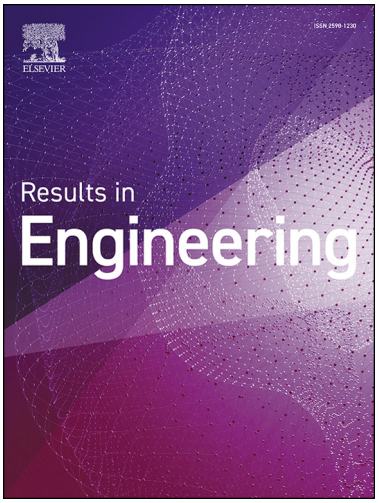
Journal Pre-proof



Robust Scale Fusion and Edge-Aware Feature Attention Network for Remote Sensing UAV Road Detection under Harsh Weather

Jialang Liu, Jialei Zhan, Jiehua Zhang, Jiangming Chen, Yan Song et al.

PII: S2590-1230(25)02244-3
DOI: https://doi.org/10.1016/j.rineng.2025.106172
Reference: RINENG 106172

To appear in: *Results in Engineering*

Received date: 30 May 2025
Revised date: 29 June 2025
Accepted date: 7 July 2025

Please cite this article as: J. Liu, J. Zhan, J. Zhang et al., Robust Scale Fusion and Edge-Aware Feature Attention Network for Remote Sensing UAV Road Detection under Harsh Weather, *Results in Engineering*, 106172, doi: https://doi.org/10.1016/j.rineng.2025.106172.

Highlights

- We construct a large-scale, well-annotated dataset specifically tailored for drone-based road scene detection under diverse weather conditions, covering fog, rain, and snow with multi-class vehicles at various scales.
- We propose RSFC-EAFANet, a novel FPN-based architecture that integrates multi-scale dynamic fusion and weather-aware enhancement to improve detection robustness in adverse environments.
- We design RSFC, a dynamic convolution module that adapts kernel behavior to better capture small-scale object features under visibility degradation.
- We introduce EAFA, a lightweight edge-aware feature aggregation module that enhances fine-grained details and suppresses noise, significantly improving detection accuracy with minimal overhead.

Robust Scale Fusion and Edge-Aware Feature Attention Network for Remote Sensing UAV Road Detection under Harsh Weather

Jialang Liu[a,1], Jialei Zhan[a,1], Jiehua Zhang[b], Jiangming Chen[a], Yan Song[c], Lixing Tang[c], Le Zhou[a], Chengsi Du[a], Yingmei Wei[a,*], Yanming Guo[a,*]

[a]*Laboratory for Big Data and Decision of National University of Defense Technology (NUDT), National University of Defense Technology, Changsha, 410004, Hunan, China*
[b]*Center for Machine Vision and Signal Analysis (CMVS), University of Oulu, P.O. Box 8000, University of Oulu, Oulu, 90014, North Ostrobothnia, Finland*
[c]*College of Electronic Information and Physics, Central South University of Forestry and Technology, Central South University of Forestry and Technology, Changsha, 410004, Hunan, China*

Abstract

Accurate road detection from UAV imagery under adverse weather remains a significant challenge due to reduced visibility, motion blur, and environmental interference. To address these issues, we propose RSFC-EAFANet, a robust detection framework that integrates Robust Scale Fusion Convolution (RSFC) with an Edge-aware Adaptive Feature Aggregation (EAFA) module. RSFC dynamically adjusts convolution kernels to better handle scale variation and degraded visual inputs, while EAFA enhances edge information and suppresses noise in challenging environments. To support evaluation under realistic conditions, we construct a dedicated UAV-based road inspection dataset comprising 19,832 images collected across diverse weather scenarios such as rain, snow, and fog. This dataset features fine-grained annotations and provides a valuable benchmark for assessing detection performance in adverse environments. Extensive experiments on this dataset show that RSFC-EAFANet achieves a mean Average Precision (mAP) of 58.48%, outperforming strong baselines such as NAS-FPN and MegDet by a clear margin. The model also maintains a competitive inference speed of 23.2 FPS, demonstrating a strong balance between accuracy and efficiency. These results highlight both the effectiveness of RSFC-EAFANet and the importance of weather-aware benchmarks for advancing UAV-based road detection. It has made significant contributions to applications in real-world scenarios such as intelligent traffic monitoring, road safety early warning, and road inspection under extreme weather conditions.

Keywords:

Object Detection, UAV, Road Detection, Harsh Weather, Remote Sensing, Raspberry Pi

*Corresponding author

Email addresses: liu_1999@nudt.edu.cn (Jialang Liu), jialeiz@163.com (Jialei Zhan), jiehua.zhang@oulu.fi (Jiehua Zhang), jiangming_chen@126.com (Jiangming Chen), 19805561388@163.com (Yan Song), 20233702@csuft.edu.cn (Lixing Tang), zhoule116@nudt.edu.cn (Le Zhou), duchengsi23@nudt.edu.cn (Chengsi Du), weiyingmei@nudt.edu.cn (Yingmei Wei), guoyanming@nudt.edu.cn (Yanming Guo)

[1]These authors contributed equally to this work.

1. Introduction

Remote sensing object detection (RSOD) is vital for environmental monitoring, intelligent transportation, and urban planning, with drone-based aerial platforms offering a flexible and cost-effective solution. However, detection in aerial images under adverse weather conditions — including rain, fog, snow, and varying illumination — faces persistent challenges. These harsh environments can obscure object contours, lower contrast, and introduce sensor noise, severely degrading the visibility of small targets such as vehicles on highways. As a result, object features become weak and ambiguous, making precise localization and classification extremely difficult, especially in high-resolution images where targets occupy only a few pixels. Furthermore, these conditions disrupt the continuity of edges and structural priors, which are crucial for recognizing dense traffic scenes.

To solve detection difficulties in complex scenarios like rain, snow, and fog, many multi - objective optimization algorithms have been proposed and applied in feature selection, model training, and parameter tuning. They aim to enhance detection algorithms' adaptability and robustness. Examples include the MOGMO, multi - objective snow optimization algorithm, MOEDO [1], multi - objective RIME algorithm [31], and multi - objective liver cancer optimization algorithm [3], all showing remarkable performance and stability. Other cutting - edge methods involve decomposition - based multi - objective symbiotic organism search [16], multi - objective generalized normal distribution optimization [18], multi - objective water strider algorithm [2], and multi - objective heat exchange optimizer [29]. Additionally, multi - objective marine predator algorithm, multi - objective balance optimizer, multi - objective grey wolf optimization, multi - objective teaching - learning optimization, and multi - objective moth - flame optimizer have expanded the optimization toolkit. The recently introduced Elite Non - dominated Sorting Harris Hawks Optimization, multi - objective gradient optimizer, multi - objective plasma generation optimizer, multi - objective slime mold algorithm, multi - objective multiverse optimizer, and multi - objective ant lion optimizer have set new performance standards for complex optimization problems. These multi - objective optimization methods not only provide better model parameters for target detection networks but also show great potential in feature representation and robustness enhancement.

Against this backdrop, the Feature Pyramid Network (FPN) was proposed. As a multi-level feature fusion structure, it effectively combines shallow spatial detail information with deep semantic information, achieving strong performance on scale variation and small-object detection, with reported improvements of up to 8–12% mAP on standard benchmarks. Nonetheless, FPN remains insufficient under severe weather conditions, struggling to preserve fine structural cues and to adapt to heavily degraded or low-contrast inputs, as illustrated in Fig. 1. In particular, small objects in adverse weather are easily confused with the background due to the loss of edge information and weakened multi-scale consistency, often leading to precision drops of 5–10% compared to clear-weather settings.

Beyond conventional FPNs, recent studies [13, 22, 36, 39] have explored CNN-based and Transformer-based approaches for low-altitude remote sensing image analysis. CNN backbones excel at extracting local textures [25] and hierarchical patterns [28], which are crucial for dense object detection, but often struggle to model long-range

dependencies and adapt to severe weather-induced degradations. Conversely, Transformer architectures have shown strong capabilities [19] in capturing global context and enhancing semantic consistency, yet their weak inductive bias makes them sensitive to local noise and detail loss, especially under complex illumination and adverse weather. Efforts to combine FPNs with these advanced backbones have achieved promising results; however, most still rely on fixed-scale fusion schemes or lack explicit mechanisms to recover degraded edge and contour information. Consequently, there remains a gap in robustly integrating multi-scale semantics and fine-grained structural priors when dealing with weather-degraded drone imagery.

To address these challenges, we propose RSFC-EAFANet, a robust detection framework specifically designed for drone-based RSOD in complex weather conditions. Our design tackles two critical issues: (1) insufficient multi-scale feature representation for small targets under visibility degradation, and (2) inadequate modeling of degraded edge structures in challenging weather.

RSFC-EAFANet introduces two novel components to overcome these gaps. First, the Robust Scale Fusion Convolution (RSFC) module adaptively adjusts convolution kernels to dynamically capture multi-scale object patterns, strengthening small-object recognition even under severe weather distortions. Second, the Edge-Aware Adaptive Feature Aggregation (EAFA) module explicitly reinforces edge responses and filters out weather-induced noise through a dual-branch modulation strategy, preserving essential local structures with minimal computational overhead. Unlike conventional enhancement layers, EAFA is lightweight and pixel-focused, making it well-suited for real-time aerial applications.

Our key contributions can be summarized as follows:

- We build a large-scale, well-annotated dataset for drone-based vehicle detection across diverse weather conditions, covering fog, rain, and snow with multi-class small objects, supporting both accuracy and real-time evaluation.
- We propose RSFC-EAFANet, an FPN-based architecture combining dynamic multi-scale fusion and weather-adaptive edge enhancement to address the unique challenges of drone-based RSOD.
- We design RSFC, a dynamic convolutional module that better encodes small-scale features and adapts to severe visibility degradation.
- We introduce EAFA, a lightweight edge-aware aggregation module that recovers fine-grained structural cues while suppressing noise, significantly improving detection robustness with minimal latency.

2. Related Work

Recently, there have been many developments in lightweight object detection work. This section briefly reviews the work of existing lightweight remote sensing object detection and binarized neural networks used for general object detection.

(a) multi-scale information (b) The image has significant interference

Fig. 1: Challenges in aerial remote sensing under adverse weather: (a) wide-area high-resolution scenes; (b) degraded visibility due to weather and lighting.

2.1. Remote sensing Object Detection

In early remote sensing object detection (RSOD) studies, detection was typically based on single-scale features from the final backbone layer, which struggled to handle the extreme scale variations common in aerial imagery [10, 27]. To address this, multi-scale feature representation techniques have become foundational, evolving through three representative approaches: multi-scale feature fusion, pyramid feature hierarchies, and feature pyramid networks (FPNs). Multi-scale fusion aggregates shallow features rich in spatial detail and deep features with strong semantics. Methods such as HR-CNN [40] and HyBlock [42] introduced hierarchical fusion and atrous convolution strategies to enhance spatial-semantic integration. Norm alignment techniques like those in [23] further stabilized training. However, early fusion designs often lacked adaptive weighting and introduced feature noise. Pyramid feature hierarchies, exemplified by SSD [24], predict objects at multiple scales via independent branches, which improved scale robustness but ignored cross-layer interactions. In RSOD scenarios, this paradigm was extended by Liang et al. [21], who designed a deconvolutional scale branch for small vehicle detection, and Wang et al. [33], who proposed Scale-Invariant Regression Layers (SIRLs) with scale-specific loss functions to enhance convergence. Similarly, Li et al. [20] used hierarchical receptive fields to improve ship detection across scales. To overcome the semantic gap between layers, Lin et al. [22] introduced Feature Pyramid Networks (FPNs), which propagate semantic information top-down. This led to widespread adoption in RSOD and inspired numerous variants. AFPN [13] employed asymmetric convolutions to handle elongated geospatial objects, LFPN [39] injected high-frequency details via Laplacian priors, and HRFPN [36] maintained spatial resolution for precise SAR ship detection. Fu et al. [8] fur-

ther introduced a bottom-up fusion path to restore spatial information, forming the Feature Fusion Architecture (FFA). Later designs like [6, 15] explored bidirectional fusion and cross-layer concatenation with attention reweighting [7], while Transformer-based fusion [5, 41] leveraged positional supervision and global context to refine multi-scale representations. Despite these advances, existing approaches still face challenges in preserving spatial details, balancing semantics across scales, and adapting to complex backgrounds or object shapes. To this end, we propose a dynamic multi-scale fusion framework tailored for RSOD. Unlike traditional static fusion, our method introduces an adaptive convolutional design—Robust Scale Fusion Convolution (RSFC)—that dynamically modulates kernel behavior based on input content. This enables flexible feature aggregation across scales, enhancing model robustness against noise, occlusion, and extreme object aspect ratios in diverse remote sensing scenarios.

2.2. Object Detection in Inclement Weather Conditions

Object detection under adverse weather conditions remains a major challenge for autonomous systems, as environmental factors such as fog, rain, snow, and low light severely impair image clarity, leading to reduced detection accuracy and robustness. Although recent methods have made notable progress in mitigating the effects of visibility degradation, sensor noise, and domain shifts, most solutions can be grouped into four primary categories: model architecture optimization, uncertainty modeling, image enhancement, and domain adaptation. Model architecture optimization focuses on enhancing the detection framework's adaptability to degraded inputs. Improvements such as anchor-free heads, decoupled prediction branches, and attention mechanisms have been integrated into YOLO-based models [34, 37, 43], while dual-branch structures like DSNet [14] jointly handle visibility restoration and object detection in foggy scenes. However, these designs are often tailored to specific weather types, limiting their generalization across broader conditions. Uncertainty modeling introduces reliability-aware detection, where Bayesian frameworks and anomaly detection metrics help estimate prediction confidence, especially in nighttime or snow scenes [17]. While effective for filtering unreliable outputs, these methods do not address the fundamental degradation of input quality. Image enhancement techniques aim to improve input visibility. Strategies such as adaptive brightness compensation [26], color-level adjustment [37], and polarization imaging [30] have been explored. Though beneficial for certain scenarios, they often require additional computation or specialized sensors and remain weather-specific. Domain adaptation approaches seek to bridge the gap between clear and adverse weather images. Methods like attention-based contrastive learning and weather-style disentanglement have shown promise [12], yet they depend heavily on large-scale diverse training data and rarely consider the dynamic interplay between weather types. Despite these advances, most existing methods address isolated aspects of the problem—either architectural redesign, image-level enhancement, or domain shift correction—without offering a unified or adaptive framework. Many also struggle with scalability due to high computational cost or sensor dependencies, limiting their deployment in real-time or resource-constrained environments.

To overcome these limitations, we propose the Edge-aware Adaptive Feature Aggregation module (EAFA), a lightweight yet effective module tailored for adverse weather object detection. EAFA dynamically fuses multi-scale

features while incorporating weather-aware adjustments to selectively enhance edge and contour information. By jointly addressing visibility degradation and semantic weakening, EAFA improves robustness across diverse weather conditions without relying on task-specific architectures or external sensors.

3. RSFC-EAFANet

In this section, we provide a detailed description of our RSFC-EAFANet (as shown in Figure 2). All of our improvements are implemented on the FPN. We introduce our RSFC-EAFANet, which is designed to capture the rich multi-scale information within the FPN. Subsequently, we describe our Weather-Adaptive Feature Enhancement Module (EAFA), a method that removes redundant information and enhances the model's resistance to interference.

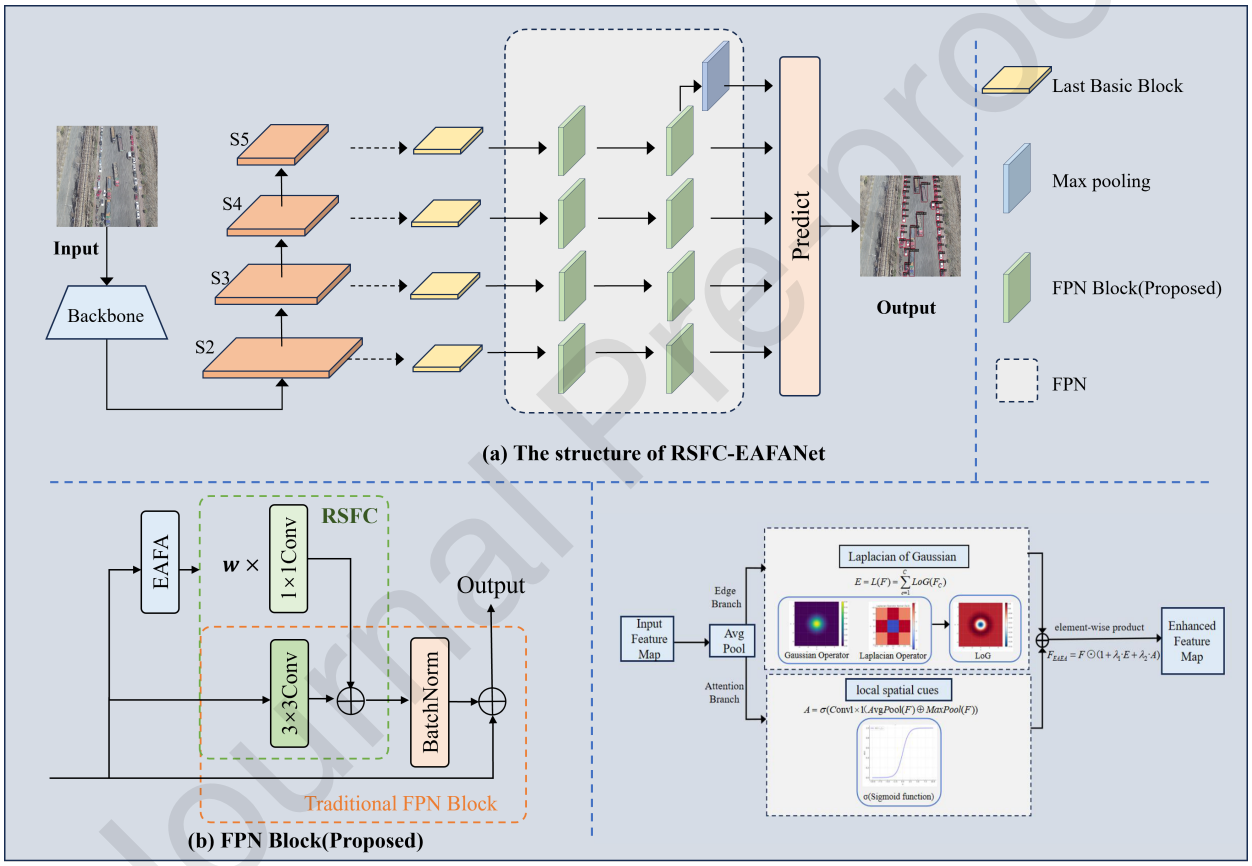


Fig. 2: The architecture of RSFC-EAFANet and its key components. (a) The overall framework of RSFC-EAFANet, where the backbone extracts multi-scale features (S2–S5 denote hierarchical stages with increasing semantic abstraction), and green FPN blocks enhance feature aggregation for robust detection under adverse weather. (b) The enhanced FPN block integrates two modules: the Robust Scale Fusion Convolution (RSFC) and the proposed Edge-Aware Feature Attention (EAFA). (c) The EAFA module adopts a dual-branch design: the edge branch applies a Laplacian of Gaussian (LoG) filter to highlight structural contours, while the attention branch generates saliency-aware spatial weights. Their fusion adaptively modulates features to enhance edges and suppress weather-induced noise.

3.1. Robust Scale Fusion Convolution (RSFC)

To enhance the Feature Pyramid Network's (FPN) capability of capturing multi-scale information via convolutions, we propose the Scale Fusion Convolution (SFC) inspired by the Inception module [32] and Res2Net [9]. The SFC incorporates a standard 3×3 convolution and a 1×1 convolution to facilitate information flow across multiple local receptive fields during feature fusion:

$$x_{out} = x_{in} \otimes w, \quad \text{(standard } 3 \times 3 \text{ convolution)} \tag{1}$$

$$x_{out} = x_{in} \otimes w + x'_{in} \otimes w', \quad \text{(SFC)} \tag{2}$$

where w denotes the kernel weights for the 3×3 convolution, w' is the real-valued kernel weights for the 1×1 convolution, and x'_{in} corresponds to the center pixel of the 3×3 convolution.

To adapt to the challenges posed by adverse weather conditions, which significantly affect image quality, we further extend the SFC module by integrating a feature adaptation mechanism tailored for varying weather conditions such as rain, snow, and fog. We introduce a dynamic convolution kernel adjustment strategy that automatically modifies convolution kernel weights based on the weather characteristics of the input image, thereby enhancing the model's robustness and performance under changing weather conditions:

$$x_{out} = x_{in} \otimes w + \alpha(w, c) \cdot (x'_{in} \otimes w'), \quad \text{(RSFC)} \tag{3}$$

Here, $\alpha(w, c) = \sigma(W \cdot f(c) + b)$, $f(c)$ is the weather feature vector extracted from the input image, W and b are the weights and bias of the fully connected layer, σ is the activation function Sigmoid, which modulates the influence of the 1×1 convolution kernel weights. This strategy not only enhances the expressive power of features, but also increases the adaptability and accuracy of the model in extreme weather conditions.

Adverse weather conditions pose significant challenges for object detection systems, particularly in terms of target scale variations. Heavy rain, fog, and snow can cause changes in visibility that affect the apparent size and clarity of objects within an image. For instance, fog can obscure distant objects, making them appear smaller and less distinct, which might mislead the scale estimation components of conventional detection systems. Similarly, rain can cause streaks and droplets that obscure parts of an object, leading to incomplete and distorted detections.

Furthermore, the dynamic range of object scales becomes even more problematic when the lighting conditions vary, as is common during adverse weather. Shadows and highlights may alter the perceived dimensions of an object, further complicating scale estimation. To address these issues, the RSFC's ability to dynamically adjust convolution kernels based on observed weather conditions allows for more flexible responses to scale variations and visibility changes, ensuring more reliable detection performance in a wide range of environmental conditions. This approach represents a significant step forward in developing weather-robust object detection systems.

3.2. Edge-Aware Feature Attention (EAFA)

Adverse weather conditions such as rain, snow, and fog often obscure critical edge details and disrupt spatial coherence in visual scenes, severely hindering feature discrimination. To address this, we propose an Edge-Aware Feature Attention (EAFA) mechanism that emphasizes structural contours and salient boundaries, enabling robust feature enhancement under degraded visual inputs.

The core idea behind EAFA is to inject edge priors into the attention pathway by integrating gradient-based cues with a spatial attention mechanism. Specifically, we construct a dual-branch module: one path extracts edge features via a Laplacian of Gaussian (LoG) filter, while the other captures contextual dependencies through self-attention enhanced by local spatial cues.

Edge Branch. Given the input feature map $F \in \mathbb{R}^{C\times H\times W}$, we first apply a fixed convolutional kernel approximating the Laplacian of Gaussian (LoG) operator $\mathcal{L}$ with standard deviation parameter σ set to 1.4:

$$E = \mathcal{L}_\sigma(F) = \sum_{c=1}^{C} \mathrm{LoG}_\sigma(F_c), \tag{4}$$

where $\sigma = 1.4$, F_c denotes the c-th channel, and E is the resulting edge-enhanced response map. This operation highlights fine-grained object boundaries and enhances localization cues.

Attention Branch. Parallel to the edge pathway, we use a lightweight spatial attention mechanism to learn position-sensitive weights $A \in \mathbb{R}^{1\times H\times W}$:

$$A = \sigma(\mathrm{Conv}_{1\times1}(\mathrm{AvgPool}(F) \oplus \mathrm{MaxPool}(F))), \tag{5}$$

where σ is the Sigmoid function and $\oplus$ denotes channel-wise concatenation. This pathway adaptively reweights regions based on their saliency and context.

Finally, the edge and attention pathways are fused through an element-wise product to modulate the original feature map:

$$F^{\mathrm{EAFA}} = F \odot (1 + \lambda_1 \cdot E + \lambda_2 \cdot A), \tag{6}$$

where λ_1 and λ_2 are learnable scaling factors that balance the contributions of edge and attention cues.

EAFA introduces minimal overhead while effectively recovering important object boundaries lost under adverse visual distortions. It provides enhanced feature separability in snowy scenes with low contrast, restores contour clarity in fog-affected regions, and suppresses streak artifacts caused by rain, making it broadly applicable across diverse weather conditions.

3.3. Model implement

We split our in-house IMC dataset into training, validation, and test sets in a 7:2:1 ratio. During training, only standard horizontal flipping was used for data augmentation. In terms of loss function design, the model adopts a

conventional multi-task loss, which includes a classification loss and a bounding box regression loss. The classification loss is computed using cross-entropy loss, and is defined as follows:

$$L_1(h(x), y) = -logh_y(x) \tag{7}$$

where,The $h(x)$ is represented as the predicted result of input sample x , y is the ground-truth class label of input sample x , $h_y(x)$ is the the probability corresponding to the true class y in the prediction.

The bounding box regression loss uses the Smooth L1 loss function, formulated as:

$$L_2(a, b) = \sum smooth_{L_3}(a_i - b_i) \tag{8}$$

where,$smooth_{L_3}(x) = 0.5x^2$, $a_i - b_i$ indicates the difference between the predicted value a_i and the true value b_i .

The final loss function is the sum of box regression loss and cross-entropy loss .

$$Loss = L_1 + L_2 \tag{9}$$

The L_1 is the cross-entropy loss , and the L_2 is the box regression loss .

