

Response to reviewer 2

Manuscript ID: electronics-3627644

Manuscript Title: A Review of Unmanned Visual Target Detection in Adverse Weather Monitoring Scenarios

Journal: Electronics

Corresponding Authors: Yanfeng Lu

Dear Editors and Reviewer:

Thank you very much for your letter and for the reviewers' insightful comments on our manuscript entitled "A Review of Unmanned Visual Target Detection in Adverse Weather" (Manuscript ID: electronics-3627644). We greatly appreciate the time and effort you and the reviewers have devoted to evaluating our work. The comments provided are highly valuable and have been very helpful in guiding us to improve the quality and clarity of the paper. We have carefully addressed each comment and made the necessary revisions to the manuscript. Please find our detailed point-by-point responses below. All changes in the revised manuscript are marked using red font and tracked changes to facilitate your review.

Comment 1: *Several figures (e.g., Fig. 1, 3, 5) are overly text-heavy and do not clearly illustrate the structural differences between model types (e.g., CNN vs. Transformer). Please revise these diagrams to visually highlight architectural components or processing pipelines. For example, a comparative flowchart showing input-output behavior between CNN-based and Transformer-based recovery models would enhance clarity.*

Since your review aims to cover the full pipeline (from degradation recovery to detection and evaluation), I strongly recommend adding a single integrated diagram

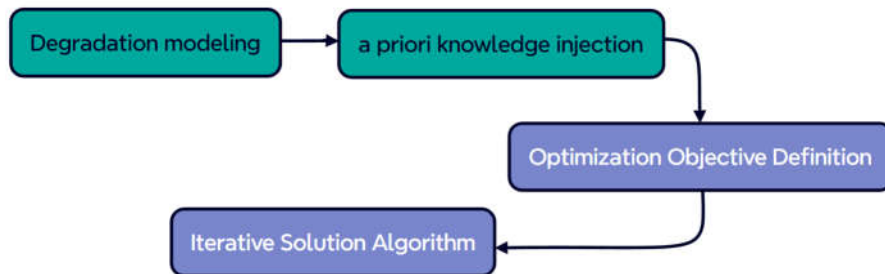
summarizing the overall framework.

Additionally, consider using visual comparison charts (e.g., bar graphs, radar plots) to complement or replace long text-based tables.

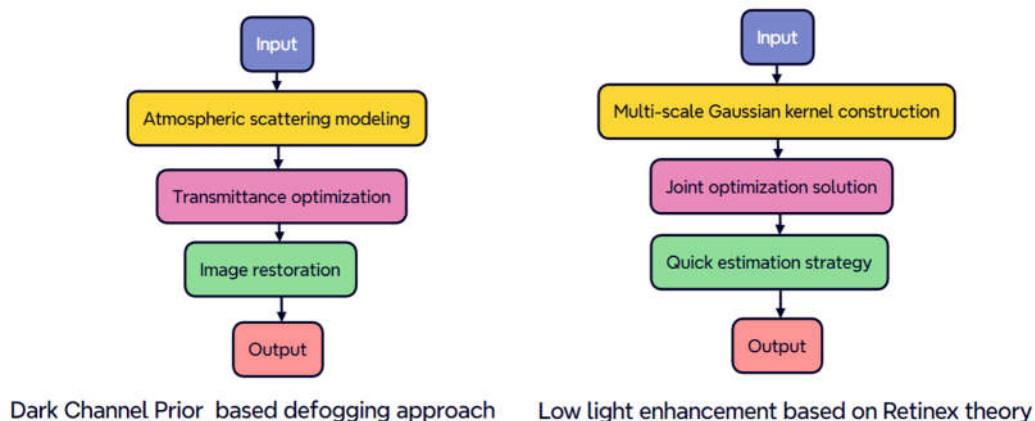
Response1: (1) Thank you for pointing out the issue with the clarity of our charts. In the previous text, some of the charts we created contained too much text and failed to show the structural differences between different model types clearly. We have taken your valuable feedback into account and revised these charts accordingly. We have replaced the previous charts with a block diagram that describes the input-output process of the image restoration and detection model robustness method. This revision has improved the clarity of the charts. We have made modifications to Figures 2, 4, 6, and 8. Below are the specific changes made:

✓In Figure 2, we have listed the input-output flowcharts for fog removal methods based on Dark Channel Priority (DCP) and low-light enhancement based on Retinex theory.

Before revision:

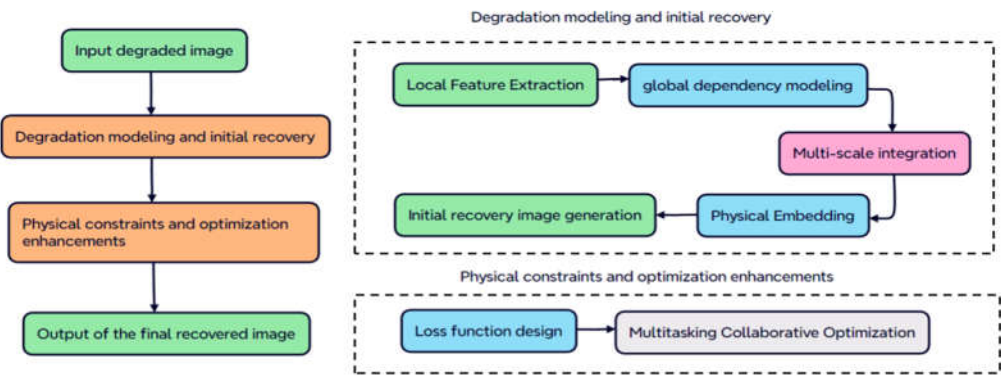


Revised:

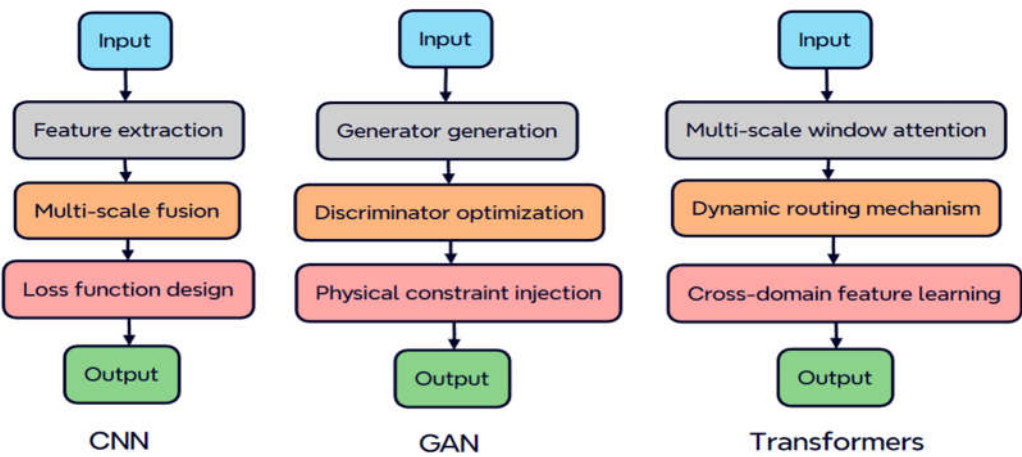


✓In Figure 4, we have listed the input-output flowcharts for image restoration methods based on Convolutional Neural Networks (CNN), Generative Adversarial Networks (GAN), and Transformers.

Before revision:

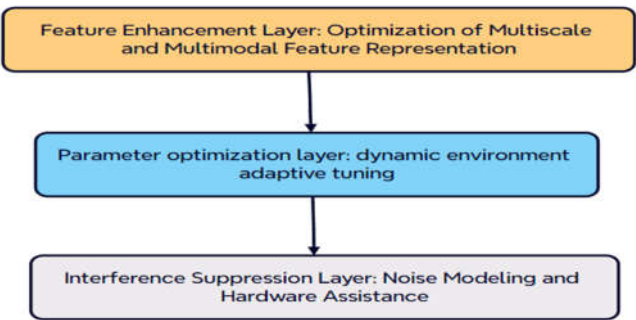


Revised:

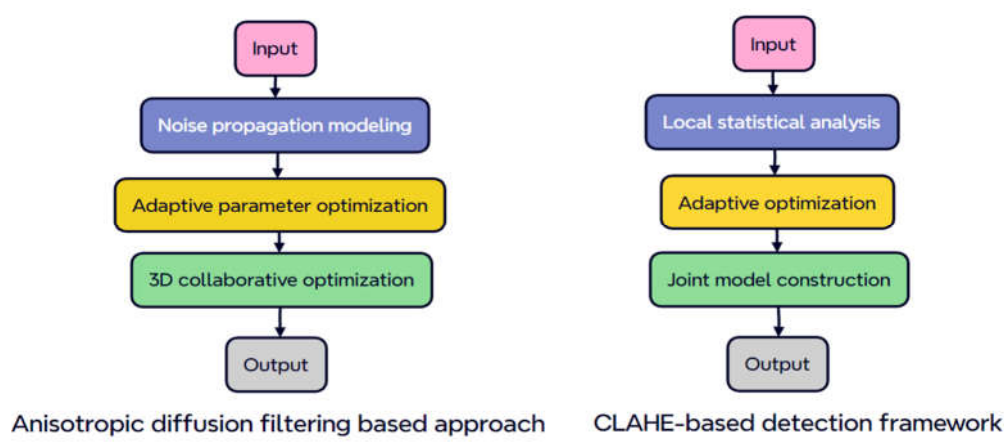


✓In Figure 6, we list the input-output flowchart for robustness improvement based on anisotropic diffusion filtering and CLAHE contrast enhancement methods.

Before revision:

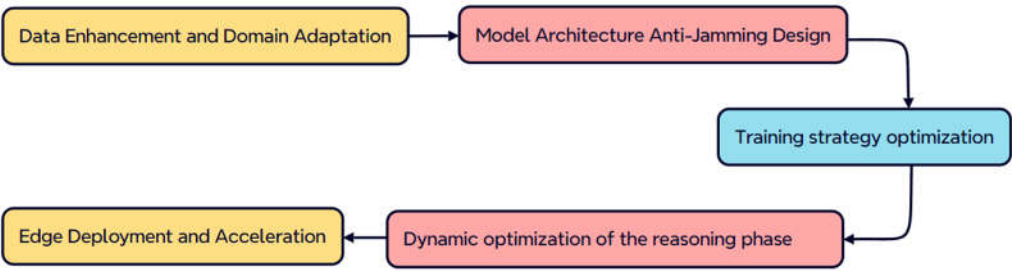


Revised:

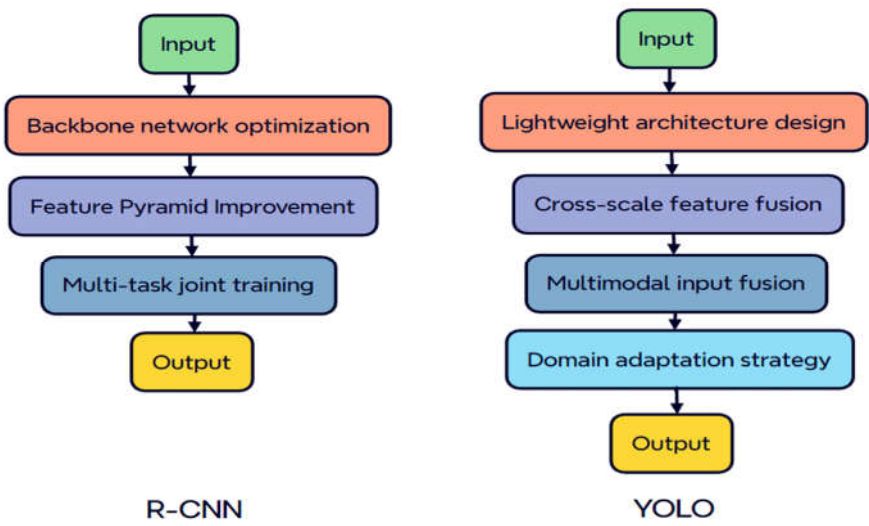


✓In Figure 8, we list the input-output flowcharts of robustness enhancement methods based on R-CNN and YOLO.

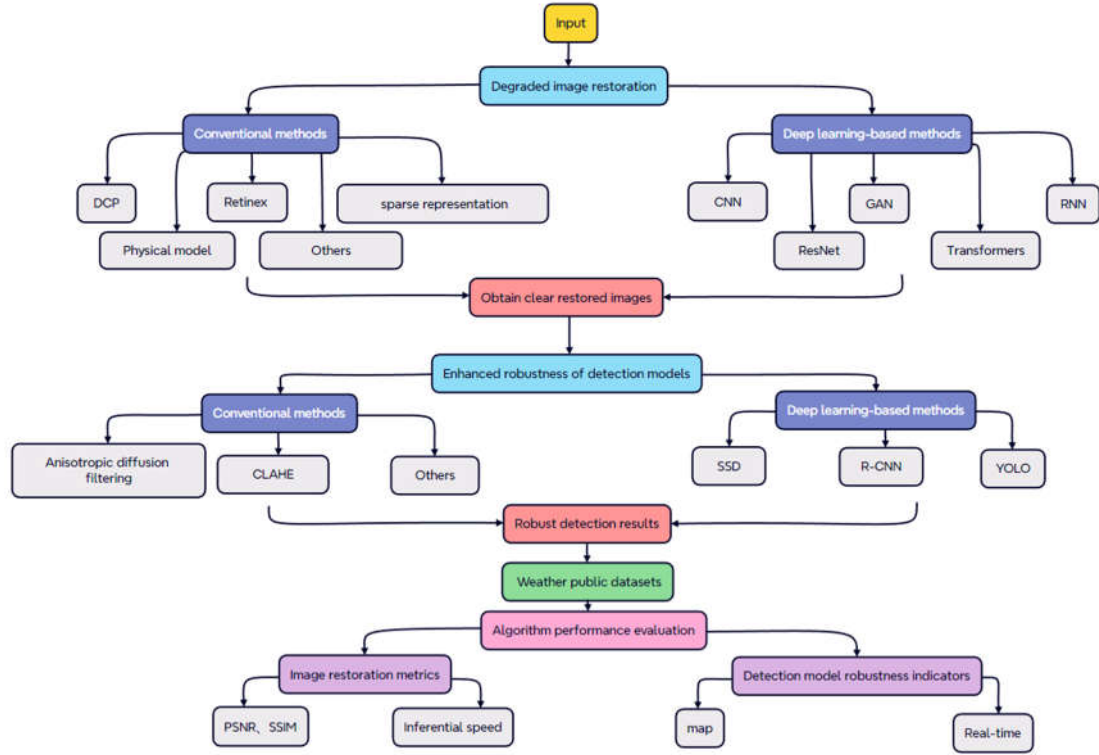
Before revision:



Revised:



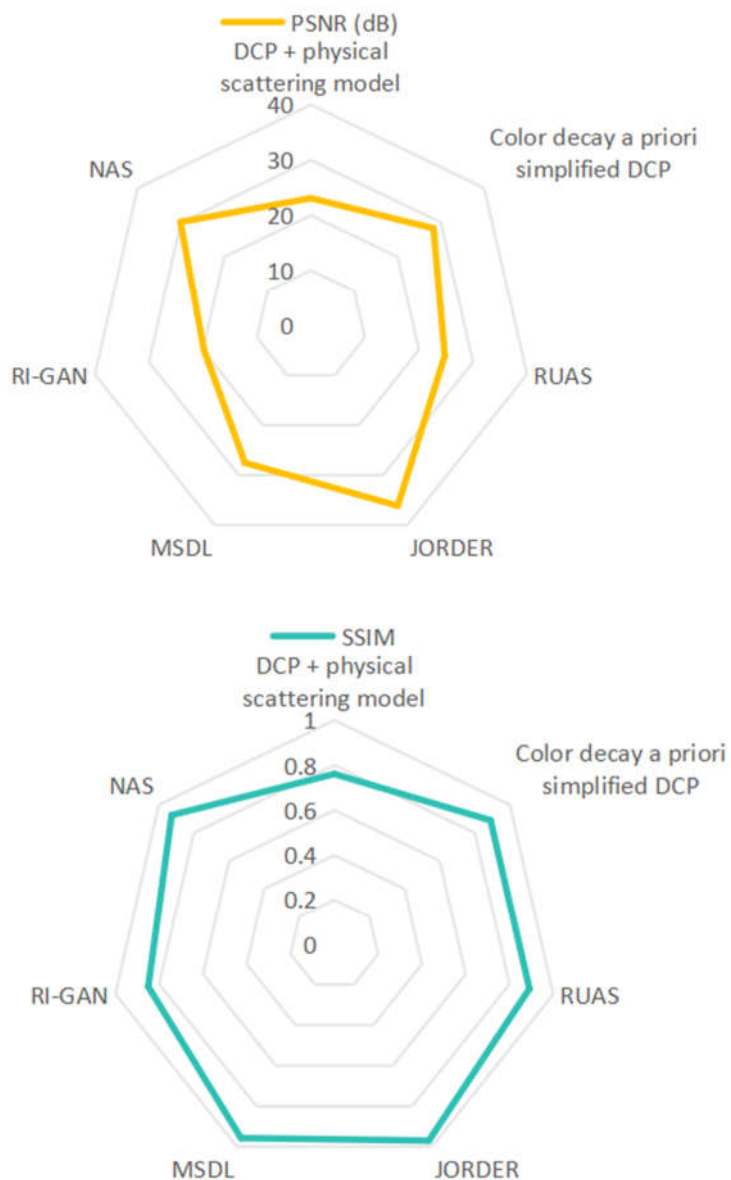
(2) Based on your suggestion, we have added an integrated summary diagram to the Introduction section. This diagram illustrates the framework of image restoration, robust detection, dataset, and algorithm evaluation. As shown below:

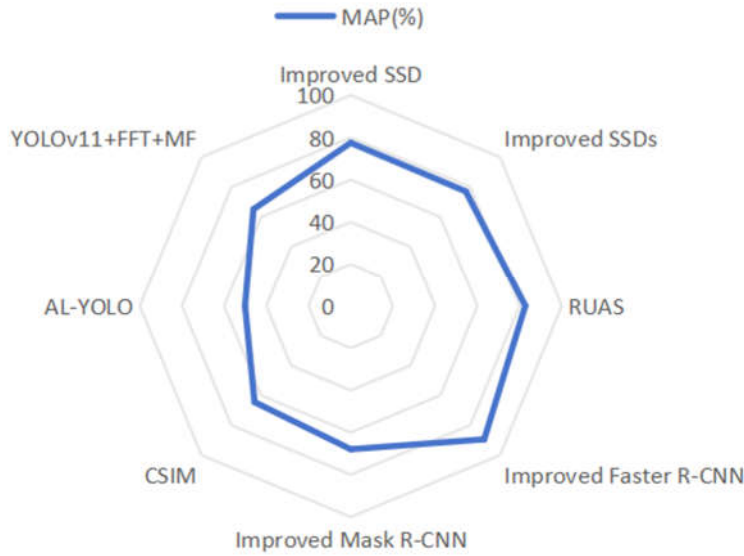


(3) Thank you very much for your valuable suggestions on the presentation of data in the paper. We have carefully considered your opinion on “using visual comparison charts (e.g., bar graphs, radar plots) to complement or replace long text-based tables.” Considering the disciplinary characteristics of the paper and the nature of the data, we would like to explain the rationale behind the current approach from the following perspectives. We kindly ask for your understanding.

After careful consideration, we believe that the long tables currently used in the paper have irreplaceable advantages in presenting multidimensional data. They can accurately display information such as algorithm models, datasets, and various metrics, and their advantages and disadvantages. This aligns with the publishing conventions for review papers in the fields of computer vision and autonomous driving. Additionally, some non-numerical information and metrics with significant magnitude differences in the tables are difficult to effectively present through visual charts. Forcing a replacement may lead to issues such as disrupted data relationships, distorted proportions, and reduced readability. For example, converting non-numerical descriptions like “driven by physical models” or the magnitude differences in inference speed into charts may result in information loss or reduced readability.

However, we also take into account your concern for data visualization. You suggested alternative optimization solutions such as adding conditional formatting to the table to highlight key information or adding small radar charts in the conclusion section for a macro-level comparison of core indicators. These approaches enhance readability without replacing the table while retaining the table to ensure data integrity and accuracy. We have supplemented the radar charts in sections 5.1.2 and 5.2.2. As shown in Figures 10, 11, and 12 below.





Comment 2:*Lack of Critical Analysis*

The paper largely summarizes existing works without critically analyzing or comparing them. Please provide clear evaluations on which methods perform best under specific conditions (e.g., fog vs. rain), and explain why — beyond listing general pros and cons.

Response2:Thank you for your valuable feedback on the article! Your comments regarding the article's shortcomings in critical analysis and methodological comparison are of great significance for further enhancing the depth of the content. We have incorporated crucial analysis and comparison into sections 2.1.6, 2.2.6, 3.1.4, and 3.2.4 of the article. We have conducted a comparative analysis of the methods proposed in each subsection and evaluated which methods are most effective under specific weather conditions (rain, snow, etc.). Below is the revised text.

✓**2.1.6.Comparison and Analysis**

In this section, traditional image restoration methods designed for adverse weather conditions are classified into several categories. These include fog removal methods based on dark channel priors, low-light enhancement methods grounded in Retinex theory, rain and snow removal strategies utilizing sparse representation, image restoration methods informed by physical modeling, as well as other traditional

restoration methods. Subsequently, a detailed review of the advancements in these methodologies is presented.

From a methodological perspective, the core contradiction inherent in traditional image restoration methods arises from the disparity between artificial prior assumptions and the diversity of real-world scenes. Dehazing methods that utilize dark channel priors (DCP) possess significant advantages due to their physical modelability and interpretability. Nevertheless, limitations such as model mismatches in high-brightness regions and issues related to computational complexity limit their application. Future research endeavors could explore the integration of these methods with deep learning approaches, leveraging neural networks to enhance transmittance estimation and mitigate domain shift challenges. Simultaneously, the introduction of lightweight structures can significantly enhance real-time performance. Weak light enhancement techniques based on Retinex theory separate illumination from reflection through the application of multi-scale Gaussian kernels, thereby achieving notable improvements in global illumination consistency. While effective, iterative solutions incur high computational costs and noise amplification issues associated with iterative solutions require urgent attention. The integration of neural architecture search with Retinex optimization to develop lightweight networks represents a significant avenue for advancement. Additionally, further improvements can be made to enhance the handling of residual noise under extreme noise conditions.

From a technical perspective, traditional methods exhibit two primary characteristics: “single-scene optimization” and “insufficient cross-domain generalization.” Rain and snow removal techniques that rely on sparse representation leverage sparse anomalies and low-rank assumptions to effectively separate degraded components. However, challenges such as sensitivity to local region selection and background blurring under extreme weather conditions must be addressed. These challenges can be overcome by improving the mask guidance mechanism and combining dynamic scene modeling to enhance adaptability to intricate rain and snow patterns. Physics-model-driven image restoration methods have gained significance due to their physical interpretability and reduced dependency

on labeled data. Despite these advantages, the application of these methods is constrained by parameter sensitivity and computational efficiency. The pre-training-fine-tuning paradigm that integrates physics priors with deep learning presents a viable approach. Future research should prioritize optimizing the flexibility of this two-stage training process while enhancing cross-domain generalization capabilities.

In foggy scenes, the Dark Channel Prior (DCP) and its improved variants demonstrate superior performance. DCP directly estimates transmittance based on an atmospheric scattering model, resulting in a PSNR improvement of 3 – 5 dB for synthetic fog images, alongside a structural similarity (SSIM) ranging from 0.76 – 0.89. This method is particularly effective in scenarios characterized by uniform fog concentration, where the statistical assumptions underlying the dark channel prior are well-aligned with the physical model, thereby facilitating accurate restoration of scene depth information. Nevertheless, certain limitations exist: in high-concentration fog or non-uniform fog scenes, deviations in the physical model can lead to a decrease in SSIM, and artifacts are likely to manifest in sky regions. The method that combines sparse representation with a physical model proves to be optimal for rainy weather conditions. Methods based on sparse representation-based consider rain streaks as sparse anomalies within images. Rain line separation is achieved through morphological component analysis (MCA) and self-learning dictionaries, achieving a PSNR of 36.11 dB and a SSIM of 0.97 on datasets such as Rain100L. This method demonstrates particular efficacy in suppressing rain lines in single-scene scenarios. Yet this approach has certain limitations: in extremely heavy rain scenarios, the density and dynamic nature of the rain lines can lead to challenges; specifically, neighborhood information reconstruction may cause background blurring, which could potentially decrease the PSNR. The physical model-driven and sparse prior fusion method demonstrates greater effectiveness in snowy weather scenarios. The physically-driven method simulates the scattering and attenuation of light caused by snow particles, achieving a PSNR ranging from 28 – 30 dB in synthesized snow images. Especially in uniformly snowing scenes, atmospheric light estimation and

transmittance optimization demonstrate efficacy in restoring scene structure. Nevertheless, these methods exhibit certain limitations; specifically, in scenarios involving dynamical snowing or snow particle-obscured scenes, the physical model struggles to capture real-time changes in scattering parameters, leading to noticeable texture distortions within the restored image.

Traditional methods perform exceptional performance in low-complexity scenarios characterized by single degradation types (e.g., uniform fog, moderate rain, static snow). Their physical interpretability and lightweight characteristics provide a solid foundation for practical deployment. However, when confronted with non-uniform degradation, multiple weather conditions (e.g., fog and rain mixed), or dynamic environments, the rigid assumptions associated with artificial prior knowledge and computational efficiency bottlenecks significantly cause their performance. Its future can be integrated with deep learning. The core value of the hybrid framework that combines "physical prior constraints + data-adaptive optimization" lies in its ability to address the inherent shortcomings of both traditional methods and deep learning approaches. On one hand, the physical model embeds prior knowledge of degradation processes, thereby mitigating the generalization blind spots of deep learning's black-box modeling in extreme scenarios. On the other hand, the data-driven nature of deep learning facilitates adaptive optimization concerning parameter sensitivity issues in physical models. This integration fundamentally addresses two critical challenges: first, while physical models often struggle to capture real-time changes in weather parameters, deep learning can learn degradation dynamics through time-series data; second, the shortcomings of physical prior knowledge in unknown scenarios can be mitigated through contrastive learning and domain adaptation techniques. Thus, the hybrid framework establishes a robust perception system characterized by both interpretability and environmental adaptability through the synergy of physical constraints and data adaptation. This represents a pivotal pathway for overcoming generalization limitations in complex scenarios.

✓2.2.6. Comparison and Analysis

In this section, methods for vehicle marking recognition based on deep learning are categorized into several approaches: image restoration methods utilizing convolutional neural networks (CNNs), image restoration methods employing generative adversarial networks (GANs), image restoration methods leveraging recurrent neural networks (RNNs), image restoration methods based on ResNet, and other image restoration methods grounded in Transformers. A detailed review of the research progress associated with each method is presented.

Unlike traditional methods, deep learning methods transcend the limitations imposed by the prior assumptions of conventional techniques through a data-driven paradigm. They have demonstrated significant advantages in the restoration of images affected by severe weather conditions. However, convolutional neural network (CNN)-based methods encounter challenges related to insufficient targeting in dynamic degradation modeling. For instance, HQS+CNN exhibits limited generalization capabilities when processing complex weatherr-induced degradations. DehazeNet exhibits limited capabilities in detail restoration due to its inadequate network depth. Nevertheless, its lightweight design and real-time processing advantages are noteworthy. In future developments, the introduction of dynamic weighting mechanisms and multi-scale feature fusion could significantly enhance adaptability to extreme weather conditions. Simultaneously, the incorporation of transfer learning can mitigate reliance on synthetic data. GAN-based methods such as Cycle-Dehaze and RI-GAN, represent significant advancements in cross-domain generalization and cascade error suppression. While innovative, they remain computationally expensive and heavily dependent on synthetic data. In future research, we can investigate the deep integration of unsupervised learning with physical constraints to leverage generative models for improving adaptability in real-world scenarios. At the same time, we can optimize the design of the discriminator to balance generation quality and efficiency. RNN-based methods, such as RVRT and NLRN, have progressed in dynamic modeling and multi-scale feature fusion, yet computational complexity and limitations in optical flow accuracy pose challenges for

practical applications. Future research may explore the integration of lightweight recurrent units and adaptive temporal constraints with neural radiance field (NeRF) technology to enhance temporal consistency in dynamic scenes. ResNet-based methods, such as NAS and RDN, have achieved multi-weather modeling and lightweight design, but a trade-off exists between the number of parameters and robustness under extreme weather conditions. In future work, neural architecture search (NAS) can be employed to optimize network structures, while integrating residual learning with dynamic weather classification may enhance adaptability across multiple scenarios. Methods based on Transformers, such as DehazeFormer and MWFormer, demonstrate strong performance in complex weather modeling due to their global self-attention mechanisms. However, many model parameters and high computational costs limit their deployment in vehicular applications. Future research may explore sparse attention mechanisms and dynamic routing optimization to address these limitations. Additionally, the integration of multimodal data -such as radar point clouds and images-can enhance robustness in extreme weather conditions. Furthermore, combining diffusion models with self-supervised learning has the potential to effectively address domain generalization bottlenecks, thereby facilitating a deeper integration of physical interpretability with data-driven approaches.

In foggy conditions, the fusion method of Transformer and GAN demonstrates superior performance. For example, DehazeFormer jointly optimizes transmittance and atmospheric light through multi-scale window attention, achieving a PSNR of 27.51 dB and a SSIM of 0.9576 on high-density fog datasets. Its global modeling capability effectively mitigates artifacts in the sky region. RestoreCUFormer leverages contrastive loss to learn domain-invariant features, resulting in a 12% improvement in generalization compared to traditional CNNs in cross-domain fog scenes, with PSNR fluctuations maintained below 1.5 dB. Combining CNN with sparse prior methods yields optimal results in rainy weather. The local sparse characteristics of rain lines are well-aligned with the local feature extraction capabilities of CNN, and the integration of sparse priors enhances robustness against complex rain patterns. For example, JORDER effectively separates positional

and intensity information through a binary rainline map, achieving a PSNR of 36.11 dB on the Rain100H dataset. The localization accuracy for dynamic rain lines is 18% higher compared to traditional sparse methods. The fusion of physical model-driven approaches with sparse priors proves to be more effective in snowy weather scenarios. The physically-driven method effectively simulates the scattering and attenuation of light caused by snow particles, achieving a PSNR ranging from 28 – 30 dB in synthesized snow images. Especially in scenes characterized by uniform snowfall, atmospheric light estimation and transmittance optimization demonstrate efficacy in restoring scene structure. Nevertheless, these methods exhibit certain limitations; specifically, in scenarios involving dynamical snowfall or snow particle-obscured scenes, the physical model encounters difficulties in accurately capturing real-time changes of scattering parameters. This limitation ultimately results in texture distortion within the restored images.

Deep learning methods have surpassed traditional methods in high-resolution image restoration for single-weather scenarios, such as uniform fog and moderate rain. Significant challenges persist, however, in extreme weather conditions, multimodal fusion, and edge deployment scenarios. Notably, , Transformer-based methods have successfully addressed previous obstacles associated with modeling complex weather degradation and have demonstrated unique advantages in this domain. Their self-attention mechanism effectively models long-range dependencies among pixels. They can address challenges through contrastive learning and domain-invariant feature extraction strategies. Additionally, the Transformer's global modeling capability enhances preservation of image structural details. Undoubtedly, deep learning-based degraded methods for degraded image restoration have become mainstream with the continuous development in deep learning technology. This method eliminates the need for manually designed features, thereby offering enhanced accuracy and robustness, particularly when applied to large datasets. Nevertheless, it is important to note that deep learning-based methods for degraded image restoration typically rely on high-quality datasets and achieve improved performance through increasing network weight parameters. This raises the challenge of balancing accuracy

and computational cost. Consequently, deep learning-based methods for degraded image restoration necessitate continuous improvement to meet real-world demands, particularly regarding dataset expansion and considerations related to computational expenses .

✓3.1.4. Comparison and Analysis

In this section, the traditional methods for vehicle identification are categorized into three categories: anisotropic diffusion filtering-based methods, CLAHE-based methods, and other traditional enhancement methods. A comprehensive review of the research advancements associated with each category is presented.

From a technical perspective, the core contradiction inherent in traditional technology arises from the disparity between the “rigidity of artificial a priori assumptions” and the “dynamic changes in environmental diversity.” Anisotropic diffusion filtering, such as OS-SART-ADF, mitigate noise diffusion via partial differential equations, mitigate achieving high edge retention under uniform fog conditions. Nevertheless, the computational complexity associated with three-dimensional iterative solutions restricts their adaptability to dynamic environments. In contrast, methods based on CLAHE, exemplified by CLAHE-GIF, enhance detection accuracy in low-light scenarios through local contrast enhancement.. Nonetheless, these methods are contingent upon fixed parameter configurations. The interplay between lighting and weather conditions within fog-rain mixed scenes contributes to improved target detection rates. Looking ahead, integrating reinforcement learning could facilitate adaptive parameter optimization tailored for dynamic environments. Combining this approach with innovative methodologies such as YOLO may pave the way for significant advancements in research development.

Traditional methods for enhancing the robustness of detection models exhibit limitations and present opportunities for innovation across various technical pathways. Future breakthroughs should prioritize the integration of physical modeling with data-driven approaches. Methods based on anisotropic diffusion filtering, such as the

OS-SART-ADF framework, leverage physical models to facilitate noise suppression while providing advantages in balancing edge retention and noise diffusion. However, these methods exhibit poor adaptability in extreme noise scenarios and lack sufficient automation in parameter optimization. Future approaches could incorporate deep learning techniques to enable dynamic adjustment of diffusion coefficients, alongside three-dimensional scene prior information to improve capabilities for suppressing complex noise.. Concurrently, it is essential to optimize iterative solution efficiency to accommodate real-time processing requirements. Contrast enhancement methods grounded in CLAHE, exemplified by the CLAHE-GIF model, improve adaptability in low-light environments through dynamic parameter adjustment. Nevertheless, they suffer from weak cross-scene generalization and are limited by dataset size. In future research endeavors, self-supervised learning can be integrated to extract lighting-invariant features from unlabeled datasets. Additionally, dynamic threshold optimization may be integrated to bolster detail recovery under extreme lighting conditions while expanding real-world scene datasets will enhance representation capabilities for complex noise. Other traditional enhancement methods, such as gradient-domain bootstrap filtering, have progressed in parameter robustness and halo suppression; however, they still lack generalization when faced with high-noise scenes. Future research could introduce attention mechanisms that adaptively allocate multi-scale weights. Combining physically driven noise models could enhance interference resistance within dynamic scenes.

Combining anisotropic diffusion filtering with a physical model yields optimal performance in foggy scenes. The LUT replacement framework achieves an inference speed of 14.3 FPS in such conditions through hardware acceleration and memory partitioning strategies, while maintaining a mAP of 72-75%. The physically driven noise suppression and edge retention mechanisms are consistent with the uniform degradation characteristics observed in foggy conditions. By optimizing the design of the diffusion coefficient, the improved Perona-Malik model realizes an exponential increase in edge retention to 92% under high-density fog scenarios. Compared to traditional methods, it effectively reduces halo artifacts by 30%. In rainy weather

scenarios, CLAHE combined with guided filtering methods performs better. CLAHE-GIF uses guided filtering to constrain the range of contrast enhancement, achieving an mAP of 68-70% in moderate rain events. This represents a notable improvement of 5-8% over the pure CLAHE method. It successfully balances brightness uniformization amidst mainline interference while preserving target features. The dynamic CLAHE fusion model, which utilizes a lightweight architecture, attains inference speeds of up to 20 FPS on edge devices. Furthermore, it improves the suppression rate of low-light-rainline mixed interference in rainy night scenes by 15%.

Traditional robust enhancement methods for detection models, while facing challenges such as a strong dependence on manual design and limited adaptability to dynamic environments, significantly outperform deep learning-based methods regarding computational cost and interpretability. In terms of accuracy, traditional methods can sometimes rival the performance of many deep learning methods. Certain robust enhancement methods may serve as optimization tools for achieving high accuracy in specific scenarios. However, given the complexity of outdoor environments and the presence of various random conditions in traffic, such as adverse weather, occlusions, and fluctuations in lighting, the effective utilization of traditional robustness enhancement methods to address these intricate scenarios and improve their generalization capabilities remains a significant challenge. In future research, it is imperative to overcome this bottleneck of combining traditional physical models with lightweight deep learning. For instance, we could incorporate the edge protection mechanism derived from anisotropic diffusion into the feature extraction layer of CNN. Alternatively, employing neural architecture search could facilitate the automatic optimization of the parameter configuration strategy for CLAHE, thereby contributing to the development of a robust, interpretable, and environmentally adaptable detection system.

✓ 3.2.4 Comparison and Analysis

In this section, we will categorize the methods for enhancing robustness through

deep learning into three main groups: SSD-based improvement methods, R-CNN-based improvement methods, and YOLO-series-based improvement methods. Additionally, we will provide a detailed review of the research progress of each method.

Deep learning-based detection models demonstrate significant advantages as well as notable bottlenecks that need to be addressed across various technical approaches. Future developments should prioritize overcoming core limitations and fostering cross-technology integration. Methods based on SSD, such as the improved SSD and M3-ECA-SSD, enhance small object detection capabilities through lightweight architectures and multi-scale feature fusion. Nevertheless, these methods face challenges related to insufficient hardware adaptability and limited generalization in extreme scenarios. Future improvements may involve the introduction of dynamic routing mechanisms to optimize the allocation of computational resources. Combining neural architecture search (NAS) could achieve hardware-aware network compression. Furthermore, expanding real-world adverse weather datasets has the potential to enhance adaptability in complex scenarios. Methods based on R-CNN, such as ECB-Mask R-CNN and improved Faster R-CNN, enhance accuracy through feature fusion and backbone networks, yet they still face challenges related to high dependency on computational resources and limited adaptability under extreme weather conditions. Future research could explore sparse attention mechanisms to mitigate computational complexity. The integration of multimodal data, such as radar and visual inputs, has the potential to enhance feature complementarity in adverse weather conditions. Additionally, self-supervised learning techniques can be utilized to improve cross-scenario generalization capabilities. Methods based on YOLO, such as Bi-STN-YOLO and CSIM, have made breakthroughs in dynamic modeling and real-time performance. Despite these prominent tradeoffs, the contradiction between data dependency and computational efficiency remains significant. Future work may involve combining diffusion models to generate diverse adverse weather data, adopt while employing knowledge distillation and model quantization to achieve lightweight deployment. At the same time, physically driven domain adaptation

mechanisms can be introduced to mitigate domain shifts between synthetic data and real-world scenarios. Additionally, the integration of spiking neural networks (SNNs) with YOLO, such as SUHD, demonstrates significant potential for achieving low power consumption. Further optimization is required to enhance adaptability to dynamic environments and improve hardware compatibility.

In foggy conditions, the YOLO series and multimodal fusion methods demonstrate superior performance. Through its lightweight architecture, AL-YOLO achieves a mean Average Precision (mAP) of 50.1% and a frame rate of 114.94 FPS on the Foggy-Cityscape dataset. Its dynamic feature calibration mechanism effectively mitigates light attenuation in foggy conditions. TransRAD integrates radar point cloud with image data. Leveraging the penetrative capabilities of radar to supplement visual input. This approach results in a 4-6% improvement in mAP compared to single-modal methods under foggy scenarios, thereby addressing the occlusion detection challenges inherent in pure visual methods. In rainy weather scenarios, SSD and domain-adaptive methods perform better. M3-ECA-SSD enhances detection capabilities through the implementation of channel attention and feature pyramids, effectively addressing the challenges associated with small object detection failures caused by mainline occlusion in the ADE20k indoor dataset. YOLOX combines a defogging module to facilitate end-to-end domain adaptation in Foggy Cityscapes, resulting in a 12% improvement in mean Average Precision (mAP) compared to traditional methods. Nevertheless, it is important to note that under heavy rain scenarios, generalization performance tends to decline due to inherent biases in synthetic data. In scenarios characterized by snowy weather, R-CNN and cross-domain learning methods perform better. Improvements to Faster R-CNN utilizing the RegNet backbone and neural architecture search achieved a mAP of 89.4% in complex background conditions typical of snowy weather. The multi-scale feature fusion mechanism effectively adapts to changes in snow particle size. Furthermore, the dual-stream architecture of R-YOLO—comprising QTNet domain transformation and FCNet feature localization—successfully addresses the domain shift bottleneck associated with snowy weather on the ACDC dataset, resulting in improved

cross-scenario generalization performance compared to traditional methods.

Robustness improvement methods based on deep learning present several advantages , including automatic feature learning, dynamic adaptability, and multimodal fusion capabilities in complex environmental perception. These attributes facilitate high-precision detection in domains such as autonomous driving. Nevertheless, it is essential not to overlook the associated drawbacks. Key limitations include data dependency, domain segmentation, significant computational overhead, and limited interpretability. Such limitations hinder the implementation efficiency of these methods in resource-constrained scenarios. To address real-world demands effectively, robust improvement methods for deep learning necessitate ongoing refinement. Future breakthroughs must maintain the benefits of end-to-end learning while addressing technical challenges through strategies such as multimodal fusion, physical knowledge embedding, and self-supervised learning.

Comment 3:*Unclear Contribution Compared to Prior Reviews*

Your review lacks a clear explanation of how it differs from existing review papers on similar topics. Please include a brief statement in both the abstract and early in the introduction explaining your review ' s unique scope or perspective.

Response 3:Thank you for pointing out the clarity of the article! Your suggestions are crucial for clarifying the research focus and enhancing the article's value. We sincerely appreciate your pointing out the shortcomings in the article's explanation of "differences from existing reviews." This is indeed an important area that we need to strengthen. Existing reviews of this type often focus on a single technical branch (such as image restoration or detection robustness). The core innovation of this paper lies in the first-ever construction of a "restoration-detection" collaborative framework, integrating physical modeling and deep learning paradigms from an interdisciplinary perspective. The following are specific supplementary ideas to highlight the uniqueness of the abstract and introduction:

(1) Following the first sentence of the abstract, we will explicitly state the unique contribution of this paper: Unlike existing reviews that concentrate on individual technical domains such as image restoration or detection robustness, this review introduces an innovative “restoration-detection” collaborative framework.

(2) In the first paragraph of the introduction, after outlining the practical background of autonomous driving, we clarify the unique scope and perspective of this review:

In contrast to prior studies, this paper offers a comprehensive examination of the entire process encompassing "degradation recovery, robustness detection, data benchmarks, and performance evaluation." This represents the first interdisciplinary review of autonomous driving visual target detection technology in adverse weather conditions.

Comment 4:*Language and Grammar*

The manuscript contains many long, awkward, or repetitive sentences, along with translation-like expressions that reduce readability.

Examples:

- “which show good interpretability and real-time performance...” (redundant phrasing)

- “chapter 1 provides...” → use “section ” instead of “chapter ” in academic papers.

(*)*A thorough proofreading by a native English speaker or a professional editing service is strongly recommended.*

Response 4:Thank you very much for your valuable feedback on the language expression and readability of the article! We are fully aware of the issues in the text, including verbose expressions, repetitive sentence structures, and academic standards. These issues have indeed affected the effectiveness of the content. Below is our specific response to your feedback and our improvement plan:

First, we have had the entire text reviewed and proofread by native English speakers within our institution. We have revised, edited, and removed redundant, awkward, and repetitive sentences and conducted a thorough grammar check. For example:

- (1) Redundant phrasing such as “which show good interpretability and real-time performance...” has been removed.
- (2) In “the inference speed is improved to video frame rate,” the word “the” was added before “video.”
- (3) In “In terms of improving the robustness of detection models, deep learning-based approaches typically encompass five key steps,” the phrase “In terms of improving” was changed to “To enhance.”

Additionally, we have revised similar translations that reduce readability. For example:

- (1) In “Chapter 1 provides...,” “chapter” has been changed to “section.”
- (2) “The mainstream deep learning-based car label recognition methods are” has been changed to “The predominant methods for car label recognition based on deep learning include.”

All other sentences proofreading and grammatical corrections are marked in red within the text.

Comment 5:Suggestions for Improvement

To strengthen this review paper, I suggest the following:

- Provide critical insights and trend analysis, rather than listing summaries.
- Clearly define your novel contribution compared to previous survey papers.
- Enhance visual clarity by improving diagrams and adding summary graphics.

Response 5:Thank you very much for your valuable suggestions on improving the quality of the paper! These suggestions provide essential guidance for enhancing the depth, innovation, and readability of the article. Below are specific responses and

improvement plans for the three suggestions you raised, combined with the content of the paper and academic norms:

(1) We sincerely appreciate your pointing out the shortcomings in the presentation of the content! Indeed, some sections of the paper previously lacked depth in their description of methods, with superficial analysis. In response to your suggestion to “Provide critical insights and trend analysis, rather than listing summaries,” we have made revisions. We have conducted critical comparisons and analyses of each method at the end of each subsection, presenting key insights. We have analyzed the breakthrough directions and future potential of methods such as Transformers mentioned in the paper. Additionally, in Section 6. Challenges and Prospects: we have included outlooks on emerging technologies such as diffusion models, self-supervised learning (SSL), and neural radiance fields (NeRF), outlining future trends.

(2) In response to your suggestion to “Clearly define your novel contribution compared to previous survey papers,” we have made revisions. We have added definitions of the new contributions in the abstract and the opening section of the introduction. In the abstract, we said: Unlike existing reviews that concentrate on individual technical domains such as image restoration or detection robustness, this review introduces an innovative “restoration-detection” collaborative framework. In the introduction, we added, “In contrast to prior studies, this paper offers a comprehensive examination of the entire process encompassing “degradation recovery, robustness detection, data benchmarks, and performance evaluation.” This represents the first interdisciplinary review of autonomous driving visual target detection technology in adverse weather conditions. ” These additions clarify the paper’ s unique contributions and scope.

(3) In response to your suggestion to “enhance visual clarity by improving diagrams and adding summary graphics,” we have made the following modifications. We have added input-output flowcharts for various models to compare and illustrate the

structural differences between model types clearly. We have supplemented the overall “restoration-detection” summary framework to clarify the paper’s overall structure. We have added visualization charts (radar charts) to more clearly demonstrate data comparison differences.

The improvements made to the above content can be referenced in the first four responses.