

Bridging the domain gap: A transferable dynamic routing enhancer for robust aerial detection under adverse weather

Yu Wan^a, Jie Li^{a,c,*}, Liupeng Lin^b, Zhaowen Lv^a, Zaiyan Zhang^a, Qiangqiang Yuan^a, Huanfeng Shen^b

^a School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China

^b School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China

^c Hubei LuoJia Laboratory, Wuhan 430079, China

ARTICLE INFO

Keywords:

Aerial object detection
Remote sensing
Adverse weather
Dynamic routing
Lightweight networks

ABSTRACT

Aerial remote sensing object detection often suffers from severe domain shift under adverse weather. Current methods primarily face two limitations. First, direct training on degraded images is computationally expensive and causes catastrophic forgetting on clear images. Second, using image restoration as a preprocessing step introduces task misalignment; these models optimize for human visual clarity and background reconstruction, wasting computation instead of preserving the essential semantic features required for detection. To bridge this gap, we propose a plug-and-play module called the Transferable Dynamic Routing Enhancer (TDRE). Designed specifically for the object detection task, TDRE uses a Dynamic Routing Mixture-of-Experts (DR-MoE) framework to handle different weather conditions without serious forgetting of clear images. Inside DR-MoE, parallel experts separate various atmospheric degradations, while a clear-sky perceptron lets clear images bypass the restoration process to prevent unnecessary interference. To further align the enhancement with the detection task, we introduce a synergistic optimization mechanism. This combines a multi-space restoration loss with a detection-oriented enhancement loss. This mechanism uses spatial masks based on detection labels to restrict gradient updates to target regions, ignoring visually redundant backgrounds. Evaluations on our CAD-ADD dataset show that our method is highly effective. With fewer than 3 K extra parameters and fast inference speeds, TDRE significantly improves the generalization of frozen detectors in complex scenarios, achieving a total improvement of 126.9 mAP50 points across 17 test sets. The source code is released at <https://github.com/KIKYOUWY/TDRE>.

1. Introduction

Aerial object detection plays an indispensable role in a wide range of real-world applications, such as intelligent transportation (Yang et al., 2023; Zhang et al., 2014), agricultural monitoring (Badgajar et al., 2024), and disaster relief (Zhao et al., 2024). However, robust detection in open-world environments remains a formidable challenge. Aerial images captured under adverse weather conditions—such as fog, dust storms, and low illumination—typically suffer from severe color distortion and reduced contrast. Consequently, pre-trained detection models experience a drastic performance drop, which is primarily driven by the severe domain shift between clear training datasets and degraded real-world inputs.

To address this issue, existing studies have primarily explored two paradigms (Wan et al., 2025), as shown in Fig. 1. The first paradigm

involves direct model training (Badgajar et al., 2024; Hu et al., 2024; Huang et al., 2021; Li et al., 2023; Liu et al., 2022; Wang et al., 2021; Xia et al., 2022), where researchers attempt to fully fine-tune pre-trained heavyweight detection models using degraded datasets. While this strategy can improve robustness under specific conditions, it poses significant limitations. Retraining large-scale detectors for every new weather scenario incurs prohibitive computational costs and lacks cross-scenario flexibility.

More critically, fine-tuning inevitably leads to catastrophic forgetting of clear images. As the model updates its weights to accommodate degraded features, its original detection capability on clear, normal-weather images is severely impaired. Alternatively, the second paradigm utilizes distributed methods (Li and Zhao, 2023; Long et al., 2014; Qin et al., 2020; Qiu et al., 2023; Zhang et al., 2026; Zhang et al., 2020), which deploy an independent image restoration network as a

* Corresponding author.

E-mail address: jili89@sgg.whu.edu.cn (J. Li).

<https://doi.org/10.1016/j.isprsjprs.2026.08.041>

Received 13 April 2026; Received in revised form 29 June 2026; Accepted 31 August 2026

0924-2716/© 2026 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

preprocessing step. While this plug-and-play approach avoids retraining the detector, existing restoration models fall into the trap of a human-visual bias, excessive focusing on the overall perception of the human. They consume massive computational resources to blindly pursue human-oriented evaluation metrics, prioritizing overall visual clarity and background reconstruction, while neglecting the high-frequency semantic features genuinely required for machine perception. This discrepancy leads to a fatal task inconsistency between image restoration and object detection: a smoothly restored image may lose the critical edge and texture information necessary for accurate target localization.

To break this deadlock, we propose a plug-and-play Transferable Dynamic Routing Enhancer (TDRE). Rather than acting as a conventional restorer to cater for human vision, TDRE operates purely as a detection-oriented visual filter, which is designed to specifically amplify the high-frequency semantic information required by frozen detectors rather than pursuing pixel-level clarity. Specifically, TDRE utilizes a Dynamic Routing Mixture-of-Experts (DR-MoE) framework to simultaneously process multiple weather conditions and relieve forgetting. Within DR-MoE, parallel experts adaptively disentangle complex atmospheric degradations. Concurrently, a novel clear-sky perceptron acts as a hard gate to distinguish degraded inputs and undegraded clear inputs. This intelligent routing mechanism effectively alleviates catastrophic forgetting.

To resolve task inconsistency, we introduce a synergistic optimization mechanism that couples a multi-space restoration loss with a detection-oriented enhancement loss, achieving both global and local alignment: while the multi-space loss corrects global color shifts across different domains, the detection-oriented loss uses spatial masks to focus gradient updates exclusively on target regions. Unlike conventional methods that reconstruct the entire image, our mechanism uses detection labels to guide these masks, ensuring that the model ignores task-

irrelevant backgrounds. This strategy prevents computational waste and forces the enhancer to specifically amplify the target-level features required for machine perception. In summary, the key contributions of this work are encapsulated as follows:

(1) **An Overall Architecture Handling Multiple Degradations and Combating Catastrophic Forgetting:** We propose the TDRE architecture and utilize the DR-MoE framework to safely navigate complex domain shifts. Its internal parallel experts effectively disentangle diverse adverse weather conditions, while the clear-sky perceptron seamlessly bypasses undegraded clear images. This routing strategy effectively alleviates degradation of the catastrophic forgetting on clear images.

(2) **A Synergistic Optimization Mechanism Resolving Task Inconsistency:** To break the human-visual bias in conventional distributed methods, we propose tightly coupling a multi-space restoration loss with a detection-oriented enhancement loss. By leveraging spatial masks guided by detection labels, TDRE proactively discards visually redundant background reconstruction and calculates gradients exclusively on target regions. This synergy ensures the enhancement process strictly and specifically serves downstream machine perception tasks.

(3) **A Comprehensive Cross-Scenario Benchmark and Extreme Lightweight Efficiency:** We construct the Cross-Application Degradation Aerial Detection Dataset (CAD-ADD), encompassing four distinct real-world domains and three typical atmospheric degradation conditions. Extensive evaluations demonstrate that TDRE, utilizing fewer than 3 K additional parameters with an extremely fast inference speed, significantly boosts the robust generalization of frozen detectors across diverse and complex scenarios.

As shown in Fig. 2, the radar chart illustrates the performance improvements achieved by the proposed TDRE over the baseline model across four distinct detection scenarios and a total of 17 test sets. With only 2930 additional parameters and an inference cost of 0.002 s, our

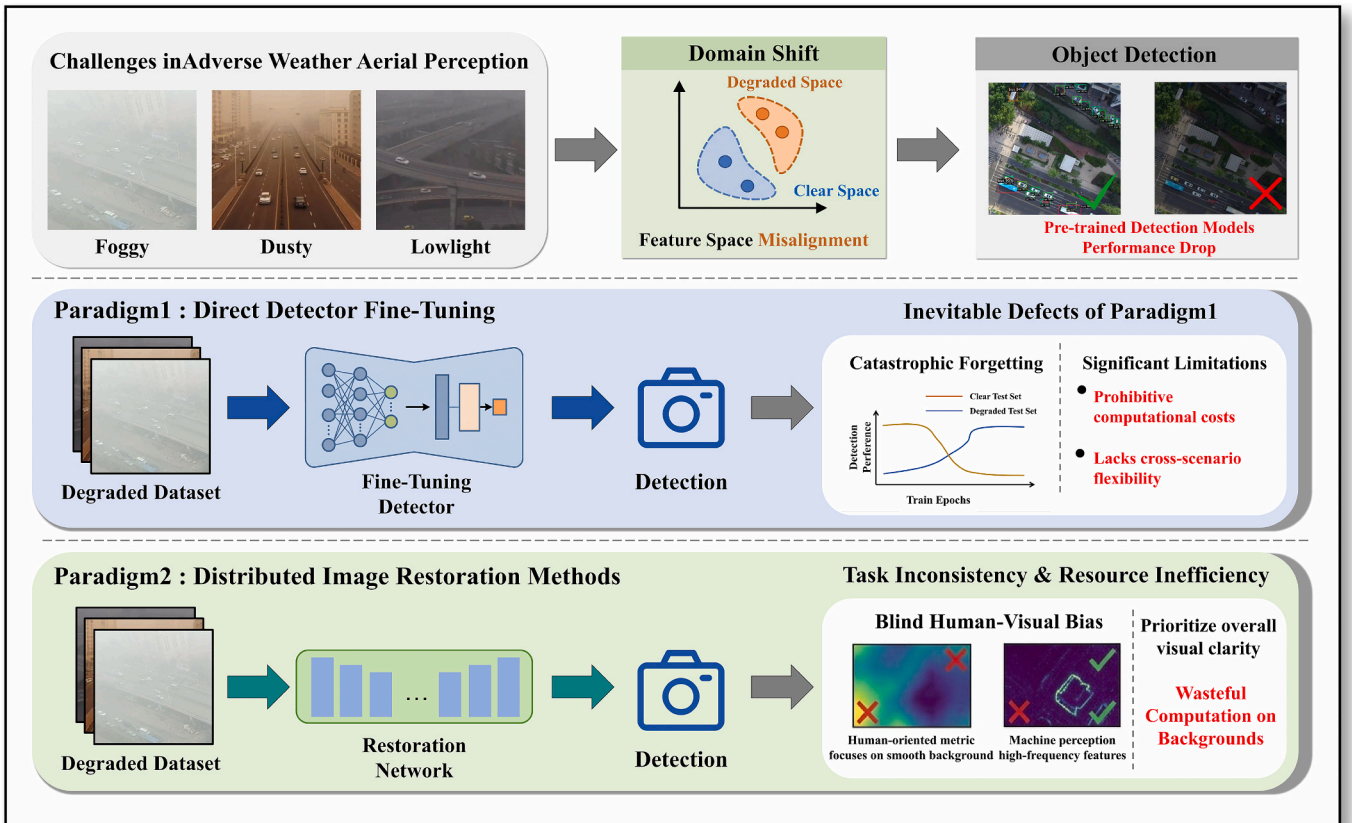


Fig. 1. Key Paradigms and Challenges in Adverse Weather Aerial Object Detection.

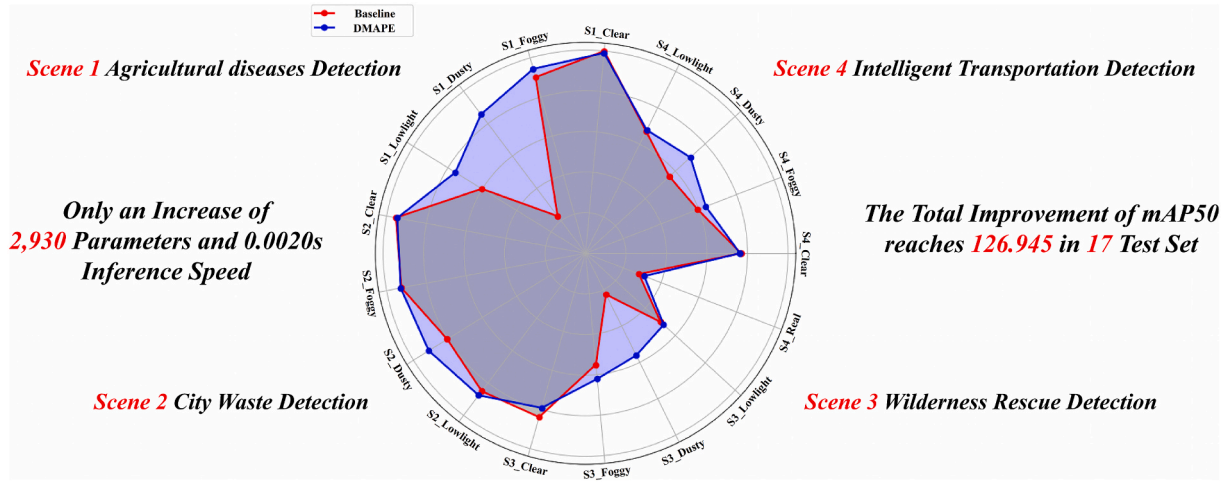


Fig. 2. The Radar Chart of Detection Improvement by TDRE under Different Detection Scenes.

method substantially improves detection performance under diverse weather conditions.

The remainder of this article is organized as follows. Section 2 reviews the related work, primarily focusing on aerial object detection and All-in-One image restoration algorithms. Section 3 details the proposed TDRE, elaborating on the Clear-Sky Perceptron, the DR-MoE framework, and the detection-driven refinement strategy. Section 4 presents the experimental setup, introduces the CAD-ADD dataset, and provides comprehensive evaluations, including comparisons with state-of-the-art methods and extensive ablation studies. Finally, Section 5 concludes this paper.

2. Related work

2.1. Object detection

In the nascent phase of research, object detection was predominantly formulated using hand-engineered low-level visual cues, such as color, texture, and shape descriptors. However, the discriminative capacity of these cues was severely limited by their fixed and domain-specific design paradigms. Seminal contributions from this era include the Viola-Jones cascade detector, the Histogram of Oriented Gradients (HOG) framework coupled with Support Vector Machines, and the Deformable Part Model (DPM)(Felzenszwalb et al., 2008). A fundamental limitation of these classical approaches lies in their restricted generalization capability, arising from the inability of rigid, hand-crafted features to adapt to diverse imaging conditions and object appearances.

The advent of deep convolutional neural networks precipitated a paradigmatic shift toward learning-based representations, ushering in modern object detectors that are broadly categorized according to their proposal-generation mechanism: one-stage and two-stage architectures (Zou et al., 2023). One-stage detectors unify localization and classification into a single, end-to-end regression problem, thereby enabling real-time inference with reduced architectural complexity. The YOLO (You Only Look Once)(Bochkovskiy et al., 2020; Redmon et al., 2016; Redmon and Farhadi, 2018; Redmon et al., 2017; Wang et al., 2023), initiated in 2015, has undergone continual methodological refinement, with YOLOv5 offering a scalable suite of models spanning multiple computational budgets to accommodate heterogeneous deployment scenarios. To mitigate the foreground-background class imbalance inherent in dense prediction tasks, RetinaNet(Lin et al., 2017) introduced the focal loss, a dynamically weighted objective that attenuates the influence of easily classified negative samples. Adaptive Training Sample Selection (ATSS) (Zhang et al., 2019) further elevates detection robustness by adaptively determining training exemplars. In the

specialized domain of aerial imagery, FFCA-YOLO(Zhang et al., 2024) incorporates foreground feature co-attention mechanisms to counteract weak feature representation and cluttered backgrounds characteristic of small object instances. SCENE-YOLO(Zhao et al., 2025) augments detection accuracy by integrating scene-level contextual supervision, thereby leveraging global semantic priors to refine local object hypotheses. Despite their computational efficiency and architectural simplicity, one-stage detectors often exhibit sub-optimal accuracy, particularly in scenes dominated by small or heavily occluded objects.

Conversely, two-stage detectors decompose the detection pipeline into a cascade of proposal generation followed by region-wise classification and regression, yielding superior localization precision at the expense of increased computational overhead. The Region-based Convolutional Neural Network(Girshick et al., 2014) pioneered this paradigm by employing Selective Search to extract approximately 2,000 category-agnostic region proposals, each subsequently classified via CNN-derived features. Faster R-CNN(Ren et al., 2017) streamlined this pipeline by integrating a Region Proposal Network (RPN) that directly generates proposals from convolutional feature maps, thereby amortizing computation across the shared backbone. Sparse R-CNN(Sun et al., 2021) further simplified the detection workflow by dispensing with hand-designed anchors and ROI-pooling operations, instead adopting a sparse set of learnable proposal embeddings. More recently, DiffusionDet(Chen et al., 2022) leverages the generative capacity of diffusion probabilistic models: departing from the static proposals of Sparse R-CNN, it initializes detection hypotheses as random bounding boxes and progressively refines them through a reverse diffusion process conditioned on image features. While two-stage detectors consistently achieve state-of-the-art accuracy, especially in complex scenes characterized by scale variation, occlusion, and dense object distributions. They incur substantial computational and temporal costs, rendering them less amenable to latency-critical applications.

Despite their impressive performance on standard datasets, these pre-trained detectors struggle in open-world environments. When confronted with adverse weather conditions, they experience a severe domain shift. To adapt to these degraded features, direct model training approaches attempt to fully fine-tune the detector. However, this strategy is inherently flawed. It not only incurs prohibitive retraining costs but also inevitably triggers catastrophic forgetting. As the detector updates its weights to accommodate hazy or low-light features, it frequently loses its original detection accuracy on clear images.

2.2. All-in-One image restoration

Currently, distributed methods frequently employ Image restoration

(IR) models as preprocessors for downstream perception tasks. Image restoration is dedicated to recovering high-quality images from observations degraded by a variety of factors, such as noise, blur, compression, and adverse weather conditions. Traditional IR methods have made notable progress by targeting individual degradation types. However, these specialized approaches often come at the cost of generalization, rendering them ill-equipped to handle the multifaceted distortions encountered in real-world applications. In response to this challenge, the All-in-One image restoration (AIR) paradigm has emerged, offering a unified framework that adeptly addresses multiple degradation types. This innovative approach enhances the convenience and versatility by adaptively learning degradation-specific features while simultaneously leveraging shared knowledge across diverse corruptions.

The objective of All-in-One Image Restoration models is to address a variety of isolated or mixed degradation tasks within a unified framework, obviating the need for task-specific retraining. Based on architectural designs, learning paradigms, and core innovations, AIR methods can be categorized into the following types. Some studies employ task-specific encoders and decoders, which involve assigning a distinct encoder-decoder pair for each type of degradation. However, this approach necessitates prior knowledge of the degradation type and exhibits limited flexibility when confronted with multiple or unknown degradations.

To this end, some research shares a single decoder but trains multiple heads for different degradation types. All-in-One Bad Weather Removal using Architectural Search (Li et al., 2020) is proposed to address the degradation of images caused by various adverse weather conditions (such as rain, fog, snow, and raindrops) using a single network. The core contributions of this method include the design of a generator with multiple task-specific encoders, each of which corresponds to a particular type of adverse weather degradation.

Meanwhile, some models utilize a unified encoder-decoder structure, which serves as a foundational design adaptable to various degradation types, such as, AirNet (Li et al., 2022), TANet (Wang et al., 2024), and AdaIR (Cui et al., 2024). AdaIR addresses the challenge of handling diverse image degradations—such as noise, blur, haze, and raindrops—within a single model. It achieves this by employing frequency mining and modulation to adaptively address the distinct impacts that different degradation types impose on image content across various frequency subbands. A prime exemplar of this paradigm is BIDE (Han et al., 2022). Models such as DINO-IR (Lin et al., 2023), Perceive-IR (Zhang et al., 2025), and DA-CLIP (Luo et al., 2023) leverage frozen vision-language models (e.g., CLIP (Radford et al., 2021), DINO (Caron et al., 2021)) to predict high-quality feature embeddings.

The mixture of expert (MoE) model has also been applied to All-in-One image restoration tasks, where inputs are routed to different experts via a gating mechanism. For instance, MEASNet (Yu et al., 2024) takes into account both pixel-level and global features, including low-frequency and high-frequency components, to select the appropriate expert for image restoration (IR). MoCE-IR (Zamfir et al., 2024) not only outperforms state-of-the-art methods, but also introduces flexible expert blocks with varying computational complexity and receptive fields (Complexity Experts). This design effectively affirms its efficiency and strong practical applicability.

Unfortunately, these restoration-centric approaches suffer from a severe human-visual bias. To achieve high scores on human-oriented metrics like PSNR, these models utilize massive architectures—often containing millions of parameters—to blindly reconstruct visually pleasing textures. This creates a critical task inconsistency. The heavily parameterized networks waste tremendous computational resources smoothing task-irrelevant backgrounds, often inadvertently discarding the high-frequency semantic edges essential for machine perception.

3. Proposed method

3.1. The architecture of The proposed method

The overall pipeline of the proposed Transferable Dynamic Routing Enhancer (TDRE) is constructed as a three-stage detection-oriented visual filter: degradation perception, condition-specific restoration, and detection-oriented enhancement as shown in Fig. 3. Unlike conventional distributed networks that blindly reconstruct entire images for human vision, TDRE is inserted as an ultra-lightweight plug-and-play module prior to a frozen downstream object detector, bridging the domain gap without modifying the detector's weights. The network parameter counts for the Clear-Sky Perceptron, DR-MoE, and Detection Enhancement stages are 273, 2406, and 251, respectively.

In the first stage (Degradation Perception), given an input aerial image I_{Input} , the Dynamic Routing Mixture-of-Experts (DR-MoE) framework evaluates the visual state through a novel clear-sky perceptron. This perceptron extracts preliminary features and outputs a clear aware score S_{ca} , which is formulated as follows:

$$F_m = f_{relu}(f_{BN}(f_{Conv1}(I_{Input}))) \quad (1)$$

$$S_{ca} = f_{Sigmoid}(f_{fc2}(f_{relu}(f_{fc1}(f_{pooling}(F_m)))))) \quad (2)$$

where I_{Input} and F_m are the input of images and the middle feature, $f_{Sigmoid}(\cdot)$ represents the Sigmoid activation, $f_{Conv1}(\cdot)$ represents the 3×3 convolution layer, $f_{BN}(\cdot)$ represents the batch normalization layer, $f_{relu}(\cdot)$ denotes the ReLU activation function, $f_{pooling}(\cdot)$ represents the pooling layer, $f_{fc1}(\cdot)$ and $f_{fc2}(\cdot)$ represent the fully connected layers. S_{ca} represents the clear aware score. In this work, the S_{ca} threshold is set at 0.5 to distinguish between clear and degraded images. If the image is classified as an undegraded clear image ($S_{ca} > 0.5$), it seamlessly bypasses the subsequent object detection model. S_{ca} threshold is set to 0.5 because it provides the best balance between avoiding unnecessary enhancement of clear images and activating TDRE for degraded inputs; this choice is further validated by the threshold sensitivity analysis in Table 6. This routing strategy guarantees less interference on clear images and substantially mitigates the risk of catastrophic forgetting.

In the second stage (Condition-Specific Restoration), if the image is identified as degraded ($S_{ca} < 0.5$), it is dynamically routed to parallel experts within the DR-MoE. These parallel experts explicitly disentangle and process diverse adverse weather conditions, such as fog and low illumination, yielding an intermediate restored image I_R .

In the third stage (Detection-Oriented Enhancement), the restored features undergo targeted synergistic optimization to explicitly resolve task inconsistency. The module processes I_R using an ultra-lightweight convolutional structure to generate the final enhanced result I_E :

$$I_E = f_{Conv3}(f_{relu}(f_{Conv2}(I_R))) \quad (3)$$

where $f_{Conv2}(\cdot)$ and $f_{Conv3}(\cdot)$ represent the 3×3 convolution layer. The enhanced results I_E along with the original clear images are fed into the subsequent detection model.

3.2. Dynamic routing Mixture-of-Experts for Condition-Specific restoration

While feeding all degraded conditions into a single, monolithic network might confer some basic restoration capabilities, the intrinsic statistical differences among diverse adverse weather scenarios are highly significant. Given our stringent constraint on minimizing model parameters, relying on a single shared feature space to handle multifaceted degradations is highly suboptimal. To address this challenge, we construct the Dynamic Routing Mixture-of-Experts (DR-MoE) framework as shown in Fig. 4. Instead of utilizing a generic architecture, this module adaptively routes degraded inputs to specialized parallel experts, explicitly disentangling diverse weather conditions in a single

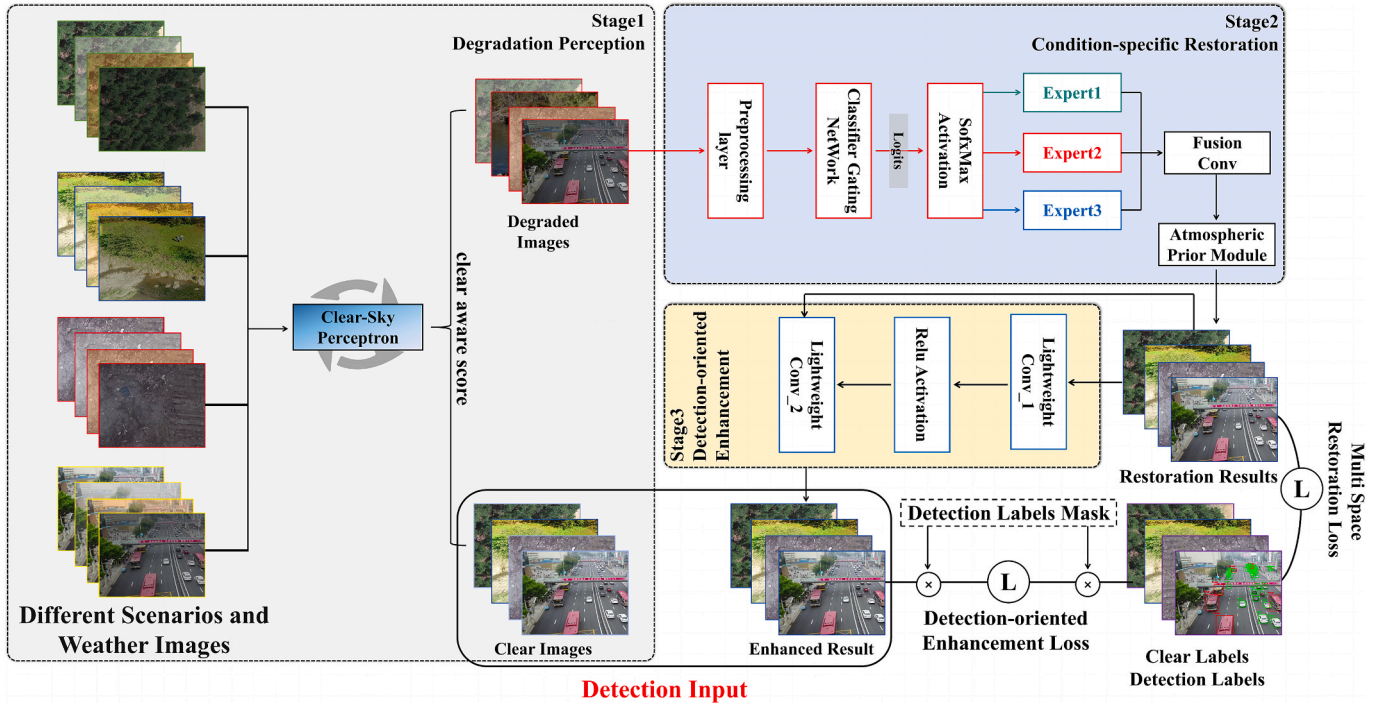


Fig. 3. The Overall Framework of TDRE.

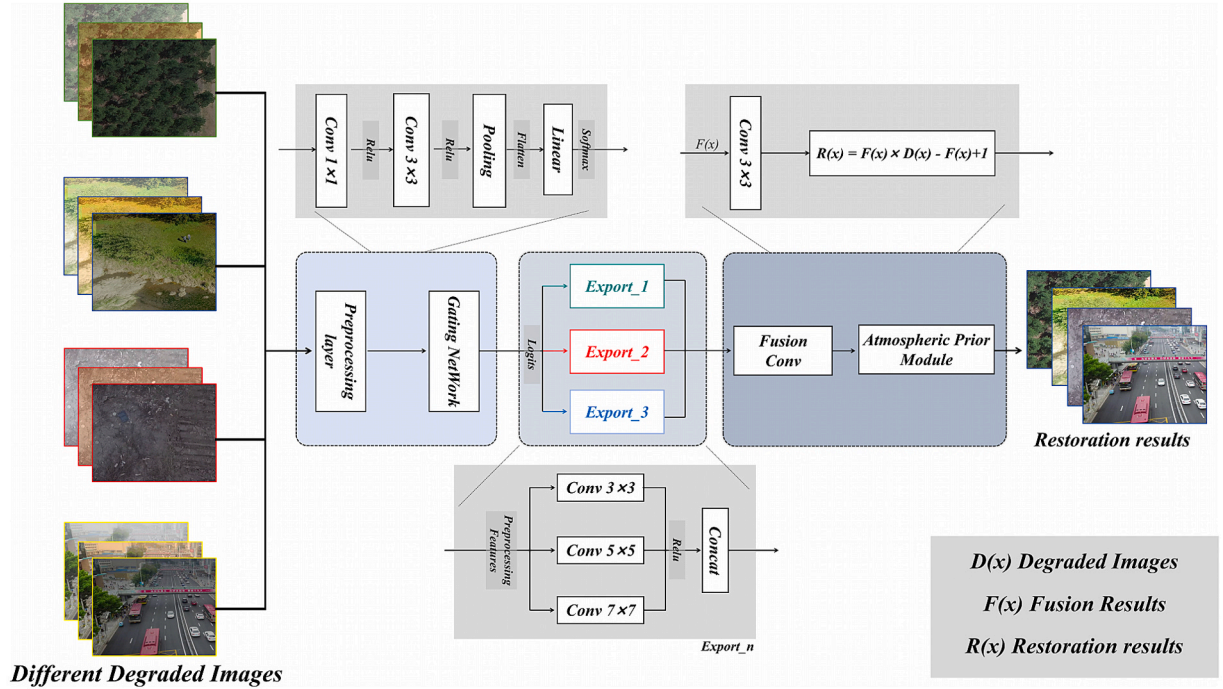


Fig. 4. The structure of the Dynamic Routing Mixture-of-Experts for Condition-Specific Restoration.

forward pass.

When an image is identified as degraded by the first-stage clear-sky perceptron, it is initially fed into a preprocessing layer to extract preliminary features. This initial feature extraction process can be mathematically represented as:

$$F_P = f_{\text{relu}}(f_{\text{Conv5}}(f_{\text{relu}}(f_{\text{Conv4}}(I_D)))) \quad (4)$$

where I_D and F_P represent the degraded images and the preprocessing features, respectively. Subsequently, the features extracted are fed into

the gating network to obtain the probability representation of degradation classification. This process can be represented as:

$$V_P = f_{\text{softmax}}(f_{\text{fc3}}(f_{\text{flatten}}(f_{\text{Pooling}}(F_P)))) \quad (5)$$

where $f_{\text{softmax}}(\cdot)$ represents the softmax activation. $f_{\text{flatten}}(\cdot)$ represents the flattening operation. V_P represents the probability representation.

Unlike conventional sparse routing designs, our framework activates all three parallel experts simultaneously during both the training and

inference phases (Top-3 routing). These three experts are specifically dedicated to resolving degradations induced by fog, dust, and low-light conditions. To effectively capture both local details and global contexts, each expert employs a multi-scale fusion strategy. The feature processing within the i -th expert is defined as:

$$F_E^i = f_{cat}(f_{Conv6}(f_{relu}(F_P)), f_{Conv7}(f_{relu}(F_P)), f_{Conv8}(f_{relu}(F_P))) \quad (6)$$

where $f_{Conv6}(\cdot)$, $f_{Conv7}(\cdot)$, $f_{Conv8}(\cdot)$ denote multi-scale convolutions with kernel sizes 3×3 , 5×5 , and 7×7 , respectively. F_E^i denote the output of the i -th expert. These three experts are respectively dedicated to restoring degradations caused by fog, dust, and low-light conditions present in the imagery. The expert restores the details and global features in the image more effectively through multi-scale fusion. The model inputs F_P into each of the three experts and then performs a weighted summation based on V_P . The fusion convolution process can be represented as:

$$F_{fusion} = f_{Conv9}\left(\sum_{i=1}^3 F_E^i \cdot V_P^i\right) \quad (7)$$

Finally, these fused features F_{fusion} are fed into an atmospheric prior module. To further enhance the robustness of the restoration process, we incorporate an atmospheric transfer equation. The intermediate restored image I_R is obtained through:

$$I_R = f_{relu}(F_{fusion} \cdot I_D - F_{fusion} + b) \quad (8)$$

where b is a bias term empirically set to 1 in this work. These restored results I_R are subsequently fed into the third-stage enhancement module.

Through this sophisticated yet parameter-efficient MoE design, the DR-MoE successfully disentangles complex atmospheric degradations. It provides a highly reliable intermediate restored image for the subsequent detection-oriented enhancement stage, ensuring robust perception augmentation while strictly adhering to our extreme lightweight constraints.

3.3. Synergistic Multi-Space detection-oriented enhancement loss mechanism

To effectively guide the model optimization and explicitly resolve the task inconsistency problem, we design a progressive, three-stage synergistic loss mechanism. Each stage is optimized independently with its specific objectives, while the parameters of the preceding stages remain entirely frozen.

Clear-Sky Perceptron Loss: In the first stage, the objective is to train the clear-sky perceptron to accurately distinguish between clear and degraded images. This ensures the hard gate can reliably bypass clear images. We employ the binary cross-entropy loss, formulated as follows:

$$L_{CSP} = \frac{1}{N} \sum_{i=1}^N [-y_i \cdot \ln(\sigma(\hat{y}_i)) - (1 - y_i) \cdot \ln(1 - \sigma(\hat{y}_i))] \quad (9)$$

where y_i is the i -th true label, $\hat{y}_i$ is the i -th predicted results. The parameters of the first stage are trained via L_{CSP} to enable the model to distinguish between clear and degraded images.

Multi-Space Restoration Loss: In the second stage, the DR-MoE framework is trained to disentangle specific weather conditions. The loss function at this stage comprises two distinct components. The first is a classification cross-entropy loss that optimizes the gating network to accurately route images to the corresponding experts. The second component is the multi-space restoration loss, which compares the intermediate restored images I_R with the clear ground-truth labels I_C . Because severe color distortion directly degrades semantic feature extraction, conventional RGB-only constraints are insufficient. Therefore, we explicitly introduce constraints within the HSV and LAB color spaces to form a comprehensive multi-space restoration objective:

$$L_{RGB} = f_{L_2}(I_R, I_C) \quad (10)$$

$$L_{HSV} = f_{L_2}(f_{HSV}(I_R), f_{HSV}(I_C)) \quad (11)$$

$$L_{LAB} = f_{L_2}(f_{LAB}(I_R), f_{LAB}(I_C)) \quad (12)$$

$$L_{MSR} = L_{RGB} + L_{HSV} + L_{LAB} \quad (13)$$

where $f_{L_2}(\cdot)$, $f_{HSV}(\cdot)$, $f_{LAB}(\cdot)$ represent Frobenius norm, the RGB to HSV transformation, and the RGB to LAB transformation, respectively. By constraining these heterogeneous color spaces simultaneously, the network recovers the authentic color distribution and structural details that are highly sensitive to machine perception.

Detection-Oriented Enhancement Loss: In the third stage, we implement the core mechanism designed to break the task inconsistency. Conventional restoration networks calculate the loss across the entire image, wasting tremendous computational resources on smoothing task-irrelevant backgrounds. To overcome this, we propose a detection-oriented enhancement loss that incorporates spatial masks guided by ground-truth detection bounding boxes. The formulation is expressed as:

$$L_{DE} = f_{L_2}(f_{Mask}(I_E), f_{Mask}(I_C)) \quad (14)$$

where L_{DE} represents the detection-oriented enhancement loss, $f_{Mask}(\cdot)$ represents the mask operation, only calculating the object part loss.

This masking strategy is the cornerstone of our synergistic optimization. By coupling the multi-space color constraints L_{MSR} with the spatial masks L_{DE} , the enhancement module proactively discards redundant background reconstruction. It forces the network to concentrate its limited capacity entirely on amplifying target-level discriminability—such as sharp boundaries and local contrast. Consequently, this task-driven synergy allows the TDRE to significantly boost downstream detection accuracy while maintaining an extremely compact architecture of fewer than 3 K additional parameters.

4. Experiments

4.1. Implementation details

To comprehensively evaluate the robustness and generalization capability of the proposed method under degraded conditions, we introduce the Cross-Application Degradation Aerial Detection Dataset (CAD-ADD) as shown in Fig. 5. This comprehensive benchmark encompasses four distinct real-world detection scenarios: agricultural disease, city waste, wilderness rescue, and intelligent transportation. Furthermore, the dataset captures diverse weather conditions, including clear, foggy, dusty, and low-light environments, alongside an additional real-world degraded testing set to validate practical reliability. The real-world degraded testing set has no geographic or temporal overlap with any training set, thereby preventing data leakage during evaluation. All scenarios in this dataset are constructed based on existing benchmark detection datasets, including PDT (Zhou et al., 2024) (agricultural detection), YQKJ-SARD (wilderness rescue), UAVwaste (Kraft et al., 2021) (city waste detection), and VisDrone (Zhu et al., 2022) (urban transport detection). The specific details of the dataset are presented in Table 1. In Table 1, $\times 4$ indicates that each training set contains four weather variants, namely Clear, Foggy, Dusty, and Lowlight. All experiments are conducted on an NVIDIA GeForce RTX 4080 GPU with 16 GB of RAM and a 13th Gen Intel(R) Core(TM) i7-13700KF CPU.

TDRE is optimized using the Adam optimizer. To ensure stable convergence, the learning rate follows a decay schedule, starting from an initial value of $2e-3$ and concluding at $2e-5$.

Following the three-stage optimization paradigm described in Section 3, we meticulously design the training batch construction to prevent data imbalance and mitigate catastrophic forgetting. In the first stage (training the clear-sky perceptron), the batch size is set to 12,

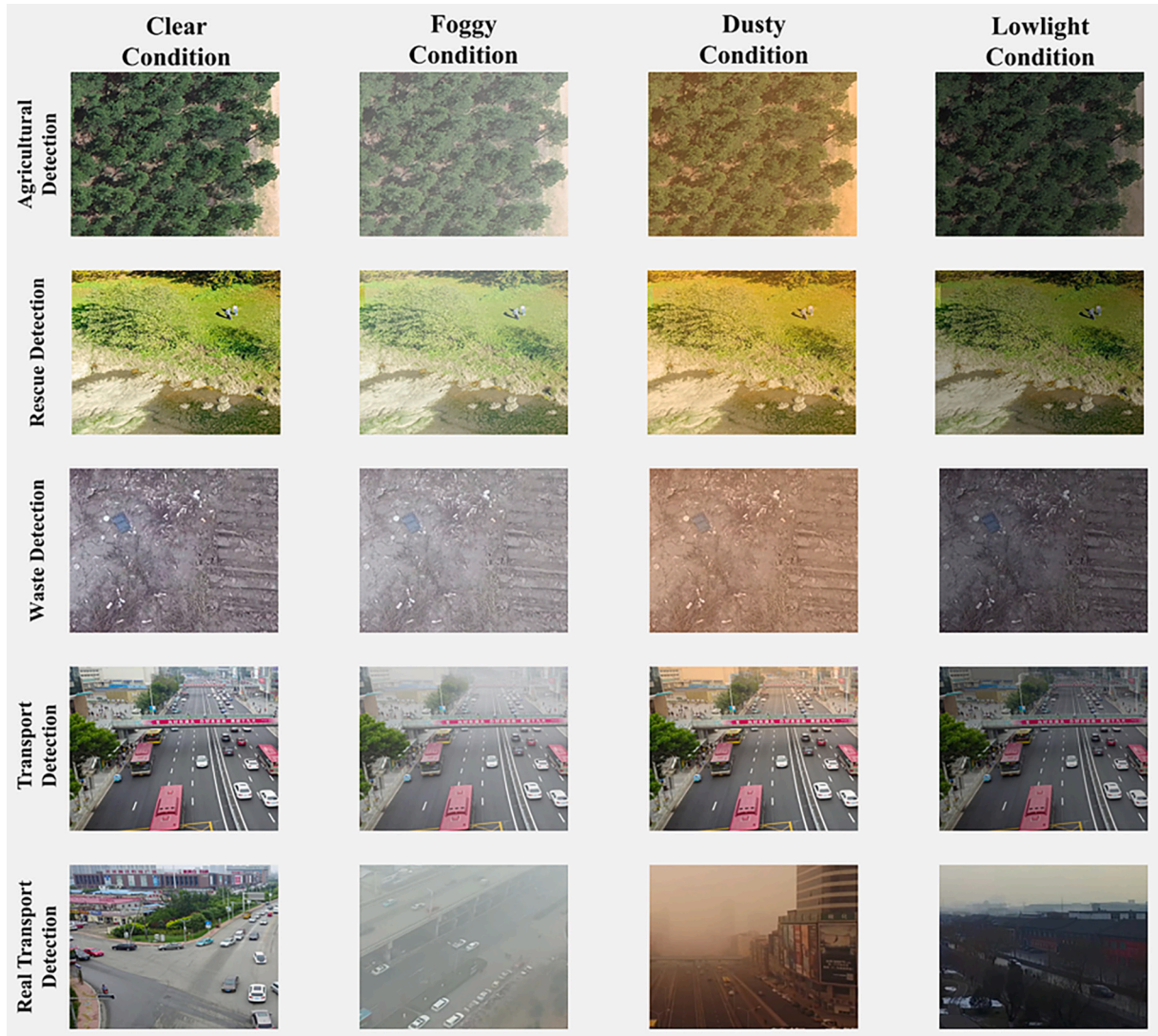


Fig. 5. The Cross-Application Degradation Aerial Detection Dataset.

Table 1
The Specific Details of CAD-ADD.

Datasets	Train Images Num	Test Images Num	Classes Num
Agricultural Detection	4536×4	567×4	1
Rescue Detection	12423×4	3550×4	5
Waste Detection	2368×4	1013×4	1
Transport Detection	3000×4	700×4	3
Real Transport Detection	N/A	2477	3

comprising a strict half-and-half split of clear and degraded images. In the second stage, the batch size is set to 16. To ensure unbiased expert routing and prevent the model from overfitting to a specific degradation, these 16 images are uniformly sampled from the four conditions (i.e., four images each from clear, foggy, dusty, and low-light sets) during every iteration. Given that clear images may also be affected by quality degradation factors, clear images are input alongside degraded images.

For the downstream perception evaluation, we employ DiffusionDet as the baseline detector. Crucially, this detector is pre-trained exclusively on clear images, and its parameters remain strictly frozen throughout all our experiments. To validate the superiority of our synergistic architecture, we select several state-of-the-art models for

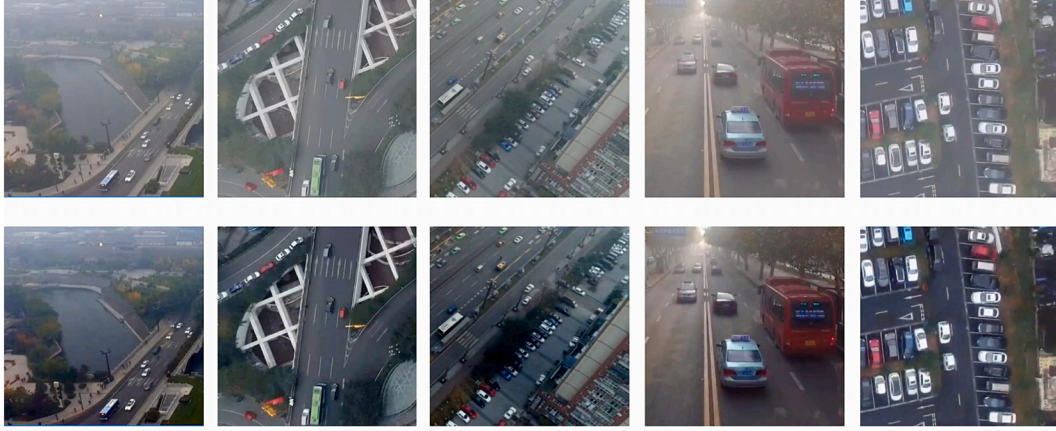
comprehensive comparison. These include specialized deep learning-based dehazing algorithms (AODNet, GridDehazeNet, FFANet, MB-TaylorFormer) and recent All-in-One restoration frameworks (IDR, MOCE-IR). To guarantee absolute fairness in comparative analysis, all baseline models are retrained from scratch on our dataset, utilizing the identical spectrum of clear and degraded images. During the evaluation phase, a strict plug-and-play protocol is enforced: the outputs from all restoration models are directly fed into the frozen detector without any task-specific fine-tuning.

4.2. Restoration performance in Real-World scenarios

To rigorously validate the practical utility of TDRE, we focus our visual analysis exclusively on aerial images captured under real-world adverse weather conditions. Conventional simulated datasets often fail to replicate the complex and unpredictable atmospheric degradations found in authentic environments. Therefore, real-world performance serves as the ultimate benchmark for a detection-oriented visual filter.

As shown in Fig. 6, we visualize the intermediate restoration results of TDRE under real-world degradations. Guided by DR-MoE, the network explicitly disentangles complex atmospheric distortions in a single forward pass. More importantly, our enhancement process is intrinsically targeted. This is achieved through the synergy of detection-oriented optimization and spatial masking mechanisms.

Foggy Scenarios



Dusty Scenarios



Lowlight Scenarios

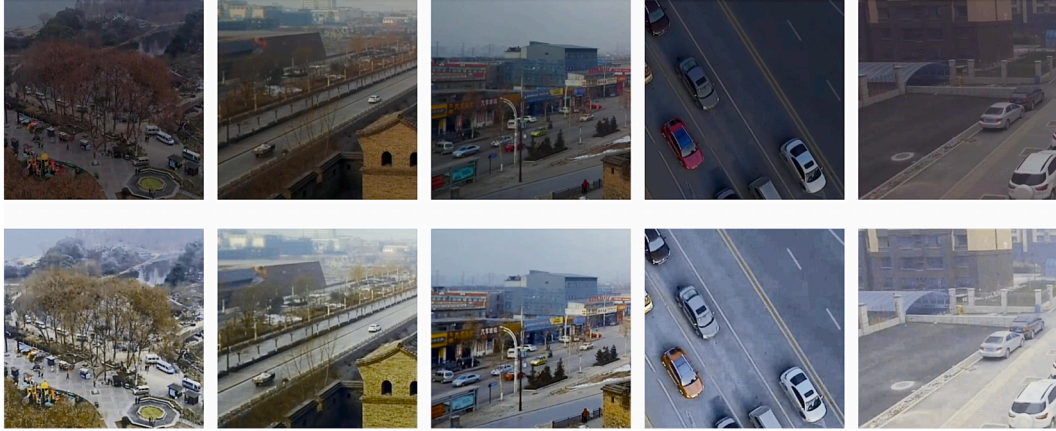


Fig. 6. Restoration Visualization Results of TDRE in real scenarios.

Visual results confirm that TDRE effectively restores global color fidelity in unpredictable environments. Furthermore, it specifically preserves and amplifies the structural details and high-frequency edges of potential targets. Rather than wasting limited computational resources on reconstructing task-irrelevant backgrounds, our method ensures that the enhanced features are strictly optimized for machine perception. This target-driven restoration strategy provides a highly reliable and robust visual foundation for downstream detection tasks. It perfectly meets the practical demands of complex aerial applications.

4.3. Detection performance comparison experiment

In the detection performance comparison experiment, we use the intelligent transportation dataset for this experiment. The intermediate restored results from each model are directly fed into the strictly frozen baseline detector, which was trained exclusively on clear images. This plug-and-play evaluation protocol rigorously verifies the actual perception gain provided by the enhancers. Detection accuracy is evaluated using the mAP50 metric, and the quantitative comparison results are summarized in Table 2. Bold indicates optimal performance, while

Table 2
Detection Performance Comparison Result.

	Clear	Foggy	Dusty	Lowlight	Real
Baseline	65.853	50.971	48.038	57.494	24.247
AODNet (CVPR2017)	61.312	48.864	55.016	40.108	20.875
GridDehazeNet (ICCV2019)	60.259	49.545	58.820	48.250	20.622
FFANet (AAAI2020)	64.675	50.368	53.296	58.970	24.100
MB-TaylorFormer (ICCV2023)	62.735	53.219	61.477	58.733	25.221
IDR (CVPR2023)	64.892	<u>54.435</u>	<u>61.550</u>	58.530	25.545
MOCE-IR (CVPR2025)	64.238	54.030	61.141	59.141	<u>25.561</u>
TDRE	65.379	55.290	61.879	<u>58.985</u>	26.668

underlined denotes second-best performance.

As demonstrated in the experimental results, the proposed TDRE comprehensively and significantly boosts the detection performance of the original baseline across various adverse weather conditions, while utilizing an extremely minimal number of additional parameters. Crucially, benefiting from the bypass mechanism of the clear-sky perception, TDRE maintains nearly identical detection accuracy on the clear image test set. In contrast, other distributed networks suffer from noticeable performance degradation on clear images. This decline directly exposes their inherent human-visual bias, where task-irrelevant background smoothing disrupts original semantic features, leading to task inconsistency. Overall, TDRE achieves the best performance on four out of the five test sets and secures a highly competitive second-best result on the low-light dataset. This strongly validates the superiority

of our detection-oriented visual filter.

The qualitative detection results are visualized in Fig. 7 and Fig. 8. As shown in the figure, under both simulated and real-world degraded scenarios, other traditional restoration methods can achieve superior visual effects for human eyes, while these methods often blur high-frequency semantic edges due to aggressive over-smoothing. This functional misalignment inevitably leads to severe missed detections and localization errors for the downstream detector. Conversely, TDRE successfully preserves and amplifies target-level feature discriminability in complex environments. Consequently, it empowers the frozen baseline detector to reliably penetrate severe domain shifts, achieving precise object localization and recognition under highly challenging conditions.

4.4. Generalization experiments in different scenarios

To thoroughly validate the cross-scenario generalization of the proposed enhancer, we evaluated it across four distinct domains. These include agricultural disease detection, city waste detection, wilderness rescue, and intelligent transportation. In these experiments, we tested two versions of our model. TDRE-A was trained on the complete dataset. In contrast, TDRE-T was trained exclusively on the transportation dataset. We designed this specific setup to see how well the model transfers to completely unseen scenarios. The quantitative results are summarized in Table 3, where D- indicates the basic baseline detector. The results clearly show that TDRE-T still achieves a highly significant performance boost when transferred to unfamiliar fields under degraded conditions. Meanwhile, TDRE-A further maximizes this robust generalization, comprehensively improving detection accuracy across all tested scenarios. Qualitative results are illustrated in Fig. 9.

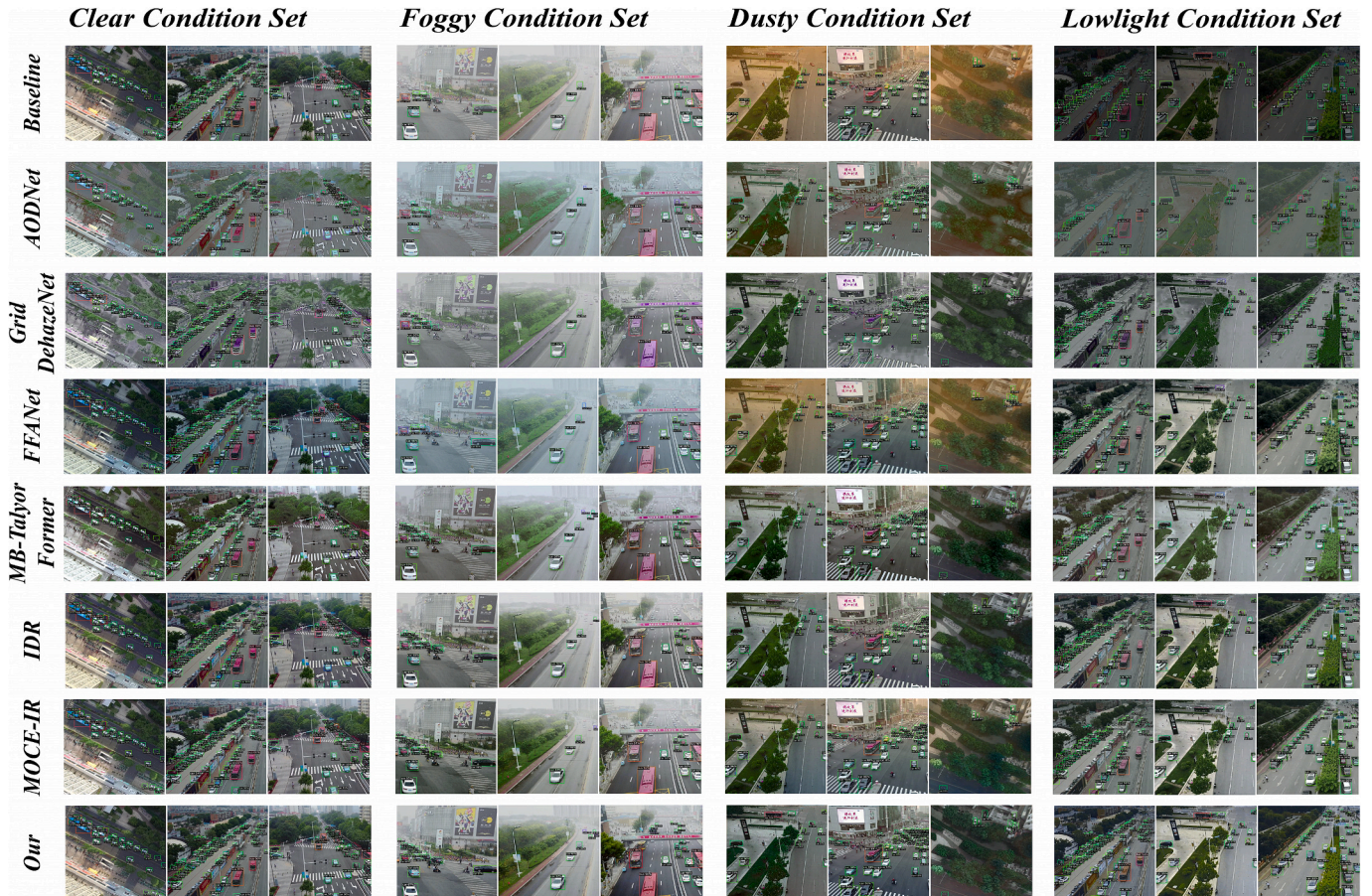


Fig. 7. Detection Visualization Result in the Simulated Test Set.

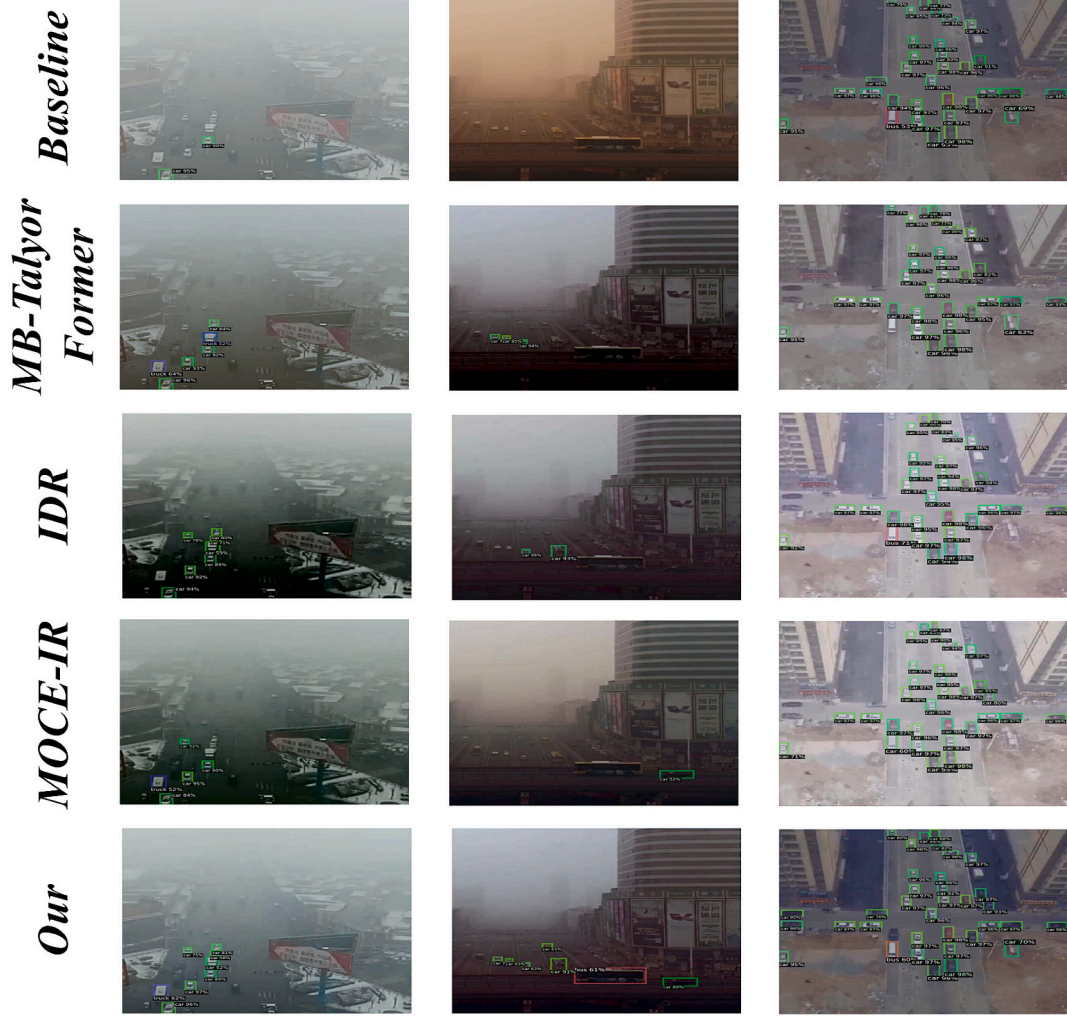


Fig. 8. Detection Visualization Result in the Real-world Test Set.

Table 3
Detection Performance Results under Different Scenes.

	Clear	Foggy	Dusty	Lowlight
D-Agricultural diseases	85.681	77.242	19.645	51.686
TDRE-T	84.892	79.096	71.517	59.333
TDRE-A	84.872	80.871	73.427	64.816
D-City Waste	81.701	79.285	68.858	72.888
TDRE-T	80.179	79.673	78.341	75.065
TDRE-A	81.100	79.606	78.062	75.072
D-Wilderness Rescue	71.850	47.169	19.395	43.059
TDRE-T	66.363	51.722	48.169	43.770
TDRE-A	67.786	53.232	48.137	44.762
D-Intelligent Transportation	65.853	50.971	48.038	57.494
TDRE-T	65.379	55.290	61.879	58.985
TDRE-A	65.618	54.602	60.061	58.126

We observed different enhancement effects across the various applications. In agricultural disease detection, the targets generally have high reflectance and lack clear geometric shapes. Because they rely heavily on subtle color features, TDRE significantly boosts detection performance under all degradation conditions. For city waste and transportation tasks, the targets are larger and have distinct geometric features. The baseline model handles them reasonably well, but TDRE still provides a considerable improvement. In wilderness rescue, the original dataset already contains mildly degraded images. This gives the baseline detector some natural robustness, meaning our enhancement

provides a positive but more limited gain.

Finally, we analyzed the results across different weather conditions. For clear images, our initial routing mechanism ensures that the original accuracy is barely affected. This actively avoids the severe performance drops frequently seen in traditional restoration models. Under foggy and low-light conditions, the images experience global color shifts, but their core distribution patterns are not entirely destroyed. TDRE effectively corrects these shifts to improve detection. Dusty conditions, however, directly and completely disrupt the RGB channels. As a result, the baseline detector suffers a massive performance drop. Yet, by applying our enhancer, the model achieves a substantial recovery and a striking improvement in final accuracy.

4.5. Generalization experiment of UAV actual flight data

To further verify the practical generalization of the proposed TDRE, we conducted additional evaluations using unconstrained, real-world UAV flight data. Standard datasets typically feature unified image sizes and fixed data distributions. These controlled conditions may not fully represent the extreme variability and unpredictability of actual aerial photography. Therefore, we selected newly captured degraded images from authentic drone flights to serve as an ultimate test bed.

The visual detection results are illustrated in Fig. 10. As shown in the figure, TDRE effectively eliminates the severe color distortions and low contrast caused by authentic adverse weather conditions. More importantly, it dramatically amplifies the structural features of targets that

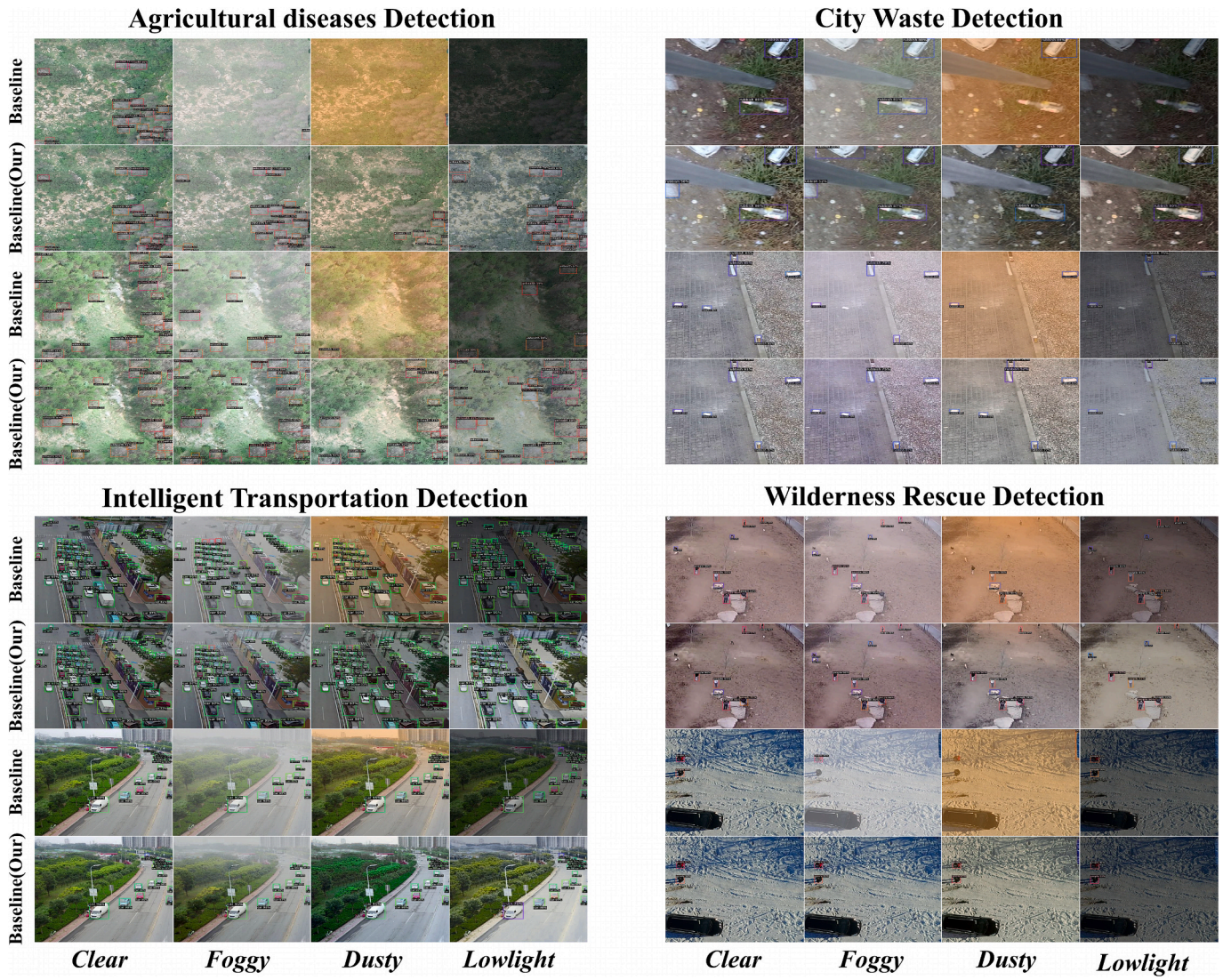


Fig. 9. Detection Visualization Result in Different Applications.



Fig. 10. Detection Results of UAV Actual Flight Data.

are extremely small or located at far distances.

By leveraging our synergistic optimization mechanism, the enhancer provides a robust and clear visual input without requiring any modifications to the frozen downstream detector. Consequently, the baseline detector successfully identifies challenging targets that were previously lost in the severe atmospheric degradation. This confirms the exceptional cross-domain generalization and high reliability of our lightweight architecture in authentic UAV applications.

5. Discussion

5.1. Component Ablation experiment

To validate the individual contributions of our core designs, we conducted comprehensive component ablation experiments. The evaluated modules include the DR-MoE framework trained solely with a standard L2 loss, the multi-space restoration (MSR) loss, the clear-sky perceptron, and the detection-oriented enhancement (DE) loss. During the training of each progressive stage, the parameters of all preceding stages are frozen. The quantitative results are presented in Table 4.

The experimental results clearly demonstrate the necessity of each component. When optimizing the DR-MoE parallel experts using only the L2 loss, the model improves detection performance in foggy and dusty scenarios. However, it suffers a noticeable performance drop under low-light conditions. By introducing the MSR loss across the RGB, HSV, and LAB domains, the model explicitly constrains complex color distortions. This addition effectively recovers and boosts detection accuracy in low-light environments.

The experimental results clearly demonstrate the necessity of each component. When optimizing the DR-MoE parallel experts using only the L2 loss, the model improves detection performance in foggy and dusty scenarios. However, it suffers a noticeable performance drop under low-light conditions. To further analyze the source of this improvement, we additionally conducted independent contribution ablation experiments on different color-space losses, including RGB, HSV, LAB, and their combinations. The results show that the RGB-domain loss helps preserve pixel-level structural consistency, the HSV-domain loss more directly constrains hue and saturation shifts caused by illumination and atmospheric degradation, and the LAB-domain loss provides a perceptually meaningful color correction constraint. By introducing the MSR loss across the RGB, HSV, and LAB domains, the model explicitly constrains complex color distortions from complementary color representations. This addition effectively recovers and boosts detection accuracy in low-light environments and provides more stable performance across different degraded conditions.

To further verify the final impact of different color-space losses on detection performance, we additionally conducted independent contribution ablation experiments on the MSR loss as shown in Table 5. Specifically, we compared the results of using the RGB, HSV, and LAB losses individually, using their pairwise combinations, and using all three color-space losses together. These comparisons allow us to evaluate how each color representation contributes to the final detection performance.

The RGB-domain loss helps preserve pixel-level structural consistency, the HSV-domain loss more directly constrains hue and saturation

Table 4
Component Ablation Experiments Result.

	Component	Clear	Foggy	Dusty	Lowlight	Real
(A)	Baseline	65.853	50.971	48.038	57.494	24.247
(B)	A + DR-MoE	61.178	54.092	61.418	49.512	25.137
(C)	B + MSR Loss	61.125	54.139	61.569	57.258	25.922
(D)	C + Clear-Sky Perceptron	65.454	53.931	61.268	57.932	26.015
(E)	D + DE loss	65.379	55.290	61.879	58.985	26.668

Table 5

Ablation Study on the Impact of Different Color-Space Losses on Detection Performance.

Loss	Clear	Foggy	Dusty	Lowlight	Real
RGB	61.178	54.092	61.418	49.512	25.137
HSV	60.683	53.931	61.023	53.229	25.209
LAB	61.277	53.666	61.009	52.490	25.459
RGB + HSV	61.101	53.906	61.441	54.987	25.710
RGB + LAB	61.037	53.964	61.186	55.814	25.698
HSV + LAB	60.891	54.021	61.228	56.376	25.733
MSR Loss	61.125	54.139	61.569	57.258	25.922

Table 6

Results of Generalization Experiments for Different Detection Baselines.

	Clear	Foggy	Dusty	Lowlight	Real
Baseline_1	54.203	39.745	44.610	42.947	18.300
Baseline_1 + TDRE	-0.228	+4.058	+5.276	+5.219	+1.727
Baseline_2	62.459	47.128	47.234	52.476	25.729
Baseline_2 + TDRE	-0.263	+3.608	+12.903	+0.932	+2.009
Baseline_3	65.853	50.971	48.038	57.494	24.247
Baseline_3 + TDRE	-0.474	+4.319	+13.841	+1.491	+2.421

shifts caused by illumination and atmospheric degradation, and the LAB-domain loss provides a perceptually meaningful color correction constraint. The results also show that different color-space losses exhibit inconsistent effects under different weather conditions, indicating that a single color representation cannot fully handle diverse degradations. In contrast, using the three color-space losses together achieves the best detection performance across multiple degraded weather conditions and the real-world degraded test set, demonstrating that RGB, HSV, and LAB provide complementary constraints for detection-oriented restoration.

Furthermore, feeding original clear images into the restoration network causes unnecessary smoothing, which leads to a severe accuracy drop on the clear test set. The proposed clear-sky perceptron successfully resolves this issue by routing normal inputs around the enhancer. Finally, incorporating the spatial mask-guided DE loss forces the network to concentrate exclusively on target regions. This explicit task-aligned mechanism comprehensively maximizes the detector's perception across all adverse conditions, particularly in real-world scenarios.

To better illustrate the effectiveness of the MSR loss, we visualized the 3D scatter plots of the color space distributions. As shown in Fig. 11, the original degraded images exhibit severe distributional shifts compared to the clear images. These shifts critically disrupt the feature extraction of the downstream detector. Conversely, the color distributions of our intermediate restored images align tightly with those of the clear images. This authentic color recovery provides a highly robust foundation for object detection.

Finally, to validate our target-driven enhancement mechanism, we visualized the feature differences using heatmaps. Specifically, the heatmaps were generated from the feature differences extracted from the Stem layer of the detector backbone. As illustrated in Fig. 12, under various simulated and real-world degradations, the final enhanced features exhibit exceptionally high activation values precisely on the potential objects. This confirms that our network successfully discards redundant background reconstruction. Instead, it dedicates its limited computational power entirely to amplifying target-level discriminability, thereby directly improving the final detection performance.

5.2. Generalizability experiments of different detection baselines

To thoroughly validate the efficacy and architectural agnosticism of TDRE, we conducted generalization experiments across various mainstream detection baselines. A robust detection-oriented enhancer must demonstrate consistent performance gains regardless of the underlying

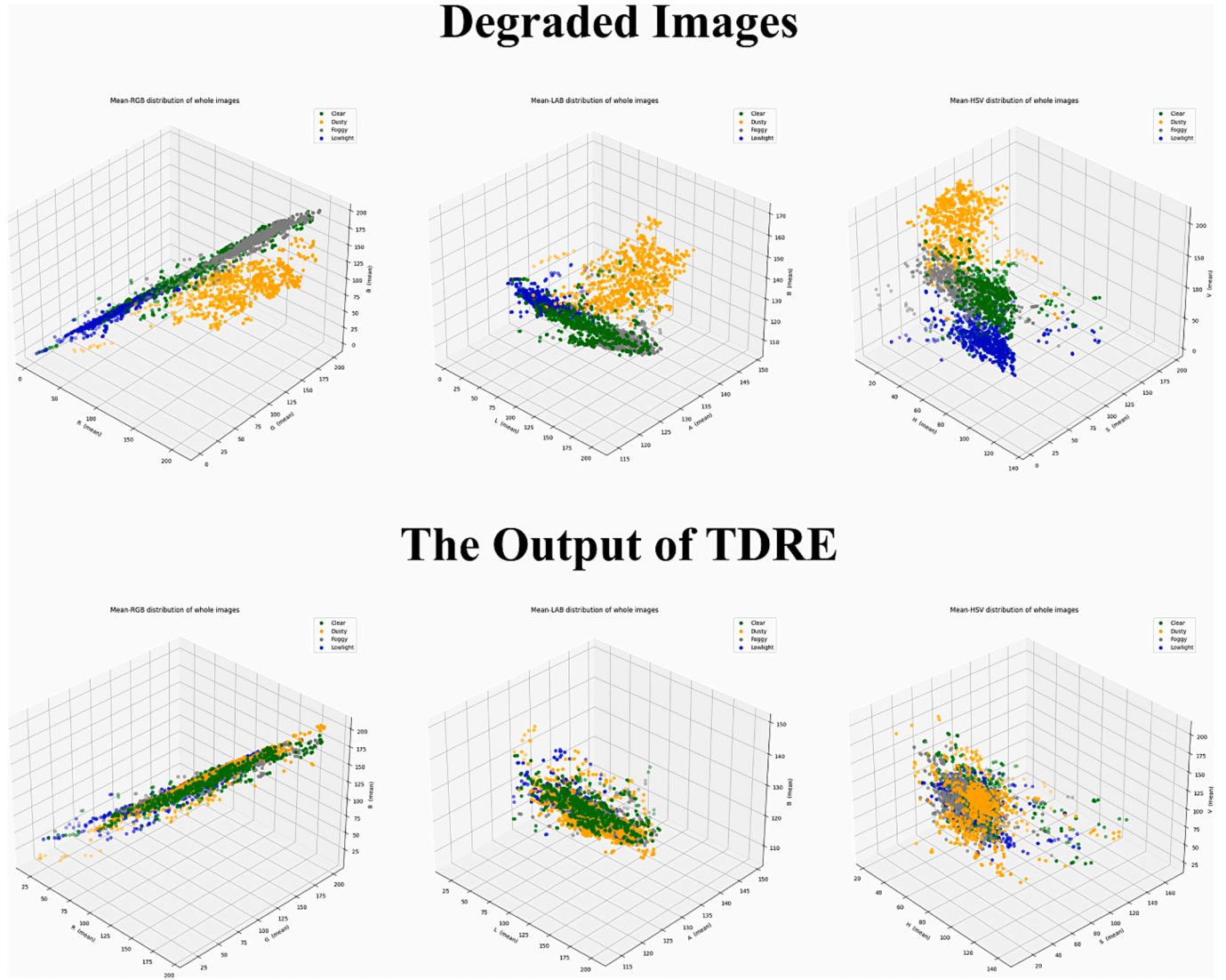


Fig. 11. Comparison of 3D Scatter Plots in Multi Color Spaces, the first row shows the original color space distribution results, and the second row shows the output of TDRE.

detection architecture. In this comprehensive evaluation, we selected three representative models: Faster R-CNN (Baseline 1) as a classic two-stage detector, Sparse R-CNN (Baseline 2) as a query-based detector, and DiffusionDet (Baseline 3) as a generative paradigm detector. The quantitative results are summarized in Table 6, where underlined indicates performance decline and bolded indicates performance improvement.

As demonstrated by the experimental data, TDRE effectively and consistently boosts both the detection accuracy and environmental robustness of all tested detectors under adverse weather conditions. The enhancements are highly significant. Specifically, when plugged into the frozen baseline architectures, the overall mAP50 metric increased by 16.052 for Faster R-CNN, 19.189 for Sparse R-CNN, and 21.598 for DiffusionDet.

These substantial performance gains confirm that our synergistic optimization mechanism is not overfitted to a specific detection framework. Instead, TDRE successfully provides a universally reliable visual foundation. By filtering out task-irrelevant degradations and amplifying structural features, it empowers diverse detection architectures to reliably penetrate severe domain shifts.

5.3. Hyperparameter experiment

To better balance the classification performance between clear images and degraded images, and to comprehensively improve the detection performance under degraded conditions while minimizing the impact on the detection results of clear images, hyperparameter experiments were conducted on the threshold of the clear-aware score S_{ca} and the number of experts in DR-MoE. Specifically, the clear-aware score threshold was ablated to determine a suitable routing criterion, while the number of experts in DR-MoE was further analyzed to evaluate its influence on the balance between degradation modeling capacity and detection performance.

The thresholds of clear-aware-score were set at 0.5, 0.6, 0.7, 0.8, and 0.9, where this value indicates that images with a probability greater than this value after the softmax activation function are classified as clear images. The experimental results are shown in Table 7. The experimental results show that when the threshold is higher, that is, when more clear images are classified as degraded images, although the detection performance under degraded conditions can be slightly enhanced in some cases, it leads to a significant drop in the detection performance of clear images. Moreover, the performance on the real test set also decreases. Therefore, TDRE sets this hyperparameter to 0.5.

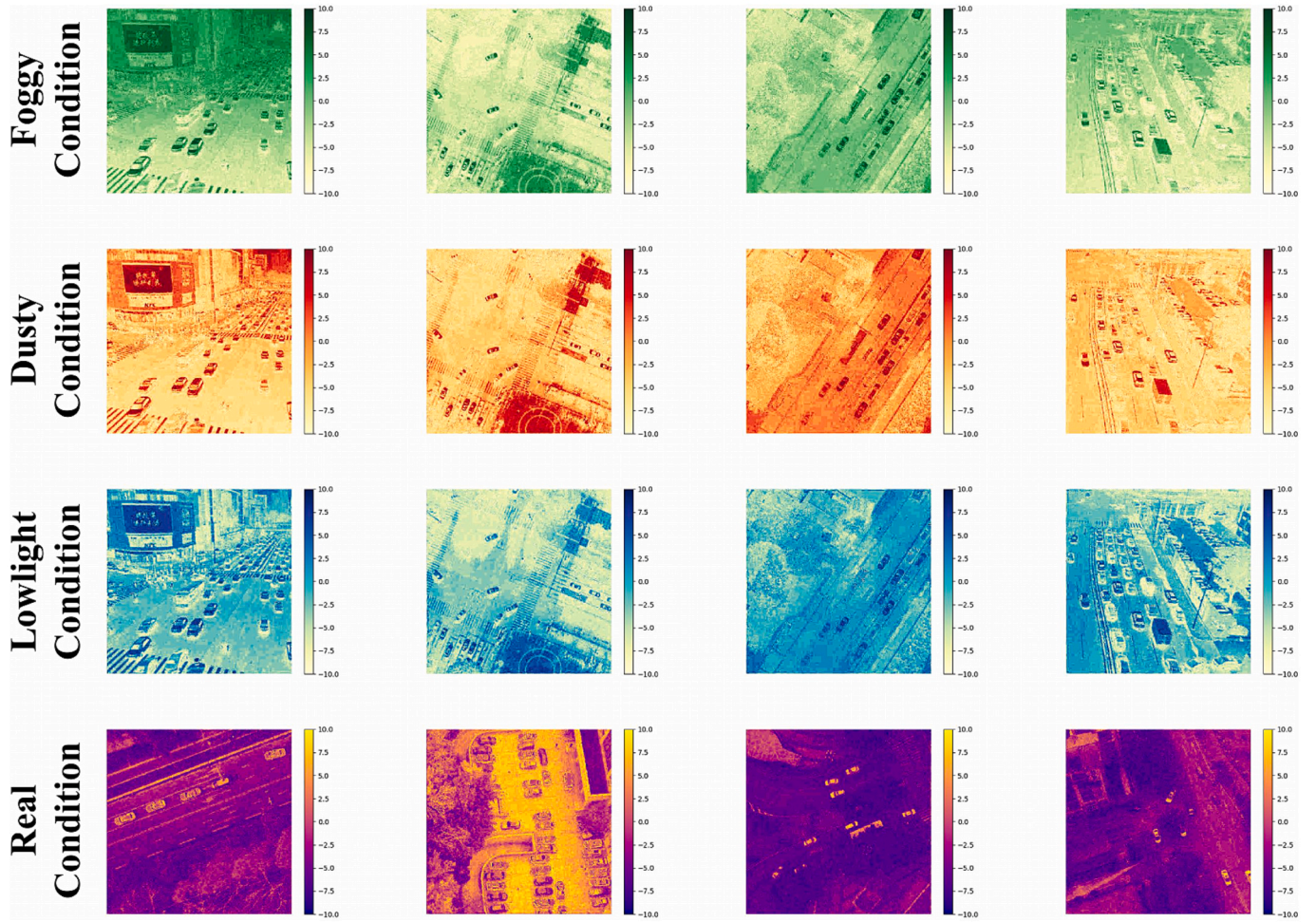


Fig. 12. Visualization of the Heatmap of the Difference between Detection Enhancement Results and Restoration Results.

Table 7
Ablation Study Results on the Threshold of the Clear Aware Score.

S_{ca}	Clear	Foggy	Dusty	Lowlight	Real
0.5	65.379	55.290	61.879	58.985	26.668
0.6	64.077	55.505	61.909	58.783	26.570
0.7	63.044	55.383	61.925	58.646	26.579
0.8	62.352	55.237	62.023	58.386	26.532
0.9	61.588	55.240	61.966	58.386	26.490

Table 8
Ablation Study Results on the Number of Experts (Num_E) in DR-MoE.

Num_E	Clear	Foggy	Dusty	Lowlight	Real
1	64.822	54.382	61.301	58.634	25.854
2	65.289	55.111	61.422	58.981	26.113
3	65.379	55.290	61.879	58.985	26.668
4	65.214	55.311	61.782	58.661	26.410

The Ablation Study Results on the Number of Experts in DR-MoE is shown in Table 8. The results show that increasing the number of experts from 1 to 3 generally improves the detection performance, especially on the real-world degraded test set. This indicates that multiple experts help DR-MoE better model heterogeneous weather degradations. However, further increasing the number of experts to 4 does not consistently improve performance. Although the 4-expert setting obtains a slightly higher result on the foggy subset, it decreases the performance on clear, dusty, low-light, and real-world degraded images. Therefore, we adopt 3

experts in DR-MoE, which achieves the best overall balance between degradation-specific modeling capacity, detection accuracy, and lightweight computation.

5.4. Model efficiency and complexity analysis

To comprehensively evaluate the performance of the proposed TDRE, we compared it with state-of-the-art methods across multiple metrics, including the number of parameters, inference speed, GPU memory usage, and detection accuracy on a real-world test set (Table 9). The reported inference time of TDRE measures the complete end-to-end forward process, including the routing calculation of the Clear-Sky Perceptron and the subsequent enhancement stage. The results demonstrate that our method achieves approximately 100 times fewer parameters and over ten times the inference speed of the high-

Table 9
Comprehensive Performance Comparison Result.

Methods	Params (M)	Inference Speed (s)	GPU Memory (MB)	Performance (mAP)
AODNet	1.76 K	0.0006	58.0	20.875
GridDehazeNet	958.05 K	0.0178	351.7	20.622
FFANet	4455.91 K	0.1701	472.2	24.100
MB-TaylorFormer	2676.52 K	0.3674	1545.5	25.221
IDR	2985.09 K	0.0803	1737.6	25.545
MOCE-IR	2950.40 K	0.0873	454.5	25.561
TDRE	2.93 K	0.0020	77.2	26.668

performing models, such as IDR / MOCE-IR / MB-TaylorFormer, while maintaining superior detection accuracy.

This efficiency is further validated by its significantly lower GPU memory footprint, making it ideal for edge deployment in real-time aerial detection tasks. These comprehensive results are visualized in Fig. 13, where the x-axis denotes inference speed, the y-axis represents detection accuracy, and the bubble size indicates the number of parameters. This visualization clearly highlights the proposed model's superior balance between detection performance and deployment efficiency.

Furthermore, we conducted extensive hardware-specific experiments to assess practical utility. In CPU-based tests, we analyzed the correlation between CPU utilization and inference latency, confirming the model's stability under varying computational loads. On GPU platforms, we evaluated the impact of input resolution on processing speed, demonstrating consistent real-time performance across different image scales.

The results for inference speed under varying CPU utilization are presented in Table 10 and Fig. 14. The data indicate that TDRE maintains acceptable inference speeds even under constrained CPU resources, which is critical for deployment on mobile platforms such as UAVs. Conversely, competing methods exhibit a sharp increase in latency as CPU availability decreases, with processing times often reaching the order of seconds. Such delays are prohibitive for real-time UAV applications, highlighting the superior efficiency and robustness of our proposed model.

With continuous advancements in UAV sensor technology, the ability to process high-resolution imagery has become a critical performance metric. The correlation between input resolution and inference speed on the GPU is illustrated in Fig. 15. The results demonstrate that the proposed TDRE maintains efficient processing speeds even at large scales, such as 4096×4096 pixels. In contrast, several compared methods, including MB-TaylorFormer, IDR, and MOCE-IR, suffer from Out of Memory (OOM) errors at these resolutions, while others require seconds per frame. These findings underscore the superior memory efficiency and real-time scalability of our model for high-definition aerial imaging.

5.5. Failure Case analysis

Although the proposed TDRE achieves consistent performance improvements under various degraded conditions, several challenging failure cases still remain. To further analyze the limitations of our method, we visualize three representative failure cases under extremely low-light and dense foggy conditions. As shown in Fig. 16, when the degradation is particularly severe, the domain gap between the degraded target images and the clear source-domain images becomes extremely large. In such cases, the restoration results produced by TDRE are still not fully satisfactory, and the detection results may contain missed detections or inaccurate localization.

This limitation mainly arises because the discriminative object cues in the original images are severely weakened by extreme illumination degradation or dense atmospheric interference. When the object texture, boundary, and color information are largely suppressed, the enhancement module cannot completely recover the missing visual information. Consequently, the detector may still fail to identify some objects, especially small or heavily degraded targets.

Nevertheless, compared with the original baseline detector, TDRE still improves the visual consistency between degraded images and clear images to a certain extent. Specifically, the enhanced results exhibit better color consistency, clearer local contrast, and more salient object regions, which help the detector capture more reliable target-level cues. Therefore, even in these difficult failure cases, the proposed method can still achieve better detection performance than the baseline model. These observations indicate that future work may further explore stronger domain adaptation strategies or multi-frame contextual information to improve robustness under extremely challenging degradation conditions.

6. Conclusion

In this paper, we identified that aerial object detection under adverse weather conditions is fundamentally hindered by severe domain shifts. Existing solutions frequently fall into two critical traps: direct model training inevitably triggers catastrophic forgetting on clear images, while traditional distributed methods suffer from a severe human-visual bias, leading to task inconsistency and wasted computational resources.

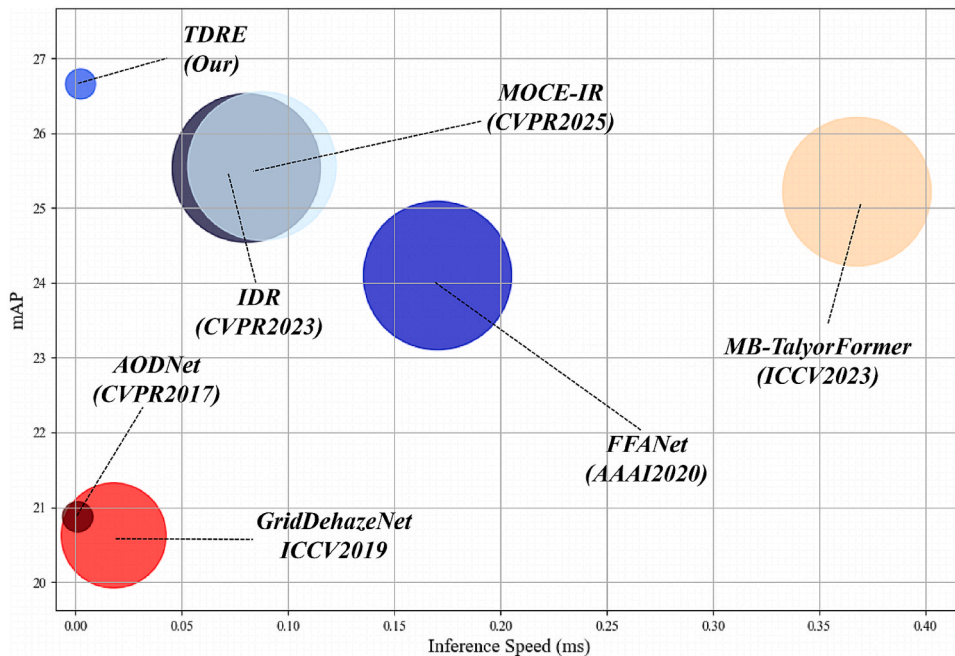
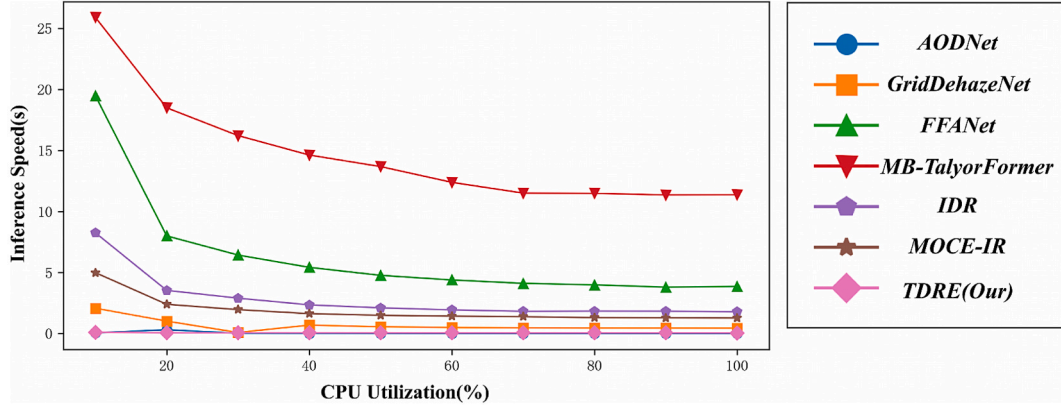
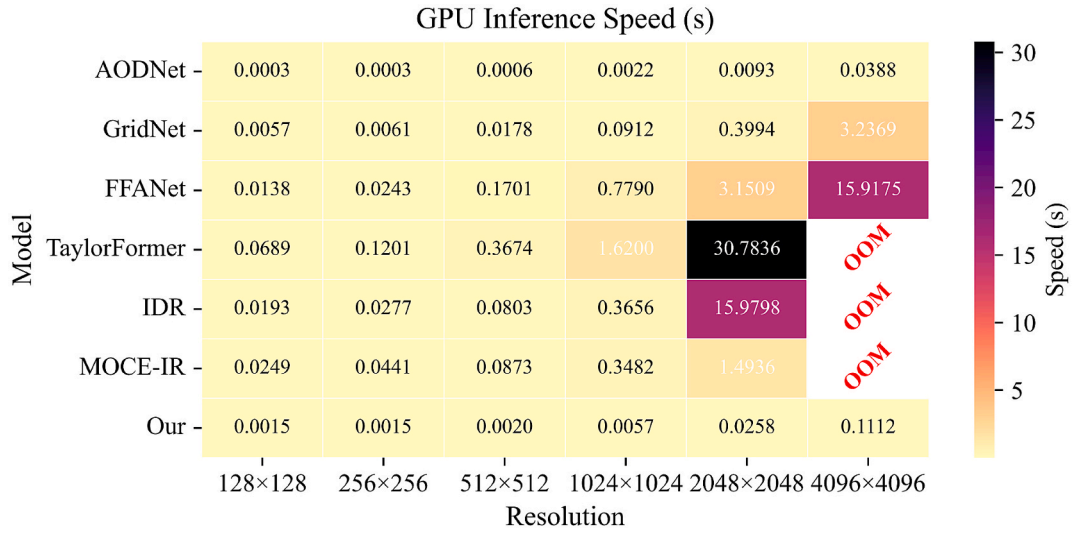


Fig. 13. Visualization of Comprehensive Performance.

Table 10

The Results for Inference Speed under Varying CPU Utilization.

Methods	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %	100 %
AODNet	0.0605	0.3128	0.0261	0.0200	0.0157	0.0121	0.0101	0.0098	0.0092	0.0095
GridDehazeNet	2.0813	1.0206	0.0873	0.6932	0.5585	0.5038	0.4661	0.4522	0.4498	0.4456
FFANet	19.4919	8.0047	6.4357	5.4297	4.7730	4.3918	4.1153	3.9901	3.8018	3.8599
MB-TaylorFormer	25.9058	18.5024	16.2235	14.6319	13.6839	12.3918	11.5153	11.4922	11.3609	11.3745
IDR	8.2574	3.5389	2.904	2.3529	2.1096	1.9436	1.8232	1.8416	1.8356	1.7912
MOCE-IR	4.9928	2.4002	1.9641	1.6333	1.4984	1.4228	1.3881	1.3002	1.2895	1.2743
TDRE	0.1106	0.0676	0.0622	0.0609	0.0581	0.0557	0.0538	0.0541	0.0529	0.0522

**Fig. 14.** Line Plot of Inference Speed as a Function of CPU Utilization.**Fig. 15.** Heatmap Visualization of Model Inference Speed as a Function of Image Resolution.

To break this deadlock, we proposed the Transferable Dynamic Routing Enhancer (TDRE), an ultra-lightweight, plug-and-play visual filter designed exclusively for machine perception. At its architectural core, the Dynamic Routing Mixture-of-Experts (DR-MoE) explicitly disentangles multifaceted atmospheric degradations in a single forward pass. Concurrently, the novel clear-sky perceptron acts as a highly reliable hard gate. By seamlessly bypassing undegraded inputs, it effectively alleviates the catastrophic forgetting problems of clear images. Furthermore, we introduced a synergistic optimization mechanism to explicitly resolve the task inconsistency problem. By tightly coupling a multi-space restoration loss with a detection-oriented masking strategy, our framework proactively discards visually redundant background reconstruction. It forces the network to concentrate its limited computational capacity entirely on amplifying the high-frequency structural

features and color fidelity of potential targets.

Extensive evaluations on our comprehensive Cross-Application Degradation Aerial Detection Dataset (CAD-ADD) thoroughly validate the superiority of the proposed method. With an incredibly compact architecture of fewer than 3 K additional parameters and extremely fast inference speeds, TDRE consistently and significantly boosts the robust generalization of various frozen baseline detectors, yielding a remarkable cumulative improvement of 126.9 mAP50 points across 17 distinct test sets. It successfully provides a highly reliable visual foundation for complex and unpredictable real-world aerial applications.

CRediT authorship contribution statement

Yu Wan: Writing – review & editing, Writing – original draft,

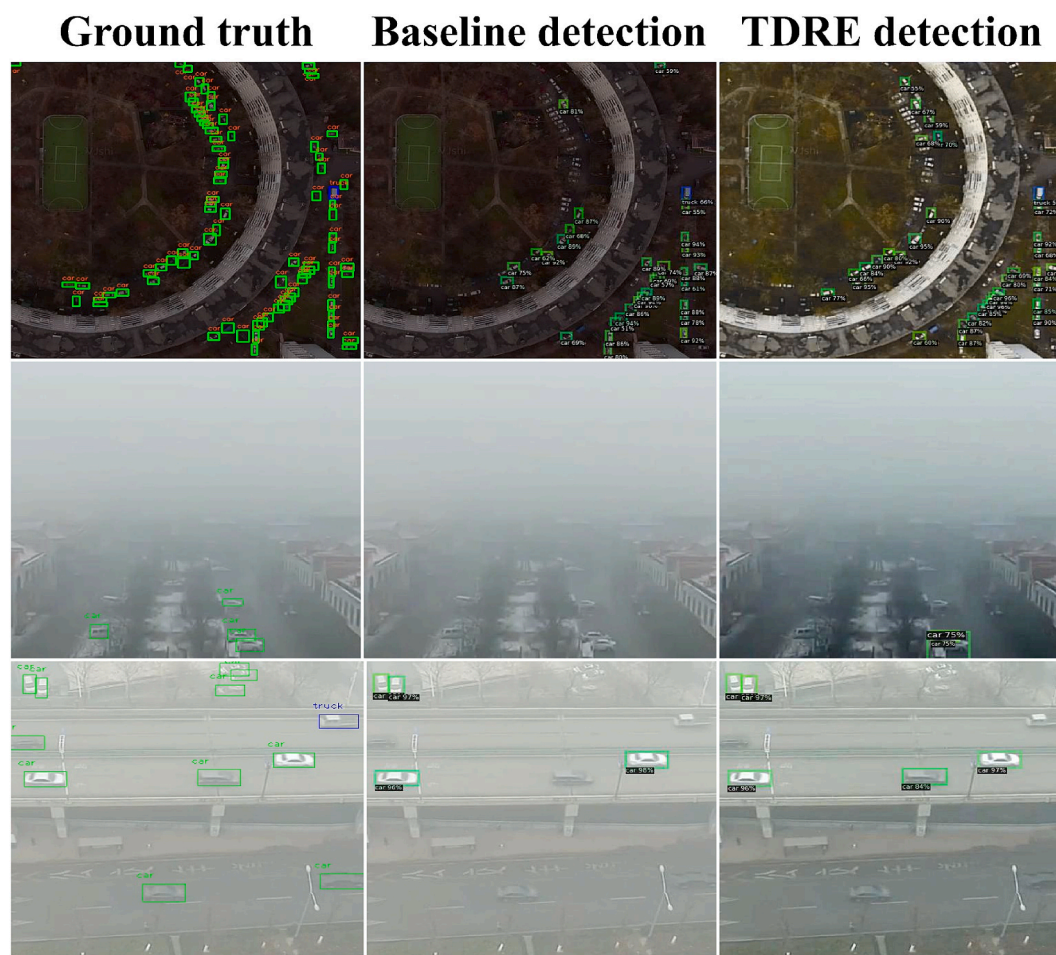


Fig. 16. Visualization of representative failure cases under extremely degraded conditions.

Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jie Li**: Writing – review & editing, Visualization, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Liupeng Lin**: Funding acquisition, Data curation, Conceptualization. **Zhaowen Lv**: Writing – review & editing, Investigation, Formal analysis. **Zaiyan Zhang**: Writing – review & editing, Investigation, Formal analysis, Conceptualization. **Qiangqiang Yuan**: Writing – review & editing, Resources, Funding acquisition. **Huanfeng Shen**: Funding acquisition, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work is supported in part by the National Natural Science Foundation of China under Grant 42471504 and 42301417 and the Fundamental Research Funds for the Central Universities under Grant No. 2042026kf001. (Corresponding author: Jie Li.).

Data availability

Data will be made available on request.

References

- Badgujar, C.M., Poulse, A., Gan, H., 2024. Agricultural object detection with you only look once (YOLO) Algorithm: a bibliometric and systematic literature review. *Comput. Electron. Agric.* 223, 18.
- Bochkovskiy, A., Wang, C.-Y. and Liao, H.-Y.M., 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. *ArXiv*, abs/2004.10934.
- Caron, M., Touvron, H., Misra, I., Jegou, H., Mairal, J., Bojanowski, P., Joulin, A. and Ieee, 2021. Emerging Properties in Self-Supervised Vision Transformers, 18th IEEE/CVF International Conference on Computer Vision (ICCV). Ieee, Electr Network, pp. 9630–9640.
- Chen, S., Sun, P., Song, Y., Luo, P., 2022. DiffusionDet: Diffusion Model for Object Detection. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 19773–19786.
- Cui, Y., Zamir, S.W., Khan, S.H., Knoll, A., Shah, M. and Khan, F.S., 2024. AdaIR: Adaptive All-in-One Image Restoration via Frequency Mining and Modulation. *ArXiv*, abs/2403.14614.
- Felzenszwalb, P., McAllester, D., Ramanan, D. and Ieee, 2008. A discriminatively trained, multiscale, deformable part model, IEEE Conference on Computer Vision and Pattern Recognition. IEEE Conference on Computer Vision and Pattern Recognition. Ieee, Anchorage, AK, pp. 1984–+.
- Girshick, R., Donahue, J., Darrell, T., Malik, J. and Ieee, 2014. Rich feature hierarchies for accurate object detection and semantic segmentation, 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Conference on Computer Vision and Pattern Recognition. Ieee, Columbus, OH, pp. 580–587.
- Han, J.L., Li, W.H., Fang, P.F., Sun, C.Y., Hong, J., Armin, M.A., Petersson, L. and Li, H. D., 2022. Blind Image Decomposition, 17th European Conference on Computer Vision (ECCV). Lecture Notes in Computer Science. Springer International Publishing Ag, Tel Aviv, ISRAEL, pp. 218–237.
- Hu, K., Wu, F., Zhan, Z.F., Luo, J., Pu, H.Y., 2024. High-low level task combination for object detection in foggy weather conditions. *J. Vis. Commun. Image Represent.* 98, 8.
- Huang, S.C., Le, T.H., Jaw, D.W., 2021. DSNet: Joint Semantic Learning for Object Detection in Inclement Weather Conditions. *IEEE Trans. Pattern Anal. Mach. Intell.* 43 (8), 2623–2633.

- Kraft, M., Piechocki, M., Ptak, B., Walas, K., 2021. Autonomous, Onboard Vision-based Trash and Litter Detection in Low Altitude Aerial Images Collected by an Unmanned Aerial Vehicle. *Remote Sens. (Basel)* 13 (5), 965.
- Li, B.Y., Liu, X., Hu, P., Wu, Z.Q., Lv, J.C., Peng, X. and Ieee Comp, S.O.C., 2022. All-In-One Image Restoration for Unknown Corruption, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Conference on Computer Vision and Pattern Recognition. Ieee Computer Soc, New Orleans, LA, pp. 17431-17441.
- Li, C.Y., Zhou, H., Liu, Y., Yang, C.D., Xie, Y.Q., Li, Z.B., Zhu, L.P., 2023. Detection-Friendly Dehazing: Object Detection in Real-World Hazy Scenes. *IEEE Trans. Pattern Anal. Mach. Intell.* 45 (7), 8284–8295.
- Li, R.T., Tan, R.T., Cheong, L.F. and Ieee, 2020. All in One Bad Weather Removal using Architectural Search, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Conference on Computer Vision and Pattern Recognition. Ieee Computer Soc, null, ELECTR NETWORK, pp. 3172-3182.
- Li, Y. and Zhao, Y.F., 2023. RSID: A Remote Sensing Image Dehazing Network, 6th Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Lecture Notes in Computer Science. Springer-Verlag Singapore Pte Ltd, Xiamen Univ, Xiamen, PEOPLES R CHINA, pp. 3-14.
- Lin, T.Y., Goyal, P., Girshick, R., He, K.M., Dollár, P. and Ieee, 2017. Focal Loss for Dense Object Detection, 16th IEEE International Conference on Computer Vision (ICCV). IEEE International Conference on Computer Vision. Ieee, Venice, ITALY, pp. 2999-3007.
- Lin, X., Yue, J., Chan, K.C.K., Qi, L., Ren, C., Pan, J., Yang, M.-H., 2023. Multi-Task Image Restoration Guided by Robust DINO Features. *ArXiv, abs/2312.01677*.
- Liu, W.Y., Ren, G.F., Yu, R.S., Guo, S., Zhu, J.K., Zhang, L. and Assoc Advancement Artificial, I., 2022. Image-Adaptive YOLO for Object Detection in Adverse Weather Conditions, 36th AAAI Conference on Artificial Intelligence. AAAI Conference on Artificial Intelligence. Assoc Advancement Artificial Intelligence, Electr Network, pp. 1792-1800.
- Long, J., Shi, Z.W., Tang, W., Zhang, C.S., 2014. Single Remote Sensing image Dehazing. *IEEE Geosci. Remote Sens. Lett.* 11 (1), 59–63.
- Luo, Z., Gustafsson, F.K., Zhao, Z., Sjolund, J., Schon, T.B., 2023. Controlling Vision-Language Models for Multi-Task image Restoration. *International Conference on Learning. Representations*.
- Qin, X., Wang, Z.L., Bai, Y.C., Xie, X.D., Jia, H.Z. and Assoc Advancement Artificial, I., 2020. FFA-Net: Feature Fusion Attention Network for Single Image Dehazing, 34th AAAI Conference on Artificial Intelligence / 32nd Innovative Applications of Artificial Intelligence Conference / 10th AAAI Symposium on Educational Advances in Artificial Intelligence. AAAI Conference on Artificial Intelligence. Assoc Advancement Artificial Intelligence, New York, NY, pp. 11908-11915.
- Qiu, Y.W., Zhang, K.H., Wang, C.X., Luo, W.H., Li, H.D., Jin, Z. and Ieee, 2023. MB-TaylorFormer: Multi-branch Efficient Transformer Expanded by Taylor Formula for Image Dehazing, IEEE/CVF International Conference on Computer Vision (ICCV). IEEE International Conference on Computer Vision. Ieee Computer Soc, Paris, FRANCE, pp. 12756-12767.
- Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G. and Sutskever, I., 2021. Learning Transferable Visual Models From Natural Language Supervision, International Conference on Machine Learning (ICML). Proceedings of Machine Learning Research. Jmlr-Journal Machine Learning Research, Electr Network.
- Redmon, J., Divvala, S., Girshick, R., Farhadi, A. and Ieee, 2016. You Only Look Once: Unified, Real-Time Object Detection, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Conference on Computer Vision and Pattern Recognition. Ieee, Seattle, WA, pp. 779-788.
- Redmon, J. and Farhadi, A., 2018. YOLOv3: An Incremental Improvement. *arXiv e-prints*.
- Redmon, J., Farhadi, A. and Ieee, 2017. YOLO9000: Better, Faster, Stronger, 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Conference on Computer Vision and Pattern Recognition. Ieee, Honolulu, HI, pp. 6517-6525.
- Ren, S.Q., He, K.M., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (6), 1137–1149.
- Sun, P.Z., Zhang, R.F., Jiang, Y., Kong, T., Xu, C.F., Zhan, W., Tomizuka, M., Li, L., Yuan, Z.H., Wang, C.H., Luo, P., Ieee Comp, S.O.C., 2021. Sparse R-CNN: End-to-End Object Detection with Learnable proposals. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Conference on Computer Vision and Pattern Recognition. Ieee Computer Soc, Electr Network, pp. 14449–14458.
- Wan, Y., Li, J., Lin, L., Yuan, Q., Shen, H., 2025. Collaboration of Dehazing and Object Detection Tasks: a Multi-Task Learning Framework for Foggy image. *IEEE Trans. Geosci. Remote Sens.* 1.
- Wang, C., Li, C., Luo, B., Wang, W. and Liu, J., 2021. RiWNet: A moving object instance segmentation Network being Robust in adverse Weather conditions, pp. arXiv: 2109.01820.
- Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M. and Ieee, 2023. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Conference on Computer Vision and Pattern Recognition. Ieee Computer Soc, Vancouver, CANADA, pp. 7464-7475.
- Wang, H.H., Tsai, F.J., Lin, Y.Y. and Lin, C.W., 2024. TANet: Triplet Attention Network for All-In-One Adverse Weather Image Restoration, 17th Asian Conference on Computer Vision-ACCV-Biennial. Lecture Notes in Computer Science. Springer-Verlag Singapore Pte Ltd, Hanoi, VIETNAM, pp. 3-19.
- Xia, J., Chen, T., Qiao, J., 2022. Vehicle Detection based on YOLOv3 in adverse Weather Conditions. In: 2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), pp. 692–696.
- Yang, S., Lu, H.M., Li, J.R., 2023. Multifeature Fusion-based Object Detection for Intelligent Transportation Systems. *IEEE Trans. Intell. Transp. Syst.* 24 (1), 1126–1133.
- Yu, X., Zhou, S., Li, H., Zhu, L., 2024. Multi-Expert Adaptive selection: Task-Balancing for All-in-one image Restoration. *IEEE Trans. Circuits Syst. Video Technol.* 35, 4619–4634.
- Zamfir, E., Wu, Z., Mehta, N., Tan, Y., Paudel, D.P., Zhang, Y., Timofte, R., 2024. Complexity experts are Task-Discriminative Learners for any image Restoration. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12753–12763.
- Zhang, S., Chi, C., Yao, Y., Lei, Z. and Li, S.Z., 2019. Bridging the Gap Between Anchor-based and Anchor-free Detection via Adaptive Training Sample Selection, pp. arXiv: 1912.02424.
- Zhang, X., Ma, J., Wang, G., Zhang, Q., Zhang, H., Zhang, L., 2025. Perceive-IR: Learning to Perceive Degradation Better for All-in-one image Restoration. *IEEE Trans. Image Process.* 1.
- Zhang, Y., Xue, J.R., Zhang, G., Zhang, Y.W., Zheng, N.N. and Ieee, 2014. A Multi-Feature Fusion Based Traffic Light Recognition Algorithm for Intelligent Vehicles, 33rd Chinese Control Conference (CCC). Chinese Control Conference. Ieee, Nanjing, PEOPLES R CHINA, pp. 4924-4929.
- Zhang, Y., Ye, M., Zhu, G.Y., Liu, Y., Guo, P.Y., Yan, J.H., 2024. FFCA-YOLO for Small Object Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* 62, 15.
- Zhang, Z., Li, J., Liang, Y., Yan, J., Xiao, Y., Su, X., Yuan, Q., 2026. ECRformer: an efficient cloud removal Transformer with semantic-decoupled learning for multimodal satellite imagery. *ISPRS J. Photogramm. Remote Sens.* 237, 323–338.
- Zhang, Z., Zhao, L., Liu, Y., Zhang, S., Yang, J., 2020. Unified Density-Aware image Dehazing and Object Detection in Real-World Hazy Scenes. *Asian Conference on Computer Vision*.
- Zhao, C.J., Liu, R.W., Qu, J.X., Gao, R.B., 2024. Deep learning-based object detection in maritime unmanned aerial vehicle imagery: Review and experimental comparisons. *Eng. Appl. Artif. Intel.* 128, 21.
- Zhao, T., Feng, R.Y., Wang, L.Z., 2025. SCENE-YOLO: a One-Stage Remote Sensing Object Detection Network with Scene Supervision. *IEEE Trans. Geosci. Remote Sens.* 63, 15.
- Zhou, M., Xing, R., Han, D., Qi, Z. and Li, G., 2024. PDT: Uav Target Detection Dataset for Pests and Diseases Tree, European Conference on Computer Vision.
- Zhu, P., Wen, L., Du, D., Bian, X., Fan, H., Hu, Q., Ling, H., 2022. Detection and Tracking Meet Drones Challenge. *IEEE Trans. Pattern Anal. Mach. Intell.* 44 (11), 7380–7399.
- Zou, Z.X., Chen, K.Y., Shi, Z.W., Guo, Y.H., Ye, J.P., 2023. Object Detection in 20 Years: a Survey. *Proc. IEEE* 111 (3), 257–276.