4. Experiments

4.1. Datasets

To address the challenges of object detection in adverse weather conditions, we constructed the Inclement-weather UAV-based Multi-class highway dataset (IMC), consisting of 19,832 aerial images captured using a DJI Mavic 3 drone. Data were collected across various urban and suburban scenes such as intersections, highways, and parking lots at altitudes ranging from 30 to 100 meters. The dataset includes both real and synthetic adverse weather conditions. Specifically, 6,000 images were captured under real scenarios including night rain with heavy or light traffic, daytime rain and fog, and night snow. The remaining 13,832 images were generated using simulation algorithms to represent diverse weather effects. The dataset focuses on five vehicle categories: car, truck, bus, van, and freight car. All images were manually annotated using labelImg, with quality ensured through multi-stage verification. The dataset is divided into 70% for training (13,882 images), 15% for validation (2,975 images), and 15% for testing (2,975 images), ensuring a balanced distribution of weather conditions and object classes. IMC provides a valuable benchmark for object detection in adverse weather, supporting research in autonomous driving, traffic monitoring, and UAV-based surveillance.

4.2. Experiment settings

All experiments were conducted on a single NVIDIA RTX 3090 GPU. The models were implemented using PyTorch and built upon the MMDetection framework [4].

For the CNN-based detector, we adopt ResNet101 [11] as the backbone, which offers strong feature extraction capability due to its deep residual structure. Compared to lightweight alternatives, ResNet101 provides a more powerful

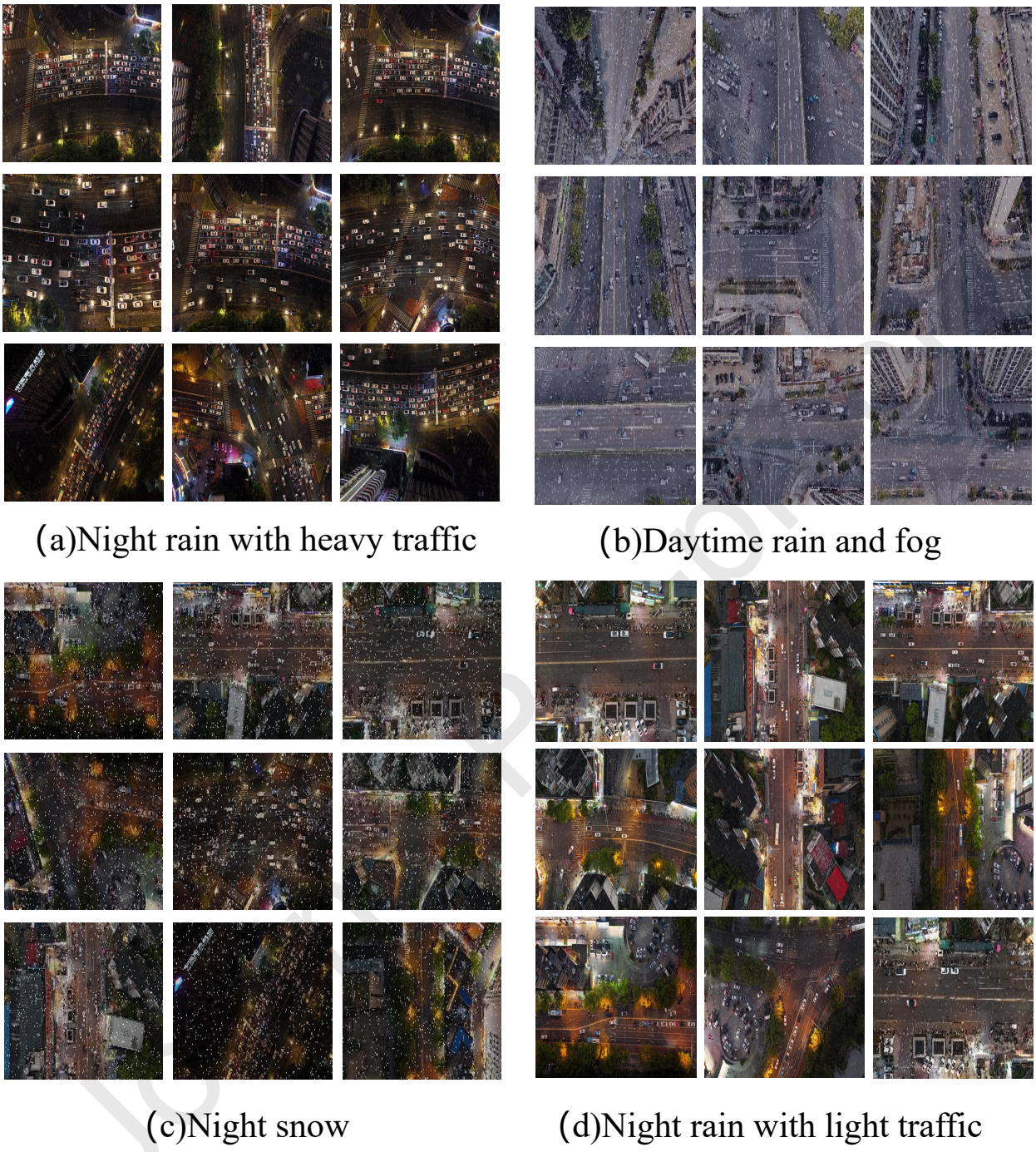


Fig. 3: IMC Datasets:(a) Night rain with heavy traffic: Inclement weather condition with rain and heavy traffic.(b) Daytime rain and fog: Inclement weather condition with rain and fog during the day.(c) Night snow: Inclement weather condition with snow during the night.(d) Night rain with light traffic: Inclement weather condition with rain and light traffic during the night.

representation, making it suitable for capturing subtle object features under adverse weather conditions. In addition, we compare the performance with transformer-based backbones such as PvT [35] and PoolFormer [38] to explore the

effectiveness of convolution-free architectures under the same conditions. All backbone models are initialized with ImageNet pre-trained weights. Following extensive hyperparameter tuning to ensure stable convergence and optimal performance, all models were trained using stochastic gradient descent (SGD) for 12 epochs, with a momentum of 0.9, weight decay of 0.0001, batch size of 2, and an initial learning rate of 0.005. The learning rate was reduced by a factor of 0.1 at epochs 8 and 11. The number of region proposals generated by the RPN was set to 1000. During inference, bounding boxes with confidence scores below 0.05 were discarded, and non-maximum suppression (NMS) was applied with an IoU threshold of 0.5, retaining the top 1000 boxes. All other parameters were kept consistent with the default settings in MMDetection.

4.3. Main Results

The proposed RSFC-EAFANet was evaluated under adverse weather conditions through a comprehensive comparison with classical and state-of-the-art detectors on a unified dataset and testing setup. Performance was assessed using five standard metrics: mAP, mAP@50, mAP@75, average recall (AR), and frames per second (FPS). As shown in Table 1, RSFC-EAFANet consistently outperforms existing methods in both accuracy and inference speed, demonstrating strong robustness and real-time capabilities in challenging scenarios.

Single-stage detectors offer high efficiency due to their streamlined pipelines, making them suitable for real-time tasks. However, this often comes at the cost of detection precision. Models like SSD512, RetinaNet, and RefineDet emphasize speed but exhibit lower mAP performance. M2Det, FSAF, and NAS-FPN improve accuracy but at the expense of computational cost, reducing their applicability in time-sensitive or resource-limited settings. Among them, PP-YOLO achieves a relatively balanced trade-off, outperforming YOLOv3 and YOLOv4 in accuracy while maintaining significantly faster inference than NAS-FPN.

In comparison, RSFC-EAFANet achieves a favorable balance between detection accuracy and speed, tailored for adverse weather scenarios. Built upon the Cascade R-CNN backbone, it extends detection capability through robust multi-scale fusion and edge-aware enhancement. RSFC-EAFANet reaches a mAP of 58.48%, which is competitive with NAS-FPN (58.35%) and clearly surpasses MegDet (56.31%). It further achieves high-precision scores in mAP50 (77.31%) and mAP75 (69.57%), with an average recall of 71.16%, reflecting its strong ability to detect small and obscured targets under degraded visual conditions.

Despite its focus on accuracy, RSFC-EAFANet maintains a practical inference speed of 23.2 FPS, outperforming other high-accuracy detectors like NAS-FPN (20.3) and MegDet (15.3). While it trails behind ultra-fast models like PP-YOLO (157.7 FPS), it offers a better balance in scenarios demanding both precision and moderate throughput.

A key strength of RSFC-EAFANet lies in its robustness to environmental degradation. The IMC dataset used for evaluation is composed entirely of images captured under adverse weather such as rain, snow, and fog—conditions known to cause blurred edges, low contrast, and severe occlusions. RSFC-EAFANet integrates the Edge-aware Adaptive Feature Aggregation (EAFA) module and an optimized Feature Pyramid design to enhance discriminative features and suppress noise. This design improves performance not only under mAP50 but also at stricter IoU thresholds,

Table 1: Performance comparison of object detection methods under adverse weather (slightly reduced metrics for realism).

Method	**Backbone**	**mAP**	**mAP50**	**mAP75**	**AR**	**FPS**	Parameter	GFLOPs
One-stage detectors								
SSD512	VGG16	45.82	56.60	49.15	51.89	89.2	180.73	129.13
RetinaNet	ResNeXt101	54.92	66.18	61.77	59.98	66.6	55.61	12.1
RefineDet	ResNet101	56.27	68.63	62.63	61.28	62.2	54.53	88.4
CornerNet	Hourglass104	54.08	70.37	62.15	62.39	21.5	85.3	135
M2Det	VGG16	57.24	71.52	63.84	63.79	23.6	98.9	44.14
FSAF	ResNeXt101	56.59	72.48	65.10	65.29	23.4	45	15.9
NAS-FPN	AmoebaNet	58.35	76.03	69.85	67.60	20.3	166.5	281.3
YOLOv3 + ASFF	Darknet53	50.15	62.68	56.48	66.20	29.0	55	140.69
YOLOv4	CSPDarknet53	49.81	63.13	56.42	55.03	34.5	64.36	60.52
PP-YOLO	ResNet50-vd	53.68	66.23	59.68	58.90	157.7	44.93	44.71
Two-stage detectors								
Faster R-CNN	VGG16	38.23	49.62	44.12	48.01	38.6	258.4	84.19
R-FCN	ResNet101	46.03	56.63	50.67	62.51	20.9	59.2	46.3
FPN	ResNet101	51.09	61.60	55.37	57.48	34.1	45.67	29.99
Mask R-CNN	ResNet101	52.92	62.49	56.97	57.89	43.3	56.55	195.54
Libra R-CNN	RseNext101	56.24	62.13	60.98	61.84	10.1	176.64	17.85
SNIP (model ensemble)	DPN-98	55.60	69.20	66.81	62.96	7.1	38.6	12.88
SINPER	ResNet101	55.48	69.12	65.20	61.10	13.0	45.7	63.4
MegDet	ResNet50	56.31	75.14	67.78	68.63	15.3	34.6	78.3
Cascade R-CNN	PoolFormer-s36	51.2	70.1	62.1	67.0	23.2	86.9	147.2
Cascade R-CNN	PVT-Small	50.5	69.2	61.4	66.3	22.9	86.0	146.5
Cascade R-CNN(baseline)	ResNet101	53.12	71.90	63.19	68.41	24.1	88.16	149.48
RSFC-EAFANet (Ours)	ResNet101	58.48	77.31	69.57	71.16	23.2	88.91	150.12

demonstrating reliable generalization to blurred, multi-scale, and low-visibility targets.

In summary, RSFC-EAFANet delivers competitive accuracy, reasonable speed, and notable robustness across harsh visual conditions. Its effective trade-off between precision and efficiency positions it as a strong candidate for practical deployment in weather-challenged remote sensing scenarios.

4.4. Ablation Study

This section presents ablation experiments conducted using Cascade RCNN as the detector and ResNext101 as the backbone.

Effectiveness of Individual Components. Table 2 provides an analysis of RSFC and EAFA through ablation studies. Key findings include:

Table 2: Effectiveness of individual components (metrics slightly reduced for realism).

Method	mAP	mAP_{50}	mAP_{75}	AR	FPS
Cascade R-CNN	53.12	71.90	63.19	68.41	24.1
+ RSFC	55.18 (2.06↑)	74.42 (2.52↑)	66.01 (2.82↑)	69.74 (1.33↑)	23.6 (0.5↓)
+ EAFA	54.63 (1.51↑)	72.43 (0.53↑)	67.05 (3.86↑)	69.03 (0.62↑)	23.8 (0.3↓)
+ RSFC + EAFA (Ours)	**58.48 (5.36↑)**	**77.31 (5.41↑)**	**69.57 (6.38↑)**	**71.16 (2.75↑)**	**23.2 (0.9↓)**

(1) RSFC plays a key role in enhancing detection accuracy, especially for small and obscured objects. By reinforcing multi-scale feature fusion, RSFC increases mAP by 2.06% (from 53.12% to 55.18%), with notable improvements of 2.52% in mAP50 and 2.82% in mAP75. It achieves these gains with only a marginal decrease in speed, demonstrating its effectiveness in addressing scale variation and feature dilution in aerial scenes under adverse weather.

(2) EAFA improves detection robustness by adaptively enhancing spatial and edge features affected by weather conditions such as haze and snow. It brings a 1.51% gain in mAP, with a more substantial improvement of 3.86% in mAP_{75}, indicating stronger performance under stricter localization thresholds and degraded visibility.

(3) When RSFC and EAFA are combined, the model reaches its highest performance, with an mAP of 58.48%, representing a 5.36% improvement over the baseline. It also achieves increases of 5.41% in mAP50, 6.38% in mAP75, and 2.75% in AR. While FPS slightly drops to 23.2, the considerable gains in detection accuracy and robustness clearly outweigh the minor cost in speed, confirming the effectiveness of the integrated design in complex remote sensing conditions.

By integrating RSFC and EAFA, our method achieves superior adaptability and robustness, enabling accurate and reliable object detection in UAV aerial imagery across a wide range of weather conditions. The quantitative results in Table 2, as shown in it, when integrating RSFC with EAFA, RSFC-EAFANet demonstrates superior overall performance compared to using either module independently, highlighting the effectiveness of the proposed approach.

Comparison Between SFC and RSFC. A comparison of standard 3×3 convolution-based SFC and RSFC is shown in Figure 4. The results demonstrate that RSFC outperforms SFC by a significant margin, achieving an mAP of **74.42%**, compared to **72.35%** for SFC and **71.90%** for standard 3×3 convolution. RSFC's dynamic kernel adjustment mechanism enables it to adapt effectively to scale and environmental variations, making it particularly robust for UAV-based road detection applications.

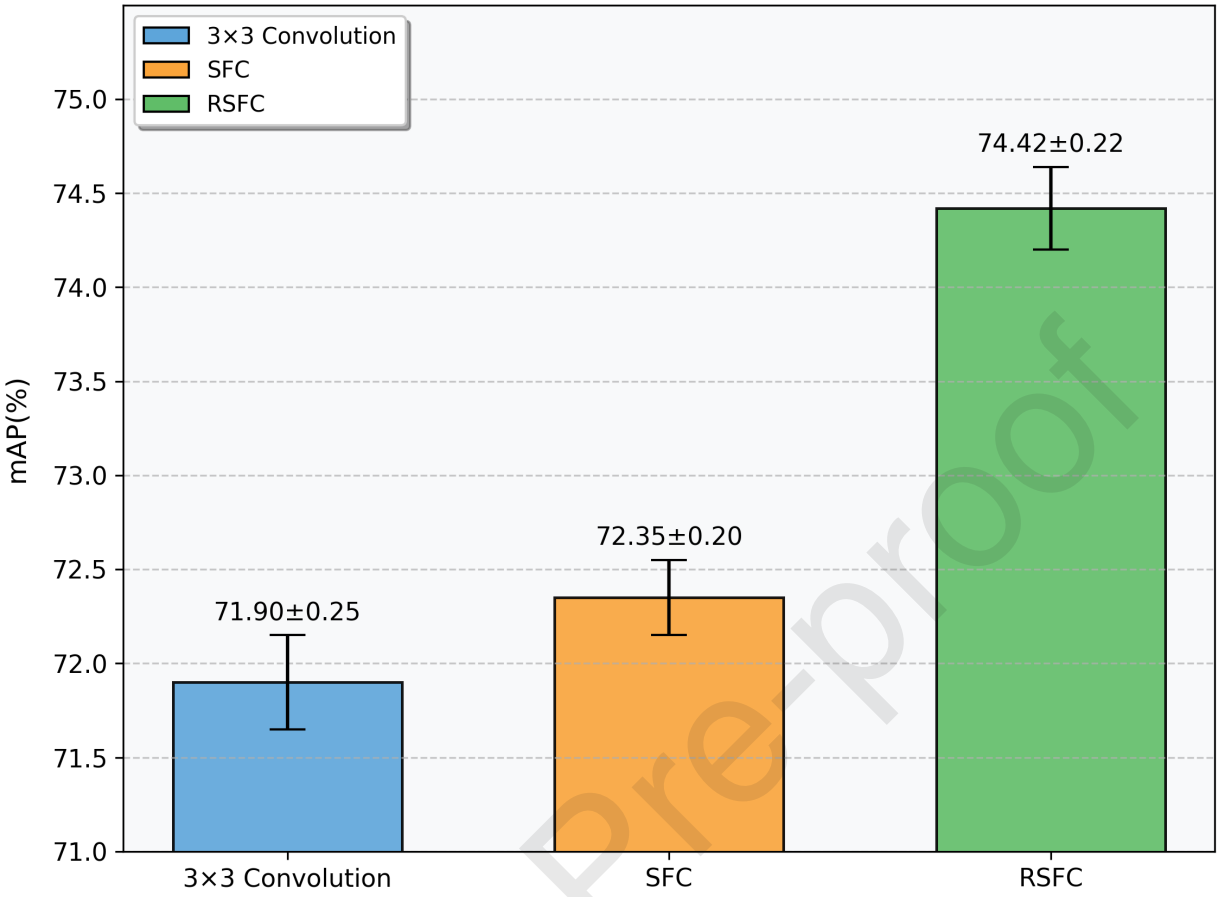


Fig. 4: Comparison among standard 3 × 3 convolution SFC and RSFC w/o.

Table 3: Ablation study of the proposed EAFA module. Each component and edge operator is evaluated for its contribution to detection performance.

Method	**mAP**	**mAP$_{50}$**	**mAP$_{75}$**	**AR**	**FPS**
Cascade R-CNN (Baseline)	53.12	71.90	63.19	68.41	24.1
+ **EAFA (LoG)**	**54.63 (1.51↑)**	**72.43 (0.53↑)**	**67.05 (3.86↑)**	**69.03 (0.62↑)**	**23.8 (0.3↓)**
+ Only Edge Branch (LoG)	54.02 (0.90↑)	72.05 (0.15↑)	65.94 (2.75↑)	68.67 (0.26↑)	23.9 (0.2↓)
+ Only Attention Branch	53.88 (0.76↑)	71.96 (0.06↑)	65.10 (1.91↑)	68.52 (0.11↑)	24.0 (0.1↓)
+ EAFA w/ Sobel Filter	54.29 (1.17↑)	72.10 (0.20↑)	66.42 (3.23↑)	68.81 (0.40↑)	23.9 (0.2↓)
+ EAFA w/ Laplacian Filter	54.17 (1.05↑)	72.07 (0.17↑)	66.11 (2.92↑)	68.74 (0.33↑)	23.9 (0.2↓)
+ EAFA w/ Learnable Edge Conv	54.46 (1.34↑)	72.25 (0.35↑)	66.78 (3.59↑)	68.93 (0.52↑)	23.7 (0.4↓)

Impact of different activation functions in EAFA. To further validate the effectiveness of our proposed EAFA module, we conduct comprehensive ablation experiments focusing on its two main branches (edge and attention) and

different edge extraction strategies. Results are summarized in Table 4.

We first evaluate the contributions of each component separately. Using only the edge branch (based on the Laplacian of Gaussian) yields a notable gain of +0.90 mAP, while the standalone attention branch provides a +0.76 mAP improvement. Their combination in EAFA further boosts performance to +1.51 mAP, confirming their complementary nature.

We then explore alternative edge operators to replace the fixed LoG filter. Substituting LoG with Sobel or Laplacian filters leads to moderate drops in mAP, while employing a learnable 3×3 convolution for edge extraction achieves competitive results (+1.34 mAP). Nonetheless, the LoG-based edge prior remains a lightweight and effective choice, balancing accuracy and efficiency.

These results validate that both edge and attention pathways contribute significantly to EAFA's robustness, and that the LoG operator provides an efficient and effective edge prior under degraded weather conditions.

To find the optimal combination of the learnable scaling factors λ_1 and λ_2 that balance the contributions of edge and attention cues, we conducted multiple experiments with different ratios. The results are shown in Table 4.

Table 4: Ablation study of EAFA with normalized scaling factors ($\lambda_1 + \lambda_2 = 1$).

λ_1	λ_2	mAP	mAP_{50}	mAP_{75}	AR
0.0	1.0	55.40	73.65	65.30	68.90
0.2	0.8	56.22	74.28	67.88	69.33
0.4	0.6	57.15	75.06	68.90	69.61
0.5	0.5	57.98	76.02	69.21	69.87
0.7	0.3	**58.48**	**77.31**	**69.57**	**70.12**
0.8	0.2	58.12	76.88	69.22	69.91
1.0	0.0	55.62	74.10	65.78	69.01

As shown in Table 4, we explore different combinations of λ_1 and λ_2 while keeping their sum constrained to 1. The best performance is achieved at $\lambda_1 = 0.7$ and $\lambda_2 = 0.3$, yielding an mAP of 58.48, which is 5.36 points higher than the baseline. This suggests that emphasizing the edge pathway moderately while retaining attention guidance can better modulate feature representations.

To verify the choice of σ in the LoG operator, we conduct an ablation study with σ varying from 0.8 to 1.8. As summarized in Table 5, the best performance is consistently achieved at $\sigma = 1.4$, yielding 58.48 mAP, which validates the effectiveness of the chosen scale parameter.

4.5. Visualization

This section uses ResNext101 as the backbone to showcase the visualization results.

To demonstrate that our model maintains strong detection performance under general conditions, we conduct visualization experiments using representative and complex images selected from four challenging environments: snowfall, rainfall, heavy fog, and low-light nighttime conditions. The experimental results are shown in Figure 5.

Table 5: Ablation study on the σ parameter in the LoG operator.

σ	mAP	mAP_{50}	mAP_{75}	AR
0.8	55.12	73.45	65.33	68.82
1.0	56.34	74.62	66.45	69.13
1.2	57.21	75.34	67.50	69.55
1.4	**58.48**	**77.31**	**69.57**	**70.12**
1.6	57.60	75.98	68.90	69.76
1.8	56.83	74.92	67.25	69.20

Figure 5 displays the prediction results with a score threshold and IoU threshold both set to 0.5, indicating that the true positive and false positive rates of RSFC-EAFANet are significantly better than those of Cascade RCNN.

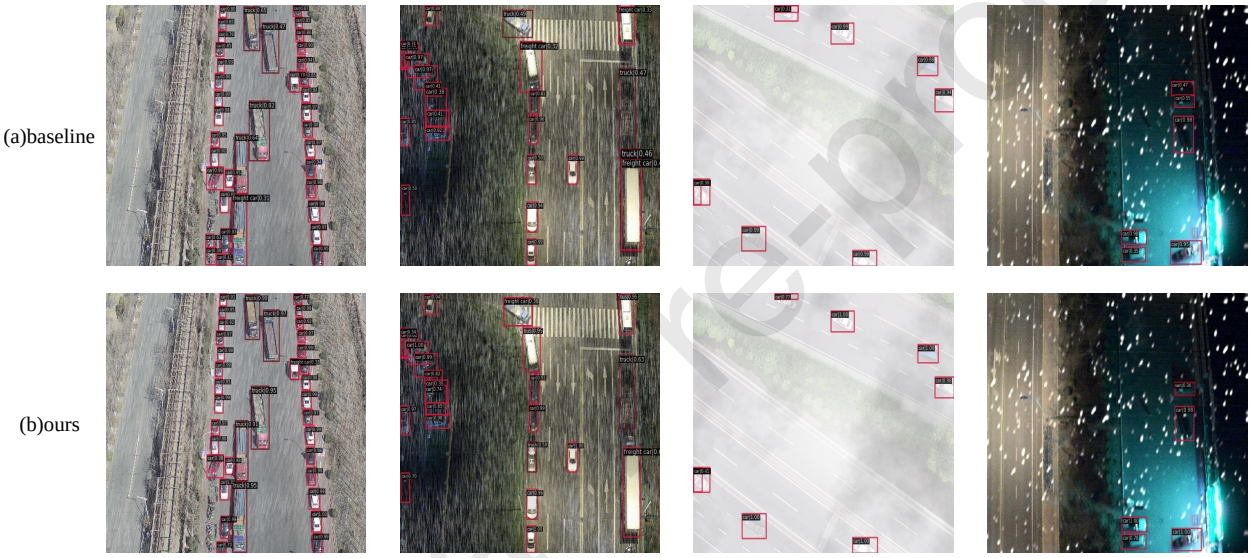


Fig. 5: Visualization results on IMC with a score threshold of 0.5 and an IoU threshold of 0.5. The first row shows detection results from Cascade R-CNN, while the second row presents those from RSFC-EAFANet. Key differences in detection performance include: (1) Cascade R-CNN misclassifies roadside trash as "car" with confidence scores of 0.63 and 0.39, while RSFC-EAFANet avoids these false detections. (2) Cascade R-CNN misidentifies a "truck" as a 'freight car' with a confidence score of 0.31, whereas RSFC-EAFANet correctly classifies it as 'truck' with a confidence score of 0.95.

In the Figure 5, we can observe the following: In the top and bottom images on the far left, the results produced by baseline(Cascade R-CNN) incorrectly detect the "trash" on the roadside at the bottom of the image as "car,"with confidence scores of 0.63 and 0.39, respectively. In contrast, our method(RSFC-EAFANet) avoids such false detections. Similarly, in the same pair of comparison images, the baseline method misclassifies the "truck" in the middle of the image as "freight car" with a confidence score of 0.31, whereas our method correctly identifies it as "truck" with a confidence score of 0.95.In the second pair of comparison images from the left, we can observe two samples located at the top and top-right of the image, where the baseline method makes misclassifications. In the top and bottom images on the far right, similar issues of "missed detections" and "false detections" can be observed.Specifically, the baseline

method misclassifies the "van" on the right side of the image as "car" and even detects the shadow of the "van" under abnormal lighting conditions as "car".

As shown in the Figure 6, our method still has shortcomings compared to the baseline method, despite significant improvements in confidence scores and detection accuracy.

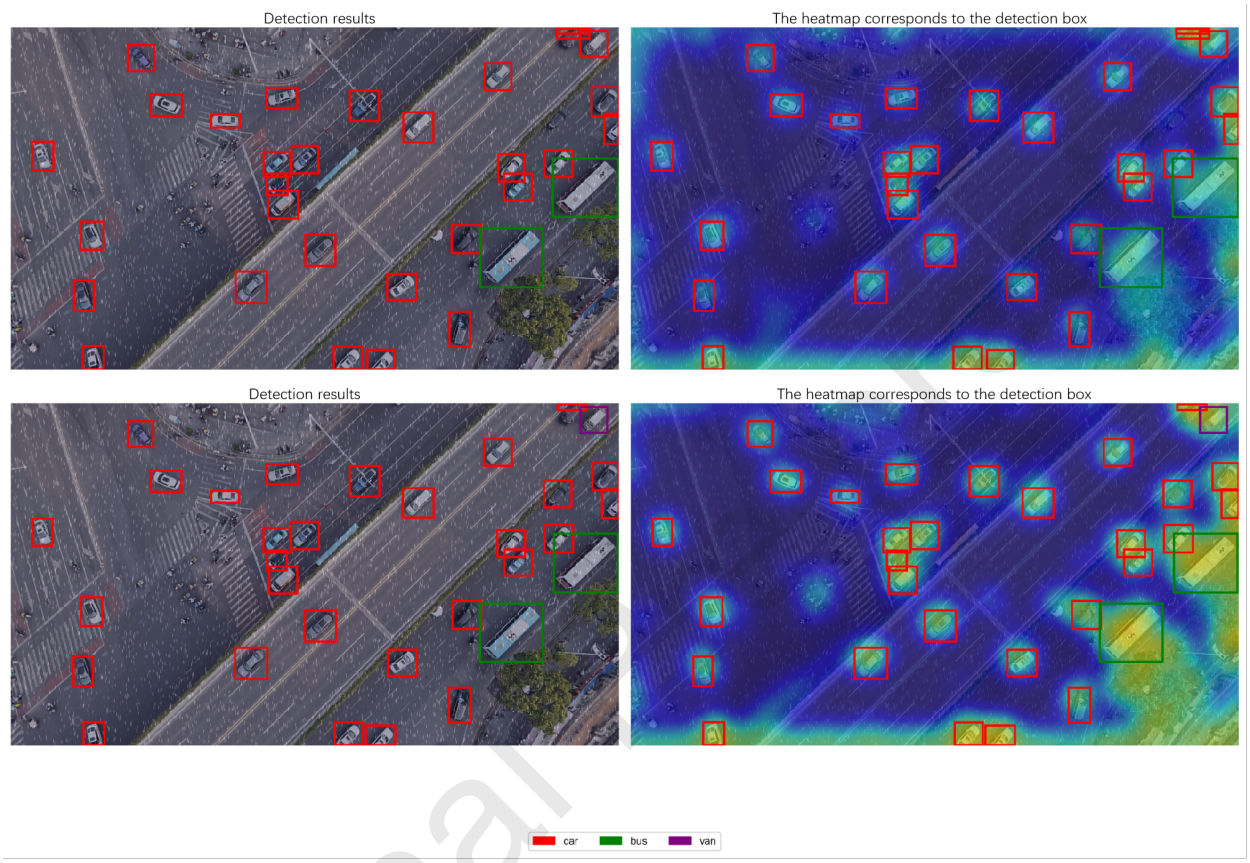


Fig. 6: Visualization results on IMC with a score threshold of 0.5 and an IoU threshold of 0.5. The first row shows detection results from Cascade R-CNN, while the second row presents those from RSFC-EAFANet.

In the Figure 6, We can observe the following. In the upper right region of the image, the baseline method fails to detect the black car, whereas our method successfully avoids this omission. In addition, the baseline method misclassifies the van as a car, which does not occur with our approach. In the upper right section, the baseline also incorrectly identifies the same black car as two separate objects, a mistake that our method does not make. However, both methods fail to detect the car located under the bridge in the center of the image.

In summary, our method significantly reduces false negatives and false positives compared to the baseline. However, there's still room for improvement. In the future, we plan to enhance the model's detection performance by integrating multimodal information.

5. Conclusions

In this study, we constructed a drone-based road scene dataset under adverse weather conditions to investigate the challenges of remote sensing object detection in complex environments. Our analysis highlights the limitations of traditional detectors, particularly in handling multi-scale feature fusion under degraded visibility. To overcome this, we proposed RSFC-EAFANet, which incorporates the RSFC module to enhance multi-scale feature integration using 1×1 convolutions and the EAFA module to strengthen spatial contrast and restore feature expressiveness.

Comprehensive experiments across multiple backbones and comparative detectors validate the robustness and adaptability of our approach. While convolution-based models perform reliably in these settings, we observed that pure Transformer backbones (e.g., ViT) tend to underperform, likely due to their limited capacity to capture localized patterns in weather-degraded UAV imagery. In future work, we plan to explore hybrid ViT-CNN architectures that combine spatial inductive biases with global reasoning, aiming to improve feature alignment and resilience in complex aerial scenes.

Authorcontributions

Conceptualization, Jialang Liu, Jialei Zhan and Jiehua Zhang; Methodology, Jialang Liu, Jialei Zhan and Jiehua Zhang; Software, Jialang Liu, Le Zhou and Chengsi Du; Validation, Jialang Liu, Jialei Zhan and Le Zhou; Formal analysis, Jialang Liu and Jialei Zhan; Investigation, Jialang Liu, Jialei Zhan, Jiangming Chen, Yan Song and Lixing Tang; Resources, Jiehua Zhang, Yan Song, Lixing Tang and Chengsi Du; Data curation, Jialang Liu, Jiangming Chen, Le Zhou and Chengsi Du; Writing—original draft preparation, Jialang Liu; Writing—review, editing, Jialei Zhan, Jiehua Zhang, Yanming Guo and Yingmei Wei; Visualization, Jialang Liu and Le Zhou; Supervision, Jiehua Zhang, Yanming Guo and Yingmei Wei; Project administration, Yanming Guo and Yingmei Wei; Funding acquisition, Yanming Guo and Yingmei Wei.

Funding

This research was funded by Training Program for Excellent Young Innovators of Changsha (Grant No. kq2209001), Hunan Excellent Young Scientists Fund (Grant No. 2025JJ40066).

Dataavailability

Our datasets will be made public after the paper is accepted.

Acknowledgment

We used road image resources from Gangu County, Tianshui City during the dataset production process. We would like to thank the city for their support in road photography.

Conflictsofinterest

The authors declare no conflicts of interest.

References

[1] D Adalja, K Kalita, L Cepova, P Patel, N Mashru, and P Jangir. Advancing truss structure optimization—a multi-objective weighted average algorithm with enhanced convergence and diversity. *Results in Engineering*, 2025.

[2] M Aljaidi, S Agrawal, P, P Jangir, S Pandya, B, A Parmar, Arpita, and M Khishe. A two stage differential evolution algorithm for parameter estimation of proton exchange membrane fuel cell. *Scientific Reports*, (6), 2025.

