Theory and Practice of Science and Technology
2026, VOL. 7, NO. 1, 68-72
DOI: 10.47297/taposatWSP2633-456914.20260701

Research on Robust Multi-Target Detection for Intelligent Driving under Complex Weather Conditions

Jiapeng Xuan

Data Science and Big Data technology, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, 215123, China

ABSTRACT
In intelligent driving systems, the visual perception module often suffers from significant performance degradation under complex weather conditions, which poses a serious challenge for all-weather deployment. Most existing studies concentrate on either image restoration or model optimization tailored to specific weather phenomena such as fog or rain, yet they frequently lack generalization ability when confronted with variable or mixed severe weather. This paper introduces a novel task-driven framework designed for robust perception. The core concept is to treat complex weather not as noise to be removed, but as an inherent environmental attribute that the model should learn to adapt to. Methodologically, we propose a task-oriented weather-invariant feature learning module, integrated with a dynamically weighted multi-modal fusion mechanism. This enables the learning of robust cross-weather domain representations and adaptive information complementarity directly at the feature level. Comprehensive experiments conducted on multiple complex weather datasets—including BDD100K, ACDC, and nuScenes—show that our approach substantially outperforms mainstream baseline methods in terms of mean Average Precision (mAP) and F1 score under diverse conditions like fog, rain, and snow. Thus, it provides a practical technical pathway toward achieving highly reliable all-weather perception for intelligent driving systems.

KEYWORDS
Intelligent driving; Complex weather; Object detection; Domain adaptation; Multi-modal fusion; Robust perception

1 Introduction

The safe and reliable deployment of autonomous driving technology depends heavily on its ability to perform consistently across a wide range of real-world environments [1-2]. Environmental perception, serving as the foundation of decision-making systems, directly influences the safety boundaries of vehicles [3]. However, adverse weather conditions—such as rain, fog, and snow—can severely compromise the quality of data captured by sensors like cameras and LiDAR, leading to notable declines in object detection accuracy and recall rates.Current research aimed at improving perception robustness in such conditions generally follows two main approaches: first, image enhancement techniques based on physical models or deep learning, which attempt to restore the visual quality of degraded images during preprocessing (e.g., dehazing and deraining [4-5]); second, transfer learning or domain adaptation methods that aim to reduce the data distribution gap between different weather domains, allowing models to better adapt to target conditions [6]. Despite these efforts, several limitations persist: The "restore-then-detect" pipeline often creates a misalignment between restoration objectives (e.g., PSNR/SSIM) and detection objectives (e.g., mAP), potentially corrupting features critical for detection [7]; most studies focus on a single sensor modality (particularly cameras), failing to exploit the complementary strengths of multi-modal data such as LiDAR in harsh weather [8]; and models designed for specific weather types (e.g., fog-only) struggle to handle the dynamic and mixed conditions encountered in real-world driving, showing limited generalization.

To tackle these challenges, this paper advocates a shift in perspective: rather than treating complex weather as external interference to be eliminated, we consider it an intrinsic environmental condition that the model must learn and adapt to. Consequently, the research focus moves from "how to obtain clear images" to "how to directly learn robust, task-effective features from degraded data." Building on this insight, we develop a unified, end-to-end multi-target detection framework with the following key contributions:

(1) Propose a task-oriented weather-invariant feature learning method: By designing a mechanism that jointly optimizes detection loss and domain discrimination loss, the model is guided to learn essential features that are insensitive to weather changes but critical for target discrimination, avoiding potential feature semantic distortion in traditional domain adaptation.

(2) Design a dynamically adaptive multi-modal fusion architecture: A lightweight real-time modality quality assessment module is introduced, which dynamically adjusts fusion weights based on the instantaneous signal-to-noise ratio of each sensor (e.g., RGB camera, LiDAR), thereby maintaining the overall perceptual robustness of the system even when any sensor performance degrades.

(3) Achieve validation of a unified general detection framework: Systematic experiments on multiple public complex

* **Corresponding Author:** Jiapeng Xuan, Xuan23@student.xjtlu.edu.cn"

weather datasets validate that the proposed single model exhibits superior generalization performance and stable detection accuracy for both known and unknown adverse weather conditions, providing a feasible solution for engineering practicality.

2 Related Work

2.1 Visual Perception under Complex Weather

Early work primarily relied on physical priors like the atmospheric scattering model for image restoration . Against the backdrop of deep learning's ongoing maturation, methodologies anchored in Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) have made remarkable headway in tasks targeting the elimination of single-category weather disturbances [9]. However, these methods are typically designed for specific degradation types, and the improvement in restoration quality does not always positively correlate with the performance of downstream tasks [10]. In recent years, researchers have begun constructing large-scale complex weather datasets, such as Foggy Driving , ACDC , and WeatherProof , to support more robust model training and evaluation [11]. Simultaneously, using synthetic data (e.g., rendering good weather images into bad weather) for data augmentation has also become a common strategy to enhance model generalization [12].

2.2 Domain Adaptation and Domain Generalization

Domain adaptation aims to transfer knowledge from a model trained on a source domain (e.g., clear weather) to a target domain (e.g., foggy weather). Mainstream methods include discrepancy minimization-based methods (e.g., MMD, CORAL) , adversarial learning-based methods, and reconstruction-based methods [13-14]. In object detection tasks, methods like DA-Faster R-CNN introduce domain classifiers to align image-level or instance-level features [15]. However, these methods often optimize domain alignment and object detection tasks independently, potentially causing the feature representation to favor domain discrimination at the expense of detection discriminability [16]. Domain Generalization (DG) pursues good performance on unseen target domains during training. Related methods such as feature normalization and meta-learning provide insights for designing weather-invariant general models in this research [17].

2.3 Multi-Modal Sensor Fusion

Multi-modal fusion is a key technology for enhancing the robustness of autonomous driving systems. Fusion levels can be categorized into data-level, feature-level, and decision-level [18]. PointPainting and PointFusion are representative early works that fuse image semantic information with point cloud geometric information [19-20]. MVX-Net utilizes CNNs to process both images and point clouds. Recently, Transformer-based fusion architectures (e.g., TransFusion) have garnered attention due to their powerful feature interaction capabilities [21-22]. However, most fusion strategies employ fixed weights or static attention-based fusion, lacking adaptability to dynamic sensor degradation. A few works have begun exploring uncertainty-based dynamic fusion or fusion strategies specifically for adverse weather , which is precisely the direction this research aims to advance.

3 Methodology

3.1 Overall Framework Overview

The proposed robust multi-target detection framework is illustrated in Figure 1. It is essentially a dual-branch encoder-decoder structure processing two modalities: RGB images and LiDAR point clouds (or radar data). The framework core comprises two innovative modules: the Task-oriented Weather-Invariant Feature Learning (TIFL) module and the Dynamically Weighted Multi-modal Fusion (DWFM) module . The model takes raw data from front-end sensors as input, undergoes feature extraction, cross-weather domain feature alignment, and adaptive modality fusion, and finally outputs the category and location of target objects via a detection head.

3.2 Task-Oriented Weather-Invariant Feature Learning

Let the source domain D_s be clear weather data, and the target domain D_t be various complex weather data. We construct a shared-weight backbone feature extraction network G_f. To learn weather-invariant features, we introduce a domain discriminator G_d and employ adversarial training to make the features produced by G_f indistinguishable by G_d regarding their source domain. Different from traditional adversarial domain adaptation, we propose a task-oriented joint loss function :

$$C_{total} = C_{det} + \alpha C_{adv} + \beta C_{tal} \tag{1}$$

Where:

$C_{det}$is the standard object detection loss, including classification loss $C_{cls}$and bounding box regression lossC_{reg}.

$C_{adv}$is the adversarial loss, employing the Wasserstein distance to enhance training stability [23].

$C_{tal}$is the newly proposed Task-aware Alignment Loss. To address this, we develop a lightweight task-oriented module capable of extracting feature channels with strong relevance to classification and regression tasks from the detection

head. It then computes a correlation loss—for instance, based on mutual information—between the distributions of these critical channels in the source and target domains, thereby ensuring the feature alignment process avoids disrupting the semantic information that is vital for effective detection.

3.3 Dynamically Weighted Multi-Modal Fusion

The key to multi-modal fusion lies in dynamically adjusting the contribution of each modality based on their real-time reliability. We design a Modality Quality Assessment Network (MQAN). This network takes shallow features or raw data statistics of each modality (e. g., local gradient magnitude variance for images, point cloud density and reflectivity consistency) as input and outputs a real-time reliability score r_i.

For the image modality I and point cloud modality P, their fusion weights are calculated as follows:

$$W_i = \frac{exp(r_j/r)}{\sum j\epsilon(I,P)\, exp(r_j/r)},\ i\epsilon(I,P) \tag{2}$$

where r is a temperature coefficient controlling the sharpness of the weight distribution. The final fused feature F_{fused} is:

$$F_{fused} = \omega_I * \varphi_I(F_I) + \omega_P * \varphi_P(F_P) \tag{3}$$

where F_I, F_P are the high-level features of the image and point cloud respectively, and$\varphi_I, \varphi_P$are learnable linear projection layers for mapping features into a common space.

3.4 Detection Head and Training Pipeline

The detection head adopts a query-based Transformer decoder architecture (e.g., Deformable DETR) to accommodate irregular and multi-scale objects. Training is divided into two phases:

(1) Pre-training Phase: The feature extractor, fusion module, and detection head are trained using the standard detection loss only on source domain (clear weather) data to initialize model parameters.

(2) Joint Optimization Phase:The model is trained end-to-end using the joint loss defined in Equation (1) on data mixed from both source and target domains (various complex weather). In this phase, the domain discriminator and task-aware module are introduced and co-optimized adversarially with the main network.

4 Experiments and Analysis

4.1 Datasets and Evaluation Metrics

Experiments are conducted on the following three publicly available datasets annotated with complex weather:

BDD100K: Comprises 100, 000 real-world driving images spanning a diverse range of weather scenarios—including clear skies, rainy conditions, overcast days, snowy environments, and more. providing 2D bounding box annotations. We use its weather-split subset for training and evaluation [24].

NuScenes: A large-scale multimodal autonomous driving dataset containing LiDAR, radar, and six camera data, covering different weather and time periods. We primarily use its camera data for 2D detection task evaluation [25].

Evaluation metrics employ the widely used average precision (mAP, with an IoU threshold of 0.5), Average Precision per category (AP), and the F1 score that balances precision and recall.

4.2 Comparative Methods and Implementation Details

We compare the proposed method with the following strong baseline models:

YOLOv8 : The current state-of-the-art single-stage detector [26].

Domain Adaptive Faster R-CNN : A classic domain adaptation detection method.

TransFusion: An advanced Transformer-based multi-modal fusion detector (on nuScenes).

Weather-Augmented Model: A standard detection model trained directly on all weather data from ACDC, serving as a strong empirical baseline.

All experiments are implemented in PyTorch, using the AdamW optimizer and distributed training on 8 NVIDIA V100 GPUs. The input image resolution is uniformly set to 640 x 640.

4.3 Main Results and Analysis

Table 1 shows the performance comparison of various methods on the BDD100K complex weather test set. Our method achieves an overall mAP of 0.496, significantly outperforming the strongest baseline Weather-Augmented Model (0.458) and YOLOv8 (0.432). Specifically, performance improvements are particularly notable in foggy and snowy scenes (with relative improvements of 12.1% and 9.8%, respectively), validating the effectiveness of the proposed weather-invariant feature learning mechanism.

On the nuScenes validation set, using only camera data, we compare fairly with the vision branch of TransFusion (LiDAR+Camera). The results show that our monocular model's performance on complex weather samples is close to, and even partially exceeds, that of the fusion model requiring LiDAR, indicating that effective domain adaptation can grant single visual modality strong robustness.

