

Article

YOLO-DH: Robust Object Detection for Autonomous Vehicles in Adverse Weather

Liye Jiang ^{1,2}, Guanqun Ma ^{1,2}, Weixuan Guo ^{1,2} and Yunhao Sun ^{3,*} ¹ Chang'an-Dublin International College of Transportation, Chang'an University, Xi'an 710064, China; 2022900712@chd.edu.cn (L.J.); guanqun.ma@ucdconnect.ie (G.M.)² School of Mechanical and Materials Engineering, University College Dublin, D04 V1W8 Dublin, Ireland³ School of Software, Dalian University of Foreign Languages, Dalian 116044, China

* Correspondence: sunyunhao@dlufl.edu.cn

Abstract

Although autonomous driving technology has advanced rapidly in recent years, the intrinsic complexities of adverse weather conditions pose significant challenges to achieving high object detection accuracy, despite extensive research on perception systems in autonomous vehicles. Accordingly, this study proposes a YOLO-based network that integrates the DHNet residual network for image dehazing with wavelet-based channel attention modules, specifically tailored for target detection in autonomous driving under adverse weather conditions. The proposed network introduces a novel channel-value attention mechanism and sequentially applies wavelet-based soft-thresholding denoising techniques, thereby improving its sensitivity to discriminative channel information. We further introduce the first DHNet Dehazing Attention Module, which sequentially combines the DHNet residual network with the MixDehazeNet hybrid structural dehazing network. This integration synergistically improves the network's image dehazing performance. To address dataset class imbalance, this study incorporates Adaptive Threshold Focal Loss (ATFL), which markedly improves training efficiency and model robustness. This optimization enhances the model's generalization ability for target detection and classification tasks under challenging weather conditions. Experimental evaluations, including ablation studies and comparative analyses, demonstrate that the proposed method achieves substantial improvements in accuracy across three benchmark datasets.



Academic Editor: Hua Zhang

Received: 17 October 2025

Revised: 10 November 2025

Accepted: 12 November 2025

Published: 17 November 2025

Citation: Jiang, L.; Ma, G.; Guo, W.; Sun, Y. YOLO-DH: Robust Object Detection for Autonomous Vehicles in Adverse Weather. *Electronics* **2025**, *14*, 4476. <https://doi.org/10.3390/electronics14224476>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: attention mechanism; YOLOv8; adverse weather conditions; image dehazing

1. Introduction

Rapid advancements in deep learning methodologies [1] have catalyzed transformative progress in object detection algorithms [2], especially within autonomous driving systems [3–5]. Within these developments, the YOLO series [6] has become a foundational technology for real-time traffic scene perception [7], due to its outstanding balance between detection efficiency and localization accuracy [8–11]. Existing approaches predominantly target vehicle, pedestrian, and traffic sign detection under clear weather conditions. By optimizing feature extraction networks and bounding box regression strategies, these methods have substantially improved the perception accuracy of backbone networks [12,13]. However, these algorithms largely depend on single-modal visual inputs. Under adverse weather conditions, such as haze or heavy rainfall, sensor-captured images experience contrast reduction and detail blurring, resulting in more than a 40% decline in detection

confidence [14]. Traffic safety studies indicate that under extreme weather with visibility below 50 m, the false detection rate of autonomous driving systems increases by a factor of 3.2 compared to normal conditions [15], posing substantial risks to decision-making processes in complex road environments. These findings highlight a critical bottleneck in deploying autonomous driving technology: the need to develop weather-robust object detection frameworks capable of enabling reliable all-weather perception [16].

Adverse meteorological conditions such as fog, rain, and snow in traffic scenarios [17] severely degrade sensor-captured images, substantially impairing the detection accuracy and response speed of autonomous driving systems for critical targets [18–21]. Consequently, developing robust object detection frameworks under inclement weather has become a pivotal research challenge to ensure autonomous vehicle safety [22]. Recent advancements have primarily focused on image enhancement techniques [23] to improve visual perception in low-visibility environments [24]. Dense block frameworks [25] enhance feature representation via multi-scale attention mechanisms and channel structural optimizations, achieving superior spatial-contextual modeling by aggregating hierarchical visual patterns. In this study, we propose a weakly supervised framework for cross-scenario imaging optimization and investigate low-visibility enhancement using a Multi-Branch Input Encoder–Decoder Network (MBINEDN) integrated with multi-modal visual deep learning [26]. These approaches facilitate feature enhancement through multi-scale attention and address imaging optimization challenges in complex scenarios with efficient computational architectures. However, current methodologies exhibit two critical limitations. First, the decoupled architecture between traditional image enhancement [27] and detection tasks introduces error propagation throughout the processing pipeline and limits the effective utilization of latent feature information in degraded images. Second, existing network architectures, while achieving notable progress in general object detection tasks, still struggle with accurate predictions, particularly for small-scale objects [28], where even state-of-the-art methods often produce suboptimal performance below acceptable thresholds.

Under adverse weather conditions such as haze and heavy rainfall, the channel values in autonomous driving image target detection systems exhibit significant spectral distortion. Specifically, in RGB channels, foggy environments induce luminance channel homogenization across the entire image, accompanied by contrast attenuation and diminished color channel differentiation. This phenomenon obscures discriminative target features that are normally distinct in clear conditions, resulting in blurred inter-channel characteristics for critical objects like vehicles and pedestrians. In inclement weather conditions involving snow or rain, white or gray pixel clusters formed by snowflakes and raindrops on captured images disrupt normal channel value distributions, which manifests particularly in the luminance channel, generating substantial noise artifacts that impair the channel value representation of critical edge and texture features in target objects. Furthermore, the varying degrees of alteration in multi-channel characteristics caused by distinct adverse weather conditions introduce significant complexity to channel-based feature analysis in object detection systems.

Under adverse weather conditions such as haze and heavy rainfall, channel values in autonomous driving image target detection systems undergo significant spectral distortion. In RGB channels, foggy environments lead to luminance homogenization across the entire image, accompanied by contrast attenuation and reduced color differentiation. This effect obscures discriminative features that are normally distinct in clear conditions, resulting in blurred inter-channel characteristics for critical objects such as vehicles and pedestrians. Similarly, in snow or rain, clusters of white or gray pixels formed by snowflakes or raindrops disrupt normal channel value distributions, particularly in the luminance channel,

generating noise artifacts that degrade the representation of edges and texture features in target objects. Moreover, the varying degrees of multi-channel alteration caused by different adverse weather conditions introduce substantial complexity to channel-based feature analysis in object detection systems.

Yongzhen Wang et al. [29] proposed USCFormer, a unified Transformer-based dehazing method with semantic contrastive learning that combines the complementary strengths of CNNs and Transformers. This framework leverages semantic priors to construct a contrastive learning paradigm, optimizing transmission rate estimation for enhanced image restoration. However, the computational complