi Lu
Department of Computer Science
William & Mary
Williamsburg, VA, USA
{yluo11, dxu05, gzhou, ysun25, sidi}@wm.edu

***Abstract*—Raindrops adhering to windshields or camera lenses substantially impair visibility, leading to significant camera-based detection challenges for software-defined vehicles in both daytime and nighttime conditions. Addressing the impact of raindrops is thus crucial. This work begins by classifying four prevalent types of raindrops within the BDD100K dataset, identifying microsphere raindrops as particularly impactful in rainy conditions. We then conduct a quantitative analysis focusing on the density and diameter of raindrops, underscoring the pronounced impacts of small-density raindrops on detection performance. To mitigate raindrop interference, we introduce and assess the SR3 model for raindrop removal, applying it to both synthetic raindrop-degraded data and real-world rainy data. Besides, we propose YOLO-RA, a novel and fast model to resolve the issues of missing small-size objects and erroneous detections in irrelevant regions. Next, a novel pipeline that combines SR3 with YOLO-RA markedly improves accuracy and processing speed. Finally, we discuss our experimental observations extensively and offer detailed explanations, contributing to understanding SDVs' operational effectiveness in adverse weather conditions.**

***Index Terms*—software-defined vehicles, adverse weather, raindrops, computer vision, impact mitigation**

I. Introduction

Software-defined vehicles. With the rapid development of automotive electronics, automotive hardware systems will gradually become standardized and unified, while automotive algorithms will become the core of smart cities [1]. In this context, software-defined vehicles (SDVs) have attracted massive attention from both academic and industry leaders [2]. For instance, Tesla [3], General Motors [4], Ford [5], BMW [6], Mercedes-Benz [7], Toyota [8], Audi [9], Arm [10], Amazon [11], and Aptiv [12] brought SDVs to the spotlight.

Camera-based algorithms. These mainstream SDVs, particularly autonomous vehicles, a popular category within SDVs, rely heavily on cameras for scenario understanding and decision-making [13]. A typical example is the "Tesla Vision Approach" [14], i.e., only using cameras to "see" the road and neural networks that are supposed to mimic the way a human brain works. Hence, ensuring the performance of camera-based algorithms under various conditions is paramount.

The influence of raindrops. However, it is widely acknowledged that adverse weather conditions diminish the quality of captured images, adversely affecting the performance of algorithms dependent on camera images [13]. Prior research

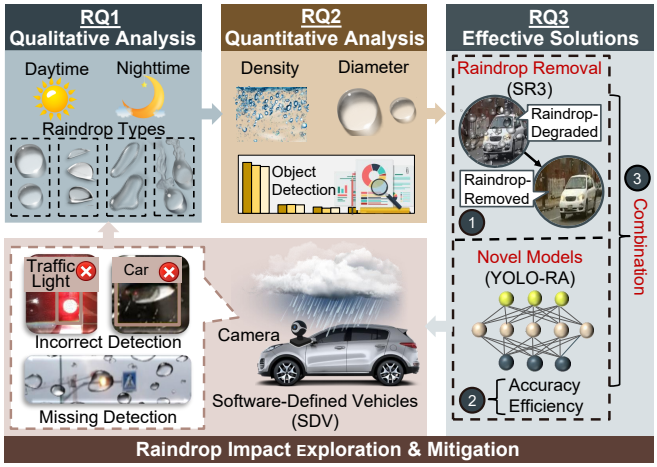


Fig. 1. A succinct overview of our structured methodology for understanding and mitigating the impacts of rain on software-defined vehicles (SDVs). It includes three main research questions (RQs) related to qualitative analysis (top-left), quantitative analysis (middle-top), and effective solutions (top-right). RQ1 examines the effects of four distinct raindrop types under daytime and nighttime conditions. RQ2 conducts a quantitative assessment based on the density and diameter of raindrops. As to RQ3, we introduce an advanced raindrop removal model (SR3) and propose a novel algorithm (YOLO-RA) for accurate and fast object detection based on raindrop-degraded images. We also synergize SR3 with YOLO-RA to assess the combined enhancement.

[15], [16] categorizes these conditions into two broad types: steady (e.g., fog, mist, and haze) and dynamic (e.g., rain, snow, and hail), in which rain is a common occurrence and significant source of camera-based perception degradation [16]–[19]. Raindrops adhering to a vehicle's windscreen significantly degrade scene visibility and obstruct the camera's view, potentially leading to inaccuracies in object recognition by computer vision models [20], [21]. This highlights the critical need for improved systems capable of functioning optimally, even in adverse weather conditions [16]–[21].

Research gaps. Despite the widespread occurrence of raindrop disturbances across numerous scenarios and their potential to severely impair visibility for both humans and algorithms, the field remains relatively underexplored. Only a few articles have focused on adherent raindrops recently [22], [23]. While hardware solutions like glass heaters and wipers can physically remove raindrops, there is a growing need for algorithm-based strategies for cost savings and autonomous processing scenes.

Time-sensitive vehicle services. Furthermore, the cornerstone of SDV systems lies in their ability to make reliable decisions in real-time. For instance, to effectively avoid an obstacle 5 meters away, the system's computing latency must not exceed 164 milliseconds [24]. Excessive delays could result in significant safety risks [25]. Consequently, any solution designed to mitigate the impact of raindrops on SDV systems must not only be effective in enhancing visibility and detection accuracy but also sufficiently rapid to adhere to the stringent temporal constraints inherent in real-time vehicular operation.

In this work, we tackle the following three main research questions (RQs) for SDVs (as shown in Fig. 1).

RQ1: Qualitative Analysis: Which type of mainstream raindrops causes the most significant impact on camera-based object detection during both daytime and nighttime conditions?

RQ2: Quantitative Analysis: For the type of raindrop with the greatest impact, which characteristic has a greater effect on camera-based object detection: variations in density or diameter of the raindrops?

RQ3: Effective Solutions: How can we effectively and quickly mitigate the impact of raindrops on the performance of object detection?

The core innovation of our study is in providing experimental evidence to answer the above three research questions for the SDV's camera-based detection under rainy conditions. Specific contributions are listed as follows.

- We first undertake a comprehensive qualitative analysis to investigate the effects of four distinct raindrop types on various object detection tasks during daytime and nighttime. We discover that microsphere raindrops exhibit the most substantial impact. The experimental results also underscore the variable impact of raindrop types under different lighting conditions (described in Sec. IV-C).
- We proceed with a quantitative analysis of the raindrops' impacts, varying in diameter and density, on diverse object detection. It reveals that low-density and large-diameter conditions predominantly influence detection outcomes. Furthermore, when examining the relative impacts of density and diameter, our results conclusively show that the effect of density on detection performance is more pronounced than that of diameter (Sec. V-C).
- To mitigate the impacts of raindrops, we introduce and assess SR3 for raindrop removal across real-world rainy datasets and synthetically generated raindrop-degraded data. Besides, we develop YOLO-RA, a novel algorithm designed for accurate and fast mitigation of raindrop impacts, which shows a superior mAP (0.85), F1 score (0.82), inference time (6.1ms), and FPS (118). The integration of SR3 with YOLO-RA and subsequent ablation studies further validate the notable improvements achieved in object detection performance under rainy conditions for SDVs (presented in Sec. VI-C).
- Finally, we provide an extensive discussion of our experimental observations, offering in-depth explanations for the results and identified trends. This detailed analysis enhances the understanding of the intricacies involved in our experiments and contributes to the broader knowledge base regarding object detection under adverse weather conditions (described in Sec. VII).

II. Related Work

A. Rain Streak Removal

Many studies have aimed at enhancing visibility in rainy conditions by primarily targeting rain streaks [26]–[29]. However, raindrops possess distinct shapes and physical effects, setting them apart from streaks. Therefore, techniques optimized for streak removal often fall short when applied to the more complex challenge of raindrop mitigation.

B. Raindrop Removal

Several previous works focus on raindrop detection. The principal component analysis (PCA) has been utilized to learn the raindrop characteristics [30], and the maximally stable extremal regions have been applied to detect raindrop candidates [31]. However, these methods primarily emphasize detection rather than the raindrop removal.

Since 2012, researchers have begun investigating raindrop removal. For instance, earlier studies have analyzed the color and texture characteristics of raindrops for identification purposes, followed by employing image inpainting techniques for their removal [32]. However, this area has been hindered by the limited availability of datasets and the challenge of controlling parameters (e.g., shape and refraction).

In this context, generative adversarial networks (GAN) have been employed to address dataset limitations [33]. However, the network may become trapped in local minima during training, resulting in overly similar samples and a reduction in diversity among the generated images.