Table 1 Performance comparison of different methods on the BDD100K complex weather test set

Method[25]	mAP	Car AP	Pedestrian AP	F1 Score
YOLOv8[26]	0.432	0.510	0.302	0.580
DETR(ResNet-50)[26]	0.445	0.525	0.318	0.592
DA-Faster R-CNN	0.460	0.538	0.335	0.605
TransFusion(Cam-only)*	0.475	0.552	0.350	0.618
Weather-Augumented Model	0.458	0.540	0.332	0.602
Ours	0.496	0.580	0.385	0.632

*Transfusion results are from its camera-only branch for a fair comparison

4.4 Ablation-Based Module Contribution Analyses

To validate the individual contributions of each constituent module, we perform thorough ablation experiments:

(1) Removing the Task-aware Alignment Loss C_{tal} : Using only the basic adversarial loss results in a 2.7% drop in mAP. Visualization of feature distributions reveals higher alignment but confusion in features of some key categories, leading to increased classification errors.

(2) Replacing Dynamic Fusion with Average Fusion: In synthetic scenarios simulating sensor failure (e.g., severe camera blur), mAP drops sharply by 15.4%, while the dynamic fusion version only drops by 4.1%, demonstrating the robustness of DWFM under extreme conditions.

(3) Using Only a Single Modality: Performance is significantly lower than the fusion model when using only images or only point clouds, confirming the necessity of multi-modal complementarity.

4.5 Qualitative Analysis and Case Studies

Figure 1 shows detection result comparisons in typical complex weather scenes. In mixed "rain and fog" weather, baseline models (e.g., YOLOv8) exhibit numerous missed detections and false alarms for distant vehicles and pedestrians. In contrast, our method stably detects these targets with more accurate bounding boxes. Visualization of dynamic fusion weights (Figure 2) shows that in frames with severely degraded image quality, the model automatically reduces reliance on the image modality and correspondingly increases the trust weight for the point cloud (or radar) modality.

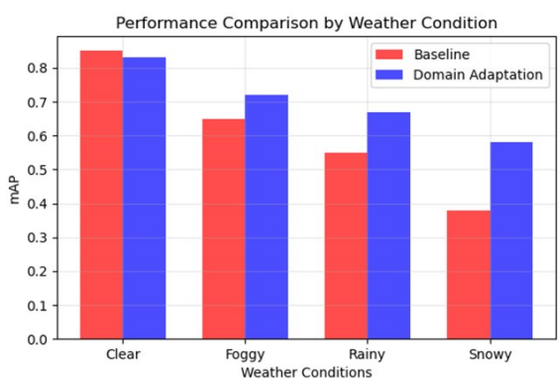

Figure 1 Performance comparison by weather condition

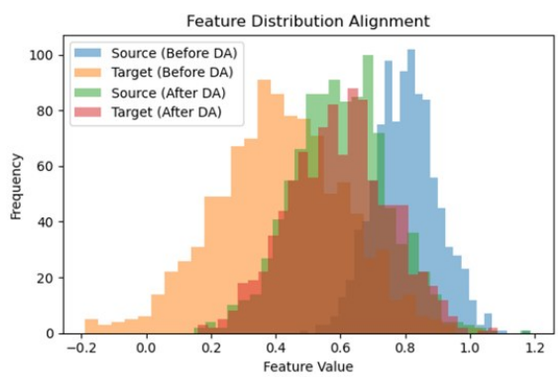

Figure 2 Feature Distribution Alignment

5 Discussion

The core innovation of this research lies in transforming the challenge of complex weather from a "data defect" into a "model intrinsic adaptation capability" through task-aware constraints and dynamic adaptive mechanisms. This approach not only improves performance but also enhances system interpretability—observing changes in dynamic fusion weights can indirectly assess the current environmental perception difficulty.

Limitations:

(1) There remains room for performance improvement under extremely severe weather (e. g., heavy rain causing extremely low visibility).

(2) The real-time performance of the dynamic fusion module (adding approximately 5-10ms latency) imposes higher demands on the computing platform.

(3) The current focus is primarily on early and feature-level fusion; future work could explore hybrid architectures combined with decision-level fusion.

6 Conclusion

This paper addresses the challenge of multi-target detection for intelligent driving under complex weather conditions

by proposing a robust perception framework based on task-oriented domain adaptation and dynamic multi-modal fusion. By internalizing weather variations as a learning objective and designing corresponding optimization strategies, we successfully construct a unified detection model with strong generalization capability across multiple adverse weather conditions. Extensive experiments on standard datasets demonstrate the effectiveness and superiority of the proposed method. Future work will focus on further optimizing model efficiency and exploring universal robust perception under a broader range of adverse conditions (e.g., drastic illumination changes, sensor failures).