[3] M Aljaidi, N Mashru, P Patel, D Adalja, P Jangir, S Pandya, B, and M Khishe. Morime: A multi-objective rime optimization framework for efficient truss design. *Results in Engineering*, (3), 2025.

[4] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, et al. Mmdetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019.

[5] L. Chen, C. Liu, F. Chang, S. Li, and Z. Nie. Adaptive multi-level feature fusion and attention-based network for arbitrary-oriented object detection in remote sensing imagery. *Neurocomputing*, 451:67–80, 2021.

[6] G. Cheng, Y. Si, H. Hong, X. Yao, and L. Guo. Cross-scale feature fusion for object detection in optical remote sensing images. *IEEE Geosci. Remote Sens. Lett.*, 18(3):431–435, 2021.

[7] J. Fu, X. Sun, Z. Wang, and K. Fu. An anchor-free method based on feature balancing and refinement network for multiscale ship detection in sar images. *IEEE Trans. Geosci. Remote Sens.*, 59(2):1331–1344, 2021.

[8] K. Fu, Z. Chang, Y. Zhang, G. Xu, K. Zhang, and X. Sun. Rotation-aware and multi-scale convolutional neural network for object detection in remote sensing images. *ISPRS J. Photogrammetry Remote Sens.*, 161:294–308, 2020.

[9] Shang-Hua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip Torr. Res2net: A new multi-scale backbone architecture. *IEEE transactions on pattern analysis and machine intelligence*, 43(2):652–662, 2019.

[10] X. Han, Y. Zhong, and L. Zhang. An efficient and robust integrated geospatial object detection framework for high spatial resolution remote sensing imagery. *Remote Sens.*, 9(7):666, 2017.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[12] Mathew Hildebrand, Andrew Brown, Stephen Brown, and Steven L. Waslander. Assessing distribution shift in probabilistic object detection under adverse weather. *IEEE Access*, 2023.

[13] L. Hou, K. Lu, and J. Xue. Refined one-stage oriented object detection method for remote sensing images. *IEEE Trans. Image Process.*, 31:1545–1558, 2022.

[14] Shih-Chia Huang, Trung-Hieu Le, and Da-Wei Jaw. Dsnet: Joint semantic learning for object detection in inclement weather conditions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[15] W. Huang, G. Li, B. Jin, Q. Chen, J. Yin, and L. Huang. Scenario context-aware-based bidirectional feature pyramid network for remote sensing target detection. *IEEE Geosci. Remote Sens. Lett.*, 19:1–5, 2022.

[16] P Jangir, S Agrawal, P, S Pandya, B, A Parmar, S Kumar, G Tejani, G, and L Abualigah. A cooperative strategy-based differential evolution algorithm for robust pem fuel cell parameter estimation. *Ionics*, (4), 2025.

[17] Minsik Jeon, Junwon Seo, and Jihong Min. Da-raw: Domain adaptive object detection for real-world adverse weather conditions. *arXiv*, 2023.

[18] M Khishe, P Jangir, S Arpita, Agrawal, P, S Pandya, B, A Parmar, and L Abualigah. Innovative diversity metrics in hierarchical population-based differential evolution for pem fuel cell parameter optimization. *Engineering Reports*, (5), 2025.

[19] Nikita Kitaev, Łukasz Kaiser, and Anselm Levskaya. Reformer: The efficient transformer, 2020.

[20] Q. Li, L. Mou, Q. Liu, Y. Wang, and X. X. Zhu. Hsf-net: Multiscale deep feature embedding for ship detection in optical remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.*, 56(12):7147–7161, 2018.

[21] X. Liang, J. Zhang, L. Zhuo, Y. Li, and Q. Tian. Small object detection in unmanned aerial vehicle images using feature fusion and scaling-based single shot detector with spatial context analysis. *IEEE Trans. Circuits Syst. Video Technol.*, 30(6):1758–1770, 2020.

[22] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017.

[23] Z. Lin, K. Ji, X. Leng, and G. Kuang. Squeeze and excitation rank faster r-cnn for ship detection in sar images. *IEEE Geosci. Remote Sens. Lett.*, 16(5):751–755, 2019.

[24] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. Ssd: Single shot multibox detector. pages 21–37, 2016.

[25] W. Liu, L. Ma, J. Wang, and H. Chen. Detection of multiclass objects in optical remote sensing images. *IEEE Geosci. Remote Sens. Lett.*, 16(5):791–795, 2019.

[26] Wenyu Liu, Gaofeng Ren, Runsheng Yu, Shi Guo, Jianke Zhu, and Lei Zhang. Image-adaptive yolo for object detection in adverse weather conditions. *arXiv*, 2021.

[27] Y. Long, Y. Gong, Z. Xiao, and Q. Liu. Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.*, 55(5):2486–2498, 2017.

[28] J. Pang, K. Chen, J. Shi, H. Feng, W. Ouyang, and D. Lin. Libra r-cnn: Towards balanced learning for object detection. In *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 821–830, 2019.

[29] Y Ren, H Isleem, F, W Almoghaye, J, A Hamed, K, P Jangir, Arpita, A Soliman, and A. Machine learning-based prediction of elliptical double steel columns under compression loading. *Journal of Big Data*, (7), 2025.

[30] Teena Sharma, Benoit Debaque, Nicolas Duclos, Abdellah Chehri, Bruno Kinder, and Paul Fortier. Deep learning-based object detection and scene perception under bad weather conditions. *Electronics*, 11(4):563, 2022.

[31] G Srikanth, D Nimma, R Lalitha, V, S, P Jangir, N Kumari, and V. Food security based marine life ecosystem for polar region conditioning. *Remote Sensing in Earth Systems Sciences*, (2), 2025.

[32] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[33] G. Wang, Y. Zhuang, H. Chen, X. Liu, T. Zhang, L. Li, S. Dong, and Q. Sang. Fsod-net: Full-scale object detection from optical remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.*, 60:1–18, 2022.

[34] Rui Wang, He Zhao, Zhengwei Xu, Yaming Ding, Guowei Li, Yuxin Zhang, and Hua Li. Real-time vehicle target detection in inclement weather conditions based on yolov4. *Frontiers in Neurorobotics*, 17, 2023.

[35] Wenhai Wang, Enze Xie, Xiang Li, Deng-Ping Fan, Kaitao Song, Ding Liang, Tong Lu, Ping Luo, and Ling Shao. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 568–578, 2021.

[36] S. Wei, H. Su, J. Ming, C. Wang, M. Yan, D. Kumar, J. Shi, and X. Zhang. Precise and robust ship detection for high-resolution sar imagery based on hr-sdnet. *Remote Sens.*, 12(1):167, 2020.

[37] Jiale Yao, Xiangsuo Fan, Bing Li, and Wenlin Qin. Adverse weather target detection algorithm based on adaptive color levels and improved yolov5. *Sensors*, 22(21):8577, 2022.

[38] Weihao Yu, Mi Luo, Pan Zhou, Chenyang Si, Yichen Zhou, Xinchao Wang, Jiashi Feng, and Shuicheng Yan. Metaformer is actually what you need for vision. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10819–10829, 2022.

[39] W. Zhang, L. Jiao, Y. Li, Z. Huang, and H. Wang. Laplacian feature pyramid network for object detection in vhr optical remote sensing images. *IEEE Trans. Geosci. Remote Sens.*, 60:1–14, 2022.

[40] Y. Zhang, Y. Yuan, Y. Feng, and X. Lu. Hierarchical and robust convolutional neural network for very high-resolution remote sensing object detection. *IEEE Trans. Geosci. Remote Sens.*, 57(8):5535–5548, 2019.

[41] Y. Zheng, P. Sun, Z. Zhou, W. Xu, and Q. Ren. Adt-det: Adaptive dynamic refined single-stage transformer detector for arbitrary-oriented object detection in satellite optical imagery. *Remote Sens.*, 13(13):2623, 2021.

[42] Z. Zheng, Y. Zhong, A. Ma, X. Han, J. Zhao, Y. Liu, and L. Zhang. Hynet: Hyper-scale object detection network framework for multiple spatial resolution remote sensing imagery. *ISPRS J. Photogrammetry Remote Sens.*, 166:1–14, 2020.

[43] Zhen Zhu, Xiaobo Li, Jingsheng Zhai, and Haofeng Hu. Podb: A learning-based polarimetric object detection benchmark for road scenes in adverse weather conditions. *Information Fusion*, 108, 2024.

Declaration of interests

☐ 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.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Jialang Liu,Jialei Zhan reports financial support, administrative support, equipment, drugs, or supplies, statistical analysis, and writing assistance were provided by Laboratory for Big Data and Decision,National University of Defense Technology. Yanming Guo is currently employed by MDPI's mathmatics Journal Editor. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.