Recent studies have increasingly treated rain in single images as a noise reduction problem. The denoising diffusion probabilistic model (DDPM) [34], [35] shows its strong capability on image inpaintment, which is designed on U-net [36]. This balanced structure is akin to that of GAN. Despite this, only a few researchers have focused on this area, often due to challenges related to image magnification.

To address the above challenges, we focus on utilizing Super-Resolution via Repeated Refinement (SR3) [37] for efficient raindrop removal, a technique that holds promise yet remains relatively untapped in this specific application domain. Based on the DDPM structure, SR3 increases residual blocks and channel multipliers at different resolutions. It works well with various magnification factors and input resolutions [37].

C. The Gap in Previous Work

While the impact of raindrops on SDVs is widely recognized, there is a limited understanding of the distinct impacts of various types of raindrops under daytime and nighttime conditions. Often, both the qualitative and quantitative analyses are overlooked, along with an in-depth exploration of different raindrop parameters.

Moreover, while current object detection algorithms exc on general benchmark datasets in identifying specific objec they face challenges when dealing with raindrops. For exa ple, while PCA [30] focuses on raindrop recognition, it oft mistakenly identifies areas locally similar to raindrops as rai drops themselves [23]. Similarly, YOLOv7 [38] demonstrat high accuracy and speed in typical scenarios but struggles wi the presence of raindrops, resulting in missed detections.

Bearing the above concerns in mind, our research direct addresses the priorities for problem-solving in diverse rai and lighting conditions affecting SDVs. We delve into t impacts of various factors such as raindrop types, density, a diameter through comprehensive quantitative analysis. We i troduce YOLO-RA, an accurate and fast algorithm specifically designed for detecting transportation-related objects on rainy days. We also integrate the SR3 method to establish a synergy between raindrop removal and object detection. This approach not only offers novel solutions for enhancing image quality but also contributes significantly to improving overall object detection performance in adverse weather.

III. Experiment Datasets and Hardware

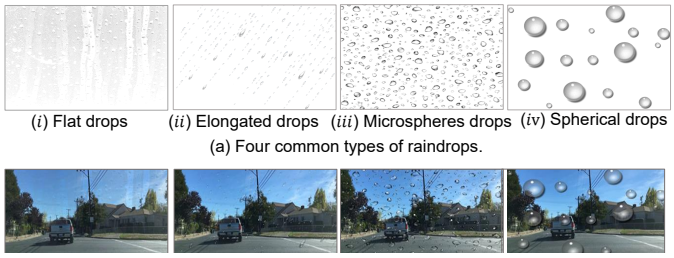
A. Dataset Selection and Related Issues

In this work, we select the BDD100K dataset [39]. Despite its broad scope of driving conditions, the dataset notably lacks substantial data records related to rainy scenarios.

However, purely relying on the BDD100K dataset is unreasonable. To the best of our knowledge, only 53 out of 3,000 records are specifically collected under such conditions. Despite it boasting 100,000 video clips, the data's value is somewhat diminished due to the high similarity between adjacent frames in continuous video collection, leading to redundancy. Most critically, our research aims to delve into the influence of raindrop density and size on SDV applications; however, the BDD100K dataset falls short in offering specific subsets that account for varying sizes and densities of raindrops, presenting a challenge in terms of data diversity and representativeness for rain-related studies.

B. Raindrop-degraded Dataset Generation

To address the aforementioned challenges, we generate a collection of raindrop-degraded images utilizing the BDD100K dataset. We first methodically select images from the 53 rainy records within BDD100K, adopting a one-frame-per-second approach to ensure diversity and minimize redundancy. Through this process, we identified four prevalent types of raindrops that typically adhere to camera lenses during rainy conditions: flat, elongated, microspherical, and spherical raindrops, as depicted in Fig. 2(a). These were added to rain-free BDD100K images, creating a diverse set of raindrop-degraded images shown in Fig. 2(b). This addition significantly enriches the dataset with challenging scenarios for robustness testing and allows for a comprehensive analysis of the different types of raindrops' impact on SDV systems.



(a) Four common types of raindrops.

	NVIDIA GPU Workstation
CPU	15 vCPU AMD EPYC 7543 32-Core
GPU	2 * 24GB GeForce RTX 3090
Frequency	2.8 GHz
Core	15
Memory	80 GB
OS	Ubuntu 18.04 LTS

Fig. 3. The configuration of NVIDIA GPU workstation.

C. Raindrop Image Formation

Investigating the effects of raindrop impacts in real-world situations is challenging because raindrops typically exhibit irregular shapes and are subject to frequent changes. Therefore, following the classification of predominant raindrop types, we extend our investigation to encompass a range of raindrop densities and sizes (diameters). Specifically, we model a raindrop-degraded image as the combination of a background image and the effect of raindrops, factoring in the fluctuations in both raindrop density and diameter:

$$I_{out} = \sum_{h=0}^{N_g-1} \sum_{w=0}^{N_g-1} (1-M) \bullet \alpha B_{hw} \odot I_{hw} \tag{1}$$

$$I_{out} = \sum_{i=1}^{\beta} \sum_{h=0}^{N_g-1} \sum_{w=0}^{N_g-1} (1-M) \bullet B_{hw} \odot I_{hw} \tag{2}$$

Where I represents the rain-free BDD100K image (background), and B denotes the binary template of the selected raindrop. In the raindrop-addition module, the image is divided into N_g grids, and each grid in the input image and the binary template is iterated. When $(1-M) = 1$, it indicates the presence of raindrops in that grid; conversely, $(1-M) = 0$ denotes no raindrops in that region. The variable α represents the diameter size, ranging from $(0, 1]$. We use β to illustrate density, simulating how many times each grid includes raindrops, and it is greater than or equal to 1. Operation $\odot$ means element-wise multiplication.

To meticulously manage variables pertinent to raindrop density and diameter, we adopt specific values for α (0.5 and 1) and β (1 and 4) to emulate varying intensities of raindrop diameter and density, respectively.

D. Hardware Setup

In this work, we assume that an SDV is equipped with an NVIDIA GPU Workstation, serving as the powerful vehicle computing platform (as shown in Fig. 3). It has an AMD

EPYC 7543, which is great for handling complex tasks efficiently by excelling at multitasking. It incorporates two NVIDIA GeForce RTX 3090 graphics cards, each with 24 GB of memory, providing robust parallel processing capabilities that are essential for tasks such as deep learning and advanced graphical computations. The RTX 3090 is renowned for its exceptional performance in professional AI workloads.

IV. Qualitative Analysis of Raindrop Types During Daytime and Nighttime

In addressing **RQ1**, this section explores which type of mainstream raindrops exerts the most significant impact on camera-based object detection under both daytime and nighttime conditions.

A. Proposed Methodology

As detailed in Section III-B, this work focuses on four typical types of raindrops: flat, elongated, microsphere, and spherical. Besides, we select a fundamental application (object detection) as our case study in the context of SDVs. The raindrop-degraded datasets are divided into daytime and nighttime segments to assess object detection performance under varied lighting conditions, comparing against the performance in the corresponding original, rain-free dataset (referred to as "rain-free").

In practical applications of SDVs, accurate and fast object detection is imperative, as inaccurate detection and prolonged latency can significantly contribute to traffic incidents. Recently, YOLOv7 [38] stands out as a recently released model boasting remarkable advancements in terms of object detection. Motivated by the discovery that YOLOv7 achieves a substantial speed improvement, specifically a 120% increase over YOLOv5 [40], we adopt YOLOv7, a cutting-edge method, as our primary object detection model.

To be concrete, we employ YOLOv7 on the five sets of data: rain-free, flat, elongated, microsphere, and spherical, each under both daytime and nighttime conditions. The objective is to accurately detect key categories of transportation-related objects, including cars, traffic lights, trucks, persons, and buses. We commence by establishing a baseline, quantifying the which the detection counts in each category of raindrops are compared.

B. Evaluation Metrics

We analyze object detection counts between original rain-free images and their raindrop-degraded counterparts across different categories. This comparison aims to calculate the decrease rate, by a metric to quantify the adverse impact of raindrop interference on object detection accuracy, a fundamental aspect of SDV functionality. The decrease rate is defined by the following formula:

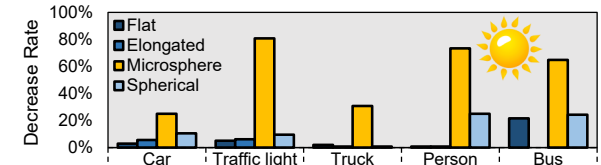
$$Decrease\ Rate = \frac{Num_{typex} - Num_{base}}{Num_{base}} \times 100\% \quad (3)$$

This rate provides a percentage indicating the reduction in detection (bounding box) counts due to raindrop effects, thus offering a clear metric for understanding the extent to which raindrops impair object detection in SDVs.

C. Experiment Results

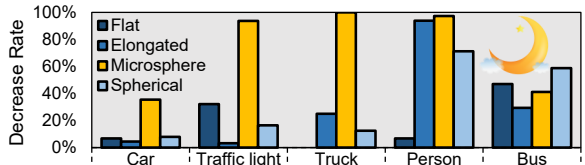
Figure 4 illustrates the variation in object detection decrease rate between the original rain-free dataset and the four types of raindrop-degraded datasets during both daytime (Fig. 4 (a)) and nighttime (Fig. 4 (b)). A significant decrease in detection rates of various objects, such as vehicles, traffic lights, and pedestrians, is observed across all types of raindrop conditions, with the most pronounced effect seen in datasets impaired by **microsphere** raindrops. Specifically, the vehicle detection rate decreases by 25% during daytime and 36% during nighttime under the influence of microsphere raindrops, underscoring the substantial impact of raindrop morphology on the efficacy of object detection systems in SDV applications.

Following microsphere raindrops, spherical raindrops exhibit the second-highest impact, though their effect is less severe, causing approximately 11% and 8% reduction in vehicle detection rates during daytime and nighttime, respectively. Notably, the flat raindrop-degraded dataset shows an anomalous increase in detected traffic lights compared to the baseline (rain-free scenario) by 32%, particularly during nighttime, highlighting that flat raindrops also exert a significant influence. Therefore, addressing the challenges posed



	Car	Traffic light	Truck	Person	Bus
Flat	3%	5%	2%	1%	22%
Elongated	6%	6%	1%	1%	0%
Microsphere	25%	81%	31%	73%	65%
Spherical	11%	10%	1%	25%	24%

(a) Detection quantity during daytime.



	Car	Traffic light	Truck	Person	Bus
Flat	7%	32%	0%	7%	47%
Elongated	5%	3%	25%	94%	29%
Microsphere	36%	94%	100%	97%	41%
Spherical	8%	16%	13%	71%	59%

(b) Detection quantity during nighttime.

Fig. 4. The decrease rate of detection for vehicles, traffic lights, trucks, persons, and buses via YOLOv7: baseline (rain-free BDD100 dataset) *versus* the other four distinct types of raindrop-degraded image datasets in daytime and nighttime.

V. Quantitative Analysis of Density and Diameter

Next, we address **RQ2**: For the raindrop types identified as having the most substantial impact (namely, microsphere and sphere raindrops, as detailed in Section IV-C), we aim to ascertain which characteristic exerts a greater influence on camera-based object detection — the density or the diameter of the raindrops. Understanding and quantifying the influence of these factors is crucial to comprehending the specific impact of raindrops on the performance of SDV detection systems.

A. Experiment Design

1) Comparative Baselines: Recognizing the significant impact of microsphere raindrops (abbreviated as "MR") and spherical raindrops (abbreviated as "SR") on object detection performance, we focus on the density of MR and the diameter of SR as our comparative metrics. Specifically, we define:

(a) Low density: the standard density of MR,
(b) High density: four times the standard density of MR,
(c) Small diameter: half the standard diameter of SR,
(d) Large diameter: the standard diameter for SR.

2) Five Groups: Drawing from the defined baselines for varying levels of density and diameter, we divide five groups of datasets for the comprehensive evaluation. These include:

- Rain-free dataset during sunny and daytime conditions,
- Low-density raindrop-degraded dataset,
- High-density raindrop-degraded dataset,
- Small-diameter raindrop-degraded dataset,
- Large-diameter raindrop-degraded dataset.

B. Proposed Methodology

1) YOLOv7 Structure: YOLOv7 is a sophisticated object detection model consisting of three primary components: i) the backbone network to extract image features; ii) the feature pyramid network (FPN) [28] to generate multiscale feature maps that effectively integrate both high-level semantic and low-level spatial features, thereby maintaining adequate spatial resolution across different object scales [41]; and iii) the head network to output the object classification and location.

Initially, the input image is resized to 640×640 and fed into the backbone network. The backbone integrates several architectural advancements, including CBS modules, Efficient Layer Aggregation Networks (ELAN) modules [42], and MP modules. Here, "CBS" is an acronym for a series of operations: convolution, batch normalization (BN), and the sigmoid-weighted linear unit (SiLU), while "MP" denotes a combination of Maxpool and CBS. The head network then produces three layers of feature maps of dimensions 20×20, 40×40, and 80×80 to capture objects at different scales. To enhance interfacing speed without compromising accuracy, a RepConv [43] operation is incorporated. Finally, each layer outputs detection results in the form of (x, y, w, h, cls), where x, y, w, and h represent the coordinates and dimensions of the bounding box, and cls indicates the class probability.

2) YOLO-RA: In the dynamic and complex driving environment of SDVs, the system frequently encounters a wide array of objects, including smaller, more difficult-to-detect items like traffic signs [44]. While FPN fuses features from different receptive fields, its effectiveness largely depends on the quality of features derived from the backbone network.

To bolster detection capabilities, particularly for small-scale vehicles, we implement a dual strategy: i) applying the Mosaic data augmentation technique [45] and ii) refining the ELAN by adding a Convolutional Block Attention Module (CBAM) [46] within the backbone of YOLOv7, culminating in an enhanced model named **YOLO-RA** (You Only Look Once-Raindrop Analysis). The respective advantages of the Mosaic method and CBAM are listed below.

Mosaic: it randomly merges four images into a single composite. The Mosaic method is able to minimize the presence of irrelevant targets within the image, significantly boosts the accuracy of detecting small-scale objects, and enhances the model's resilience to complex backgrounds. To ensure an equitable performance comparison with YOLO-RA, we exclusively integrate the Mosaic technique into YOLOv7, referred to as **YOLO-Mosaic** in Section VI.

CBAM: Since integrating attention mechanisms has been proven to enrich feature acquisition, we incorporate a CBAM into the ELAN module of YOLOv7 to address the limited contextual understanding associated with small objects. According to prior research, 7×7 convolutions in CBAM outperform 3×3 convolutions [46]. Therefore, we employ a 7×7 kernel in the spatial attention module of CBAM, enhancing the model's potential for improved performance. Similarly, for a direct comparison with YOLO-RA, we exclusively implement CBAM in YOLOv7, and this variant is referred to as **YOLO-CBAM** in Section VI.

Integrate CBAM with ELAN: To be specific, we integrate CBAM with the ELAN module to get a modified ELAN denoted as ELAN-CBAM (as shown in Fig. 5), which is designed to optimize the network's gradient length using a stacked structure within the computational block.

Figure 5 showcases the integration of the CBAM into the

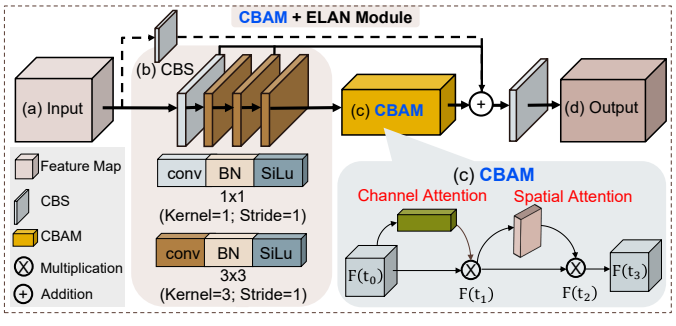


Fig. 5. Integration of CBAM with the ELAN module (ELAN-CBAM). (a) represents the input feature map alongside corresponding legends; (b) details the high-level CBS operations;(c) depicts the dual attention mechanisms within CBAM; and (d) shows the resulting output feature after processing.

passing through a series of convolutional layers, specifically through two types of CBS modules. The first CBS module employs a 1×1 kernel size, and the second uses a 3×3 kernel size. After these initial convolutions, the feature map $F(t_0)$ is processed by the CBAM's channel attention mechanism, which produces a dedicated attention map $F(t_1)$. Subsequently, it is refined by the spatial attention mechanism within CBAM and produces $F(t_2)$. Parallel to this, additional branches of the network with various kernel-sized CBS operations merge with the main flow. The network narrows down into four branches, and the final output feature map is synthesized through a concluding 1×1 kernel CBS operation.

C. Experiment Results

1) Density: We first train both YOLOv7 and the proposed YOLO-RA models using a standard dataset degraded by microsphere raindrops. Subsequently, we then test the models on the unseen datasets affected by low-density and high-density microsphere raindrops to determine their efficacy in identifying transportation-related objects under varied rain conditions. As a benchmark, we consider the detection results of YOLOv7 on the corresponding rain-free datasets as the ground truth (referred to as GT-Sunny).

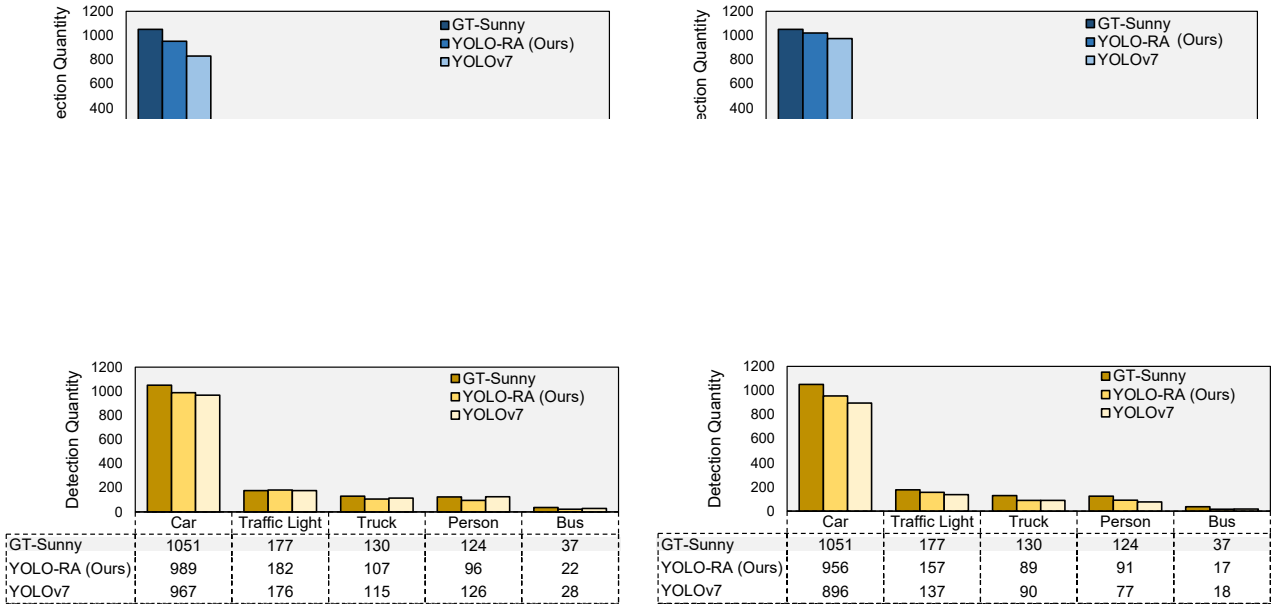
Figure 6 illustrates the detection quantities of five prevalent transportation-related objects (such as vehicles, traffic lights, trucks, pedestrians, and buses) within both low-density (Fig. 6(a)) and high-density (Fig. 6(b)) microsphere raindrop-degraded datasets. These quantities are compared against the ground truth (GT-Sunny) condition. For instance, vehicle detection rates increase by approximately 14% in low-density scenarios, highlighting the significant impact of raindrop density on object detection.

Moreover, when comparing the performance of YOLO-RA and YOLOv7 across the five object types in Fig. 6 (a) and (b), our YOLO-RA model consistently detects more objects than YOLOv7. For example, YOLO-RA detects approximately 120 more vehicles than YOLOv7, underlining the superior efficacy of our proposed YOLO-RA model.

2) Diameter: Mirroring our approach in assessing the impact of raindrop density, we evaluate the trained YOLOv7 and YOLO-RA models on datasets affected by both small-diameter and large-diameter spherical raindrops, as illustrated in Fig. 7.

The comparison between Fig. 7 (a) and Fig. 7 (b) reveals that for both YOLOv7 and YOLO-RA, an increase in raindrop diameter corresponds to a decrease in the number of detected objects. For instance, in scenarios involving large-diameter raindrops (Fig. 7(b)), YOLOv7 fails to detect over 100 vehicles and 40 traffic lights, suggesting that larger raindrops significantly impair object detection for SDVs.

Additionally, when comparing the performance of YOLO-RA and YOLOv7 across various object categories in Fig. 7 (a) and (b), the YOLO-RA model demonstrates a consistent ability to detect more objects than YOLOv7. Notably, the YOLO-RA model significantly reduces missed detections to only 56 vehicles and closely approaches the ground truth in detecting smaller objects and traffic lights, affirming the effectiveness of our proposed model.



(a) Detection on small-diameter raindrop-degraded datasets. (b) Detection on large-diameter raindrop-degraded datasets.

Fig. 7. Detection quantities of five transportation objects (cars, traffic lights, trucks, persons, buses) using YOLO-RA and YOLOv7 on small-diameter and large-diameter raindrop-degraded datasets, compared to the ground truth (GT) of the corresponding sunny dataset (denoted as GT-Sunny). Here, we treat the detection results of YOLOv7 on the original datasets without any raindrops as the ground truth.

(*ii*) Low-density and large-diameter conditions individua exert greater disruptive influences on detection perf mance, as opposed to their respective counterparts, hig density and small-diameter.

(*iii*) Across five primary categories of transportation-rela objects, our YOLO-RA model consistently outperfor YOLOv7 in terms of the number of objects detected.

VI. Raindrop Impact Mitigation

In this section, we address **RQ3**: How can we effectiv and quickly mitigate the impact of raindrops on the perf mance of object detection?

A. Proposed Methodology

1) Raindrop Removal for Visibility Enhancement: To hance the performance and robustness of vision algorithms SDVs during rainy conditions, we initially focus on identifying and implementing effective techniques to remove raindrops from degraded images, thereby restoring them to a clean state. These rain-free images are then input into the SDV's vision algorithms, such as the proposed YOLO-RA, to facilitate decision making under rainy scenarios. The objective is to determine whether this integrated approach, combining raindrop removal with our newly proposed object detection method

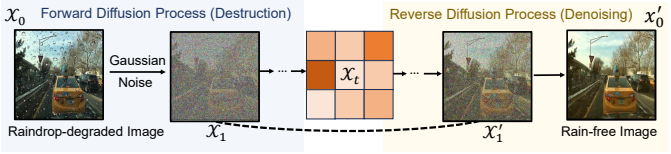


Fig. 8. An illustration of SR3's forward diffusion (destruction) and reverse diffusion (denoising) processes. The destruction process introduces Gaussian noise into the initial raindrop-degraded image X_0, and the denoising process restores the image to its rain-free state X_0' through the reversible Markov chain [47]. The variable t represents an arbitrary point in the timeline.

To remove raindrops, we adopt SR3, solving the generation slow problem for large target resolution tasks in DDPM, through adapting techniques from [48]. Fig. 8 presents the raindrop removal process of SR3, which consists of forward diffusion (destruction) and reverse diffusion (denoising) processes.

2) SR3-Enhanced YOLO-RA Processing: To further enhance the performance of object detection in images impaired by raindrops, which are commonly captured by cameras during inclement weather, we integrate SR3 with YOLO-RA. Specifically, we first remove raindrops using SR3 and then feed the SR3 output into YOLO-RA. This results in more accurate object detection while still maintaining a fast processing speed for SDV applications.

Figure 9 illustrates a comprehensive raindrop removal process integrated with a modified ELAN denoted as ELAN-CBAM, which is part of our proposed YOLO-RA network. The pipeline begins with the raindrop-addition module, where a video is segmented into one-second frames. These frames are subsequently merged with representative raindrop templates to simulate raindrop effects. Following this, the images pass

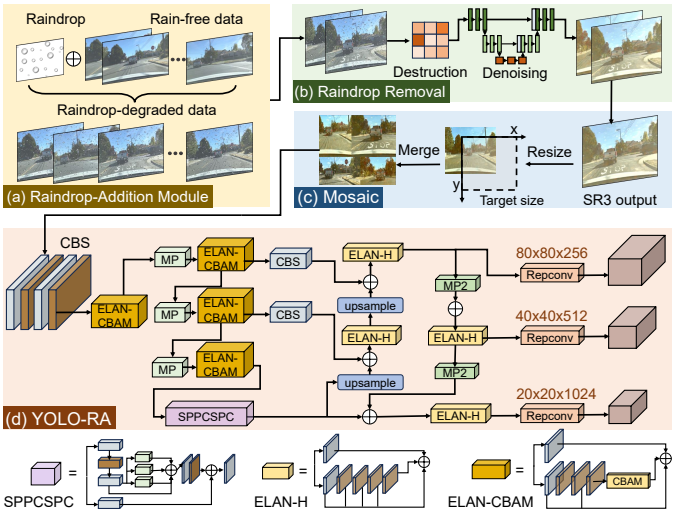


Fig. 9. A pipeline of raindrop removal and enhanced object detection. Initially, in the raindrop-addition module (a), video images are amalgamated with representative raindrop templates. Subsequent to the raindrop removal process (b), four SR3 outputs are combined in the Mosaic stage (c) to form a single image. This pre-processed composite then proceeds to the YOLO-RA module (d) to facilitate precise and fast object detection.

through the raindrop removal process and yield four SR3 outputs, which are then fed into a mosaic algorithm. This mosaic step amalgamates four images into one composite image. The purpose of this mosaic formation is to bolster the model's robustness against complex and cluttered backgrounds. Finally, the images enter the YOLO-RA model, which uses the pre-processed and enhanced images to train for more accurate object detection under raindrop-degraded conditions.

B. Evaluation Metrics

1) Raindrop Removal Metrics: Raindrops are often categorized as a source of visual noise in images. Therefore, we consider the PSNR of the SR3 as the key evaluation metric for raindrop removal. Here, PSNR [49] refers to the peak-signal-to-noise ratio between the original image and the raindrop-removed image), and we use it to evaluate the performance of the raindrop removal model (SR3) [37]. More specifically, given a clean image I of size $m \times n$ and a noisy image K, the mean squared error (MSE) is defined as follows, and the average PSNR of the video group is given by:

$$PSNR = 10 \cdot log_{10}(\frac{MAX_I^2}{MSE}) \tag{4}$$

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i-j) - K(i-j)]^2 \tag{5}$$

Where MAX_I^2 denotes the maximum possible pixel value in the image I, i and j are indices that are used to iterate over the pixels. This formulation allows for the adaptation of the PSNR calculation to various image formats by accurately defining the maximum pixel intensity based on the bit depth.

The aforementioned formulas pertains to grayscale images. For color images, it is imperative to calculate the PSNR for each of the RGB channels independently and subsequently compute the average. Usually, the higher the PSNR (smaller error), the better the quality of the SR3 output.

2) Detection Metrics: Regarding detection metrics, a common way is to compute the intersection-over-union (IoU) [50] between ground truth and prediction:

$$\text{IoU} = \frac{\text{area}\left(BBox_p \cap BBox_{gt}\right)}{\text{area}\left(BBox_p \cup BBox_{gt}\right)}, \tag{6}$$

Where $BBox_{gt}$ represents the bounding box of the ground truth (GT), and $BBox_p$ refers to the predicted bounding box. Predictions whose IoUs are larger than 0.5 are considered as true positives (TP).

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{TP}{\text{all detections}}, \tag{7}$$

$$\text{F1 score} = 2 \times \frac{\frac{TP}{(TP+FN)} \times \frac{TP}{(TP+FP)}}{\frac{TP}{(TP+FN)} + \frac{TP}{(TP+FP)}}, \tag{8}$$

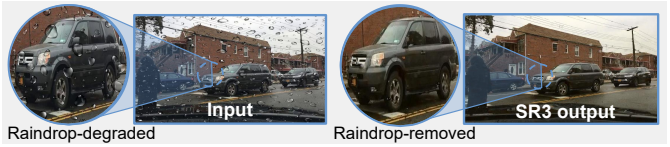
Where TP is the number of detection frames with IoU > 0.5 and FP represents the number of detection frames with IoU $\leqslant 0.5$ detection frames, or the number of redundant detection frames detecting the same GT. FN refers to the number of missing detections.

3) Processing Speed Metrics: As the safety-critical systems, near real-time inference speed is crucial for SDV. To this end, we assess the processing speed of models. We define the total processing time (ms) as $T_{total} = T_{inf} + T_{NMS}$. The term T_{inf} (ms) specifies the time required for the model to process an input image and generate an output, exclusive of any additional post-processing duration. T_{NMS} denotes the time taken for the NMS operation, a critical post-processing phase where the model consolidates its predictions to guarantee singular object detection within multiple bounding box forecasts. T_{total} encompasses the entire processing interval from input to the finalized output after NMS. Besides, we consider another evaluation metric of frame per second (FPS), which signifies the quantity of images processed by the model per second. A higher FPS indicates a faster processing speed, which is desirable for real-time applications of SDVs.

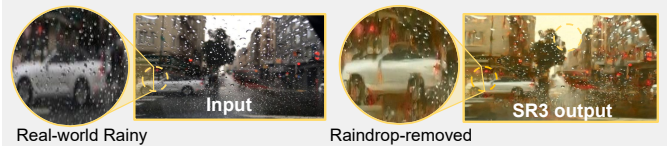
C. Experiment Results

1) Raindrop Removal: We first aim to assess the performance of raindrop removal. Hence, we train and test the SR3 based on the raindrop-degraded datasets and the real-world rainy BDD100K datasets, and we denote the output of the raindrop removal as the SR3 outputs. During the training phase, we discover that during the 80,000 iterations, the value of PSNR was at its peak of 18.75 dB, indicating a significant enhancement in image quality. This peak performance model was subsequently applied to the datasets for raindrop removal.

Figure 10 showcases the performance of the SR3 model in addressing raindrop interference in images. It presents a side-by-side comparison of the original inputs and the resulting clarity SR3 outputs across two distinct datasets: a raindrop-degraded dataset designed for controlled testing (Fig. 10(a)) and the real-world rainy BDD100K dataset, known for its challenging conditions (Fig. 10(b)). The efficacy of our raindrop removal method is further emphasized through close-up views within the blue and yellow circles.



(a) Raindrop removal results of SR3 on raindrop-degraded dataset



(b) Raindrop removal results of SR3 on real-world rainy dataset

Fig. 10. Comparative examples of raindrop removal effectiveness of SR3 on raindrop-degraded data (a) and real-world rainy BDD100K dataset (b). The blue and yellow circles draw attention to specific areas, magnifying the details to showcase the successful removal of raindrops that previously obscured critical features of the scene.

Specifically, Fig. 10(a) displays the transformation achieved on the synthetic raindrop-degraded dataset. The images on the left demonstrate the initial quality of the visuals marred by raindrops, while the images on the right reveal the enhanced clarity following SR3 processing. This stark contrast highlights the model's ability to effectively discern and eliminate raindrop distortions.

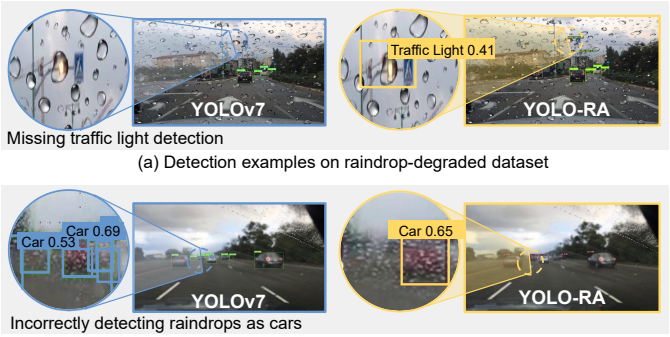
Fig. 10 (b) further validates the model's capabilities under real-world conditions, featuring images from the BDD100K dataset. The left images show the complex scenes as captured under natural rainy circumstances, where obtaining a clear rain-free image is inherently difficult due to the consistent presence of raindrops. In comparison, the right images depict a remarkable improvement in visibility after SR3 processing.

Overall, the before-and-after comparisons in Fig. 10 not only demonstrate the SR3 model's ability to purge the smallest raindrops but also its effectiveness in recovering important details in both simulated and real-world rainy visuals. The model proves to be invaluable in enhancing image clarity, thereby extending the potential for more accurate object detection and scene interpretation in adverse weather conditions.

2) Detection under Rainy Scenarios: In our study, we initially train the YOLOv7 and YOLO-RA models using microsphere raindrop-degraded images to detect objects across 80 classifications. Preliminary results from these experiments reveal that a significant detection error in rainy conditions is the false recognition of non-existent objects in vehicle driving scenarios. With this insight, we refine our focus to three critical types of object detection for SDVs: cars, persons, and traffic lights. For our purposes, the "car" category encompasses both trucks and buses, allowing for a more streamlined and targeted analysis in these prevalent SDV contexts.

Exemplar Visualization Results. Before performing a statistical analysis of the performance of our object detection model (YOLO-RA), we first compare it with the detection results of YOLOv7 based on both the raindrop-degraded dataset and the real-world rainy dataset.

Figure 11 illustrates a comparative analysis of object detection performance between our proposed model, YOLO-RA, and the baseline model, YOLOv7, under conditions of raindrop



(a) Detection examples on raindrop-degraded dataset

(b) Detection examples on real-world rainy dataset

Fig. 11. The effectiveness comparison of object detection between YOLOv7 and the proposed YOLO-RA on the raindrop-degraded dataset (a) and real-world rainy datasets (b).

interference. This comparison is critical to understanding the enhancements in detection accuracy that YOLO-RA offers over existing methods.

Figure 11(a) highlights a scenario where YOLO-RA accurately identifies a traffic light within a raindrop-degraded dataset, a detail that YOLOv7 fails to detect. This demonstrates YOLO-RA's superior ability to discern relevant objects amidst visual distortions caused by rain on the lens. Fig. 11(b) showcases detection examples from a real-world rainy dataset. Here, YOLOv7 incorrectly attributes the presence of cars to distortions created by raindrops on the camera lens, resulting in false positives. In contrast, YOLO-RA exhibits a more discerning judgment, correctly ignoring the raindrops and reducing the incidence of such false identifications.

Figure 11 effectively captures the enhanced detection capabilities of YOLO-RA, particularly in challenging weather conditions where rain can significantly degrade the quality of visual data. By mitigating the impact of raindrops, YOLO-RA shows promise in improving the reliability and accuracy of object detection systems in real-world applications.

Statistical Analysis. To rigorously assess the performance of our proposed YOLO-RA model, we undertake an ablation study [51], which involves the strategic removal of components from the model to observe the resultant effects on its efficacy.

As previously explained in Section V-B2, to bolster detection capabilities, particularly for small-scale vehicles, we implement a dual strategy based on YOLOv7 to obtain YOLO-RA: i) the integration of the Mosaic method and ii) the refinement of the ELAN module with the inclusion of the CBAM. Therefore, to ensure an equitable performance comparison, our ablation study involves the evaluation of four parts: YOLOv7 (baseline), YOLO-Mosaic (exclusively integrated the Mosaic technique into YOLOv7), YOLO-CBAM (exclusively implement CBAM in the ELAN module of YOLOv7), and our full-fledged YOLO-RA (integrate both Mosaic and CBAM).

Throughout the training iterations, the mean average precision (mAP) progression is monitored for YOLOv7, YOLO-Mosaic, YOLO-CBAM, and YOLO-RA. A peak in mAP is observed near the 50th epoch, succeeded by a decline indicative of early training instability. Hence, the selection of model weights based on this early performance would be premature. As training advances, mAP increases gradually, with YOLO-RA ultimately exhibiting superior performance, and YOLOv7, YOLO-Mosaic, and YOLO-CBAM trailing.

For testing, the Intersection over Union (IoU) threshold is set at 0.45, and the non-maximum suppression (NMS) [52] threshold is maintained at 0.6. NMS plays a crucial role in object detection by reducing redundant bounding boxes; when multiple bounding boxes are detected around the same object, NMS algorithmically selects the one that most accurately encapsulates the target, enhancing the precision of the detection.

Table I. Evaluation metrics of YOLOv7, YOLO-Mosaic, YOLO-CBAM, and YOLO-RA for three types of object detection.

Model	YOLOv7	Mosaic	CBAM	Precision	mAP	F1
YOLOv7	✓			0.78	0.82	0.81
YOLO-Mosaic	✓	✓		0.71	0.67	0.65
YOLO-CBAM	✓		✓	0.85	0.83	0.78
YOLO-RA (Ours)	✓	✓	✓	**0.89**	**0.85**	**0.82**

Table I presents the comparative results. YOLO-RA emerges as the leading model, securing the highest scores in precision and mAP. Notably, when contrasted with YOLOv7, YOLO-RA shows superior mAP and F1 scores (84.6% and 81.9%, respectively), evidencing the pronounced impact of the Mosaic augmentation. These findings affirm the effectiveness of YOLO-RA in object detection, substantiated by rigorous statistical analysis.

3) Combining SR3 with YOLO-RA: After proving the effectiveness of YOLO-RA methods for the object detection on raindrop-degraded images, we next explore the performance in object detection when combining SR3 for raindrop removal with our YOLO-RA model. This integrated pipeline, referred to as "SR3 + YOLO-RA," is tasked with initially cleansing the images of raindrop artifacts and subsequently deploying the refined images for object detection.

Table II. Evaluation metrics of YOLOv7 and "SR3 + YOLO-RA".

Model	Raindrop Removal	mAP	F1 score
YOLO-RA (Ours)		0.68	0.67
SR3 + YOLO-RA (Ours)	✓	**0.76**	**0.73**

Table II illustrates a clear improvement in performance metrics when employing the SR3 pre-processing step. Specifically, the integration of SR3 with YOLO-RA leads to an increase in the mAP and the F1 Score compared to using YOLO-RA alone. The mAP sees a boost from 0.68 to 0.76, while the F1 Score is elevated from 0.67 to 0.73. These increments translate to a substantial 12% improvement in mAP and a 10% enhancement in the F1 score, indicating that the preprocessing of images with SR3 not only aids in the removal of raindrops but also significantly enhance the subsequent object detection efficacy of YOLO-RA. This synergy between SR3's raindrop removal and YOLO-RA's detection capabilities underscores the potential of our composite pipeline in dealing with weather-degraded imagery.

4) Processing Speed: Accurate and fast processing model is the key to the success of SDVs. Table III provides a comparison of processing times and frame rates for several

object detection models and pipelines (YOLOv7, YOLO-RA, SR3 + YOLOv7, and SR3 + YOLO-RA), all of which are critical for SDVs systems that require both accuracy and speed. Each model is evaluated under consistent conditions, using a kernel size of 7×7 and an inference image size of 640×640. The kernel size refers to the dimensions of the convolutional filters within the layers, which are matrices utilized to perform convolutional operations on the input images for feature extraction.

Table III. Comparative processing times of different models using identical Kernel (7×7) and inference image size (640×640).

Models	T_{inf} (ms)	T_{NMS} (ms)	T_{total} (ms)	FPS
YOLOv7	8.1	3.4	11.5	86.96
YOLO-RA (Ours)	6.1	2.4	8.5	**117.65**
SR3 + YOLOv7	8.2	3.6	11.8	84.75
SR3 + YOLO-RA (Ours)	6.4	4.2	10.6	**94.34**

We define the total processing time (ms) as $T_{total} = T_{inf} + T_{NMS}$. The term T_{inf} (ms) specifies the time required for the model to process an input image and generate an output, exclusive of any additional post-processing duration. T_{NMS} denotes the time taken for the NMS operation, a critical post-processing phase where the model consolidates its predictions to guarantee singular object detection within multiple bounding box forecasts. T_{total} encompasses the entire processing interval from input to the finalized output after NMS. Besides, we consider another evaluation metric of frame per second (FPS), which signifies the quantity of images processed by the model per second. A higher FPS indicates a faster processing speed, which is desirable for SDVs.

The results in Table III reveal that the YOLO-RA model, without additional preprocessing, is exceptionally efficient, boasting the lowest total processing time at 8.5 ms and the highest frames per second (FPS) at 117.65. This performance suggests that YOLO-RA is highly suitable for real-time object detection, a necessity in SDV applications.

Furthermore, when SR3 is used in conjunction with YOLO-RA for pre-processing raindrop-degraded images, the composite model "SR3 + YOLO-RA" not only diminishes the total processing time T_{total} but also sustains a high FPS rate. With a T_{total} of 10.6 ms and 94.34 FPS, "SR3 + YOLO-RA" outperforms the "SR3 + YOLOv7" configuration, which underscores the efficiency and potential of our YOLO-RA model even when dealing with adverse weather conditions. In addition, to balance the image quality and time consumption, we find that 512×512 size is the best choice in our work.

VII. Observations and Discussions

In this section, we present and summarize our answers to **RQ1**, **RQ2**, and **RQ3**, discuss the key observations, and give our explanation for our experiment results and observed trends.

A. Observations and Discussions for RQ1

First, we present and discuss our answers to RQ1, drawing upon the outcomes of our experimental investigations. The supporting evidence is referenced in Section IV-C.

1) Observations and Answers: We make several interesting observations and answers:

⋆ *Microsphere* raindrops exhibit the most pronounced effect on camera-based object detection, with *spherical* and *flat* raindrops following in severity (Fig. 4).

⋆ Both *microsphere* and *spherical raindrops* similarly impede the detection of vehicles and traffic lights, albeit with diminishing severity in that order (Fig. 4(a)).

⋆ During the daytime, the influence of flat raindrops on detection results is marginal; however, at night, these raindrops considerably degrade the visibility of traffic lights, underscoring *the variable impact of raindrop types under different lighting conditions* (Fig. 4(a) and Fig. 4(b)).

2) Explanations and Discussions: **Microsphere impact**. Our analysis of raindrop impact, as visualized in Fig. 4, suggests that small and densely distributed raindrops can exert a considerable adverse effect on camera-based object detection systems. This is particularly pertinent for SDVs, which face the challenge of detecting smaller objects at greater distances, where the object's representation is constrained to a minimal number of pixels. YOLO typically struggles with small target detection due to the paucity of contextual information these targets present [53]. Raindrops further compound this issue by masking the real object data and potentially introducing misleading information to the detection model.

Raindrop impacts during daytime. The impact of raindrops on object detection varies significantly between daytime and nighttime conditions. Models from the YOLO series are typically trained on extensive public datasets such as COCO [54] and VOC [55], which comprise over 300,000 images and 1.5 million object instances, predominantly captured during daylight. Consequently, these models are well-equipped with prior knowledge to mitigate the effects of flat raindrops encountered during daytime scenarios. However, microsphere and spherical raindrops are underrepresented in these datasets, leading to a lack of critical information for the model to learn from.

The impacts of flat raindrops during nighttime: At night, the primary difficulty arises from the inherently low-light environment, which is a separate concern from any dataset-related issues. Flat raindrops can severely exacerbate visibility problems, often causing the model to erroneously identify a vehicle's rear lights as a red traffic light. This misinterpretation is frequently due to the facula effect created by light reflecting off the raindrops, which the model inadvertently captures.

B. Observations and Discussions for RQ2

Next, we present our responses to RQ2. The related supporting evidence is referenced in Section V-C.

1) Observations and Answers: We make several interesting observations and answers related to the density and diameter:

⋆ As to the raindrop density, low-density (scattered) rain has a more detrimental effect on detection efficacy compared to high-density scenarios (Fig. 6).

⋆ Regarding the raindrop size, the effect of small diameters is less significant than that of large diameters (Fig. 7).

⋆ The impact of raindrop diameter is notably less significant than that of raindrop density (Fig. 6(b) and Fig. 7(a)).

⋆ YOLO-RA consistently outperforms YOLOv7 across the board, demonstrating superior capability in accurately detecting five categories of transportation-related objects, irrespective of raindrop density or size (Fig. 6 and Fig. 7).

2) Explanations and Discussions: **Low-density impact:** Low-level features, such as edges and angles, are more likely to be obscured by isolated raindrops, offering challenges to extract key visual information. Conversely, while a high-density raindrop scenario packs more raindrops into each grid, it tends to create a uniformly cluttered visual field, which may not disrupt low-level feature detection to the same extent. This highlights the necessity for models to effectively handle a range of conditions, including the less intuitive scenarios where sparser raindrops can also impact detection abilities.

Large-diameter impact: Large size of raindrops, especially over a temporal sequence where these raindrops coalesce and grow, covering larger portions of the visual field. As the raindrops expand, they blur and obscure objects within the frame, consequently diminishing detection accuracy. Additionally, raindrops can act as lenses that refract and scatter incoming light, often resulting in glare, a phenomenon well-documented in studies [56].

Diameter versus Density: Raindrop diameter's impact is less pronounced than that of raindrop density. This is attributed to the fact that large-diameter raindrops may only partially obscure target objects, thereby affecting their recognition probability to a lesser extent. Conversely, low-density raindrops, being sparse and scattered, interrupt the model's ability to understand the context of neighboring regions. This disruption can mislead the detection algorithm into falsely identifying raindrops as objects, leading to a more substantial decrease in detection accuracy.

The effectiveness of YOLO-RA: The YOLO-RA model is specifically designed to overcome the limitations that YOLOv7 has with detecting small-sized objects. Leveraging an advanced attention mechanism, YOLO-RA allows for the effective extraction of intricate features and reducing the loss of low-level feature information. This strategic focus also aids in avoiding the misidentification of irrelevant areas or noise within the image. In tasks involving different raindrop diameters, the enhanced detection capability of YOLO-RA can be attributed to the model's ability to discern strong relational patterns even in the presence of microsphere raindrops. YOLO-RA is better equipped to address and resolve the challenges posed by adverse weather conditions.

C. Observations and Discussions for RQ3

Finally, we discuss our observations for RQ3. The corresponding supporting evidence is referenced in Section VI-C.

1) Observations and Answers: We make several interesting observations related to the algorithms:

⋆ SR3 demonstrates its effectiveness in eliminating raindrops from both raindrop-degraded images and real-world rainy dataset (Fig. 10).

⋆ YOLO-RA achieves superior performance in terms of mAP, also completing detection tasks more swiftly (Fig. 11).

⋆ "SR3 + YOLO-RA" boosts the mAP and F1 score, and speeds up the FPS as well (Table II and Table III).

2) Explanations and Discussions: **Effectiveness of SR3:** The success of SR3 depends on the similarity between microsphere raindrop-degraded data patterns to real-world raindrops, enabling the model to remove raindrops effectively. It describes the possibility of integration of image quality with multi-tasking. Besides, SR3 utilizes an iterative refinement approach by identifying raindrops as noise and integrating this aspect into the destruction phase. Consequently, it successfully removes raindrop interference.

Effectiveness of YOLO-RA: It effectively resolves the challenges associated with small-sized object detection and mitigates the errors commonly observed in YOLOv7, thus affirming the advanced capabilities of YOLO-RA in object detection amidst adverse weather conditions. Besides, the ELAN architecture effectively addresses the issue of gradient propagation length, allowing for the preservation of high inference speeds due to its well-designed computational blocks. Concurrently, the integration of CBAM in place of the original CBS modules enables more efficient feature extraction, contributing to the rapid processing capabilities of YOLO-RA.

VIII. Concluding Remarks

In this work, we first conduct a comprehensive qualitative and quantitative analysis to understand the influence of raindrops on camera-based object detection, a critical application of SDVs, under varying conditions of daytime and nighttime. Our exploration into the effects of four characteristic raindrop types across different diameters and densities provides valuable insights into their impact on diverse object detection. Significantly, we introduced the SR3 model for raindrop removal and evaluated its effectiveness across real-world rainy datasets and synthetically generated raindrop-degraded data. Furthering our contribution, we developed and proposed YOLO-RA, a novel algorithm designed for rapid and accurate mitigation of raindrop impacts. The integration of SR3 with YOLO-RA and subsequent ablation studies have effectively demonstrated the enhancements in object detection, as evidenced by improved metrics in mAP, F1 score, and FPS (or processing time).

Acknowledgement

This work is supported in part by the National Science Foundation (NSF) grant CNS-2348151 and Commonwealth Cyber Initiative grant HC-3Q24-048.

References

[1] S. Lu and W. Shi, "Vehicle computing: Vision and challenges," *Journal of Information and Intelligence*, vol. 1, no. 1, pp. 23–35, 2023.

[2] S. Lu and W. Shi, "Vehicle as a mobile computing platform: Opportunities and challenges," *IEEE Network*, 2023.

[3] L. Lowery, "OEMshift to OTArecall fixes predicted to occur by 2028 (online)," https://www.repairerdrivennews.com/2023/05/09/oem-shift-to-ota-recall-fixes-predicted-to-occur-by-2028/, accessed: 2023-05-09.

[4] L. VanHulle, "GM joins collaborative for vehicle software development (online)," https://www.autonews.com/automakers-suppliers/gm-joins-effort-develop-shared-software-auto-industry, accessed: 2023-04-27.

[5] E. Himes, "NHTSA up in the clouds: The formal recall process & over-the-air software updates," *Mich. Tech. L. Rev.*, vol. 28, p. 153, 2021.

[6] J. Henle, M. Gierl, H. Guissouma, F. Müller, G. B. Ramesh, and E. Sax, "Concept for an approval-focused over-the-air update development process," SAE Technical Paper, Tech. Rep., 2023.

[7] O. Haslam, "Mercedes-Benz rolls its flagship EV's infotainment system out to other cars via ota update (online)," https://www.pocket-lint.com/mercedes-benz-rolls-its-flagship-evs-infotainment-system-out-to-other-cars-via-ota-update/, accessed: 2023-02-10.

[8] S. Lu, N. Ammar, A. Ganlath, H. Wang, and W. Shi, "A comparison of end-to-end architectures for connected vehicles," in *2022 Fifth International Conference on Connected and Autonomous Driving (MetroCAD)*. IEEE, 2022, pp. 72–80.

[9] F. Bonomi and A. T. Drobot, "Infrastructure for digital twins: Data, communications, computing, and storage," in *The Digital Twin*. Springer, 2023, pp. 395–431.

[10] K. Arakadakis, P. Charalampidis, A. Makrogiannakis, and A. Fragkiadakis, "Firmware over-the-air programming techniques for IoT networks-a survey," *ACM Computing Surveys (CSUR)*, vol. 54, no. 9, pp. 1–36, 2021.

[11] J. Ohlsen, "The software-defined vehicle is overwhelming the automotive industry," *ATZelectronics worldwide*, vol. 17, no. 6, pp. 56–56, 2022.

[12] APTIV, "White paper: Smart vehicle architecture overview (online)," https://www.aptiv.com/en/insights/article/white-paper-smart-vehicle-architecture-overview, accessed: 2020-03-10.

[13] S. Lu, R. Zhong, and W. Shi, "Teleoperation technologies for enhancing connected and autonomous vehicles," in *2022 IEEE 19th International Conference on Mobile Ad Hoc and Smart Systems (MASS)*. IEEE, 2022, pp. 435–443.

[14] I. Dnistran, "Elon musk overruled tesla engineers who said removing radar would be problematic: Report (online)," https://insideevs.com/news/658439/elon-musk-overruled-tesla-autopilot-engineers-radar-removal/, accessed: 2023-05-22.

[15] K. Garg and S. K. Nayar, "Vision and rain," *International Journal of Computer Vision*, vol. 75, pp. 3–27, 2007.

[16] Y. Hamzeh and S. A. Rawashdeh, "A review of detection and removal of raindrops in automotive vision systems," *Journal of imaging*, vol. 7, no. 3, p. 52, 2021.

[17] T. Brophy, D. Mullins, A. Parsi, J. Horgan, E. Ward, P. Denny, C. Eising, B. Deegan, M. Glavin, and E. Jones, "A review of the impact of rain on camera-based perception in automated driving systems," *IEEE Access*, 2023.

[18] C. Zhang, Z. Huang, M. H. Ang, and D. Rus, "LiDAR degradation quantification for autonomous driving in rain," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 3458–3464.

[19] J. Hu, J. Li, Z. Hou, J. Jiang, C. Liu, L. Chu, Y. Huang, and Y. Zhang, "Potential auto-driving threat: Universal rain-removal attack," *Iscience*, vol. 26, no. 9, 2023.

[20] S. You, R. T. Tan, R. Kawakami, and K. Ikeuchi, "Adherent raindrop detection and removal in video," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1035–1042.

[21] D. He, X. Shang, and J. Luo, "Adherent mist and raindrop removal from a single image using attentive convolutional network," *Neurocomputing*, vol. 505, pp. 178–187, 2022.

[22] D. Eigen, D. Krishnan, and R. Fergus, "Restoring an image taken through a window covered with dirt or rain," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 633–640.

[23] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, "Attentive generative adversarial network for raindrop removal from a single image," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2482–2491.

[24] B. Yu, W. Hu, L. Xu, J. Tang, S. Liu, and Y. Zhu, "Building the computing system for autonomous micromobility vehicles: Design constraints and architectural optimizations," in *2020 53rd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*. IEEE, 2020, pp. 1067–1081.

[25] A. Papadoulis, M. Quddus, and M. Imprialou, "Evaluating the safety impact of connected and autonomous vehicles on motorways," *Accident Analysis & Prevention*, vol. 124, pp. 12–22, 2019.

[26] S. Li, I. B. Araujo, W. Ren, Z. Wang, E. K. Tokuda, R. H. Junior, R. Cesar-Junior, J. Zhang, X. Guo, and X. Cao, "Single image deraining: A comprehensive benchmark analysis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3838–3847.

[27] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3855–3863.

[28] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. S. Brown, "Single image rain streak decomposition using layer priors," *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 3874–3885, 2017.

[29] G. Wang, C. Sun, and A. Sowmya, "Erl-net: Entangled representation learning for single image de-raining," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 5644–5652.

[30] H. Kurihata, T. Takahashi, I. Ide, Y. Mekada, H. Murase, Y. Tamatsu, and T. Miyahara, "Rainy weather recognition from in-vehicle camera images for driver assistance," in *IEEE Proceedings. Intelligent Vehicles Symposium, 2005.* IEEE, 2005, pp. 205–210.

[31] K. Ito, K. Noro, and T. Aoki, "An adherent raindrop detection method using mser," in *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. IEEE, 2015, pp. 105–109.

[32] Q. Wu, W. Zhang, and B. V. Kumar, "Raindrop detection and removal using salient visual features," in *2012 19th IEEE International Conference on Image Processing*. IEEE, 2012, pp. 941–944.

[33] X. Yan and Y. R. Loke, "Raingan: Unsupervised raindrop removal via decomposition and composition," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*, January 2022, pp. 14–23.

[34] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[35] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *International conference on machine learning*. PMLR, 2015, pp. 2256–2265.

[36] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[37] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 4713–4726, 2022.

[38] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 7464–7475.

[39] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "BDD100K: A diverse driving dataset for heterogeneous multitask learning," *arXiv: 1805.04687*, 2018.

[40] L. Cao, X. Zheng, and L. Fang, "The semantic segmentation of standing tree images based on the yolo v7 deep learning algorithm," *Electronics*, vol. 12, no. 4, p. 929, 2023.

[41] G. Zhao, W. Ge, and Y. Yu, "Graphfpn: Graph feature pyramid network for object detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 2763–2772.

[42] W. Chen, X. Wang, B. Yan, J. Chen, T. Jiang, and J. Sun, "Gas plume target detection in multibeam water column image using deep residual aggregation structure and attention mechanism," *Remote Sensing*, vol. 15, no. 11, p. 2896, 2023.

[43] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "Repvgg: Making vgg-style convnets great again," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 13 733–13 742.

[44] Q. Xu, R. Lin, H. Yue, H. Huang, Y. Yang, and Z. Yao, "Research on small target detection in driving scenarios based on improved yolo network," *IEEE Access*, vol. 8, pp. 27 574–27 583, 2020.

[45] P. Kaur, B. S. Khehra, and E. B. S. Mavi, "Data augmentation for object detection: A review," in *2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE, 2021, pp. 537–543.
[46] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.
[47] D. Koller and N. Friedman, *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
[48] N. Chen, Y. Zhang, H. Zen, R. J. Weiss, M. Norouzi, and W. Chan, "Wavegrad: Estimating gradients for waveform generation," *arXiv preprint arXiv:2009.00713*, 2020.
[49] D. Poobathy and R. M. Chezian, "Edge detection operators: Peak signal to noise ratio based comparison," *IJ Image, Graphics and Signal Processing*, vol. 6, no. 10, pp. 55–61, 2014.
[50] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 658–666.
[51] B. Zheng, G. Jiang, W. Wang, K. Wang, and X. Mei, "Ablation experiment and threshold calculation of titanium alloy irradiated by ultra-fast pulse laser," *AIP Advances*, vol. 4, no. 3, 2014.
[52] A. Shapira, A. Zolfi, L. Demetrio, B. Biggio, and A. Shabtai, "Phantom sponges: Exploiting non-maximum suppression to attack deep object detectors," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 4571–4580.
[53] G. Chen, H. Wang, K. Chen, Z. Li, Z. Song, Y. Liu, W. Chen, and A. Knoll, "A survey of the four pillars for small object detection: Multiscale representation, contextual information, super-resolution, and region proposal," *IEEE Transactions on systems, man, and cybernetics: systems*, vol. 52, no. 2, pp. 936–953, 2020.
[54] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. Springer, 2014, pp. 740–755.
[55] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, pp. 303–338, 2010.
[56] S. You, R. T. Tan, R. Kawakami, Y. Mukaigawa, and K. Ikeuchi, "Adherent raindrop modeling, detection and removal in video," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 9, pp. 1721–1733, 2015.