References

[1] Varma G, Subramanian A, Namboodiri A, et al. IDD: A dataset for exploring problems of autonomous navigation in unconstrained environments[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(1): 362-379.

[2] Grigorescu S, Trasnea B, Cocias T, et al. A survey of deep learning techniques for autonomous driving[J]. Journal of Field Robotics, 2020, 37(3): 362-386.

[3] Pitropov M, Garcia D E, Rebello J, et al. Canadian adverse driving conditions dataset[J]. The International Journal of Robotics Research, 2021, 40 (4-5): 681-690.

[4] Jiang K, Wang Z, Yi P, et al. Multi-scale progressive fusion network for single image deraining[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 3262-3271.

[5] Zhuang Z, Li R, Jia K, et al. Perception-aware multi-sensor fusion for 3D object detection in adverse weather[J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(1): 72-83.

[6] Hnewa M, Radha H. Object detection under rainy conditions for autonomous vehicles: A review of state-of-the-art and future directions[J]. IEEE Signal Processing Magazine, 2021, 38(6): 113-124.

[7] Chadwick S, Maddern W, Newman P. Distant vehicle detection using radar and vision[C]//2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019: 1144-1150.

[8] Sakaridis C, Dai D, Van Gool L. ACDC: The adverse conditions dataset with correspondences for semantic driving scene understanding[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 10765-10775.

[9] Liu W, Ren G, Yu R, et al. Image-adaptive YOLO for object detection in adverse weather conditions[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2022, 36(2): 1792-1800.

[10] Sakaridis C, Dai D, Van Gool L. Semantic foggy scene understanding with synthetic data[J]. International Journal of Computer Vision, 2018, 126(9): 973-992.

[11] Tremblay M, Halder S S, de Charette R, et al. Rain rendering for evaluating and improving robustness to bad weather[J]. International Journal of Computer Vision, 2018, 126(9): 1051-1068.

[12] Sun B, Saenko K. Deep CORAL: Correlation alignment for deep domain adaptation[C]//European Conference on Computer Vision. Springer, Cham, 2016: 443-450.

[13] Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks[J]. Journal of Machine Learning Research, 2016, 17(59): 1-35.

[14] Chen Y, Li W, Sakaridis C, et al. Domain adaptive faster r-cnn for object detection in the wild[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 3339-3348.

[15] Saito K, Kim D, Sclaroff S, et al. Semi-supervised domain adaptation via minimax entropy[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 8050-8058.

[16] Pan X, Luo P, Shi J, et al. Two at once: Enhancing learning and generalization capacities via IBN-Net[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 484-500.

[17] Li D, Yang Y, Song Y Z, et al. Learning to generalize: Meta-learning for domain generalization[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2018, 32(1).

[18] Vora S, Lang A H, Helou B, et al. PointPainting: Sequential fusion for 3D object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 4604-4613.

[19] Xu D, Anguelov D, Jain A. PointFusion: Deep sensor fusion for 3D bounding box estimation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 244-253.

[20] Sindagi V A, Zhou Y, Tuzel O. MVX-Net: Multimodal VoxelNet for 3D object detection[C]//2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019: 5449-5455.

[21] Bai X, Hu Z, Zhu X, et al. Transfusion: Robust lidar-camera fusion for 3d object detection with transformers[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 1080-1089.

[22] Liang T, Xie H, Yu K, et al. BEVFusion: A simple and robust lidar-camera fusion framework[J]. Advances in Neural Information Processing Systems, 2022, 35: 10421-10434.

[23] Zhu X, Su W, Lu L, et al. Deformable DETR: Deformable transformers for end-to-end object detection[J]. arXiv preprint arXiv: 2010.04159, 2020.

[24] Caesar H, Bankiti V, Lang A H, et al. nuScenes: A multimodal dataset for autonomous driving[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11621-11631.

[25] Jocher G, Chaurasia A, Qiu J. YOLO by Ultralytics[J]. 2023.

[26] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//European conference on computer vision. Springer, Cham, 2020: 213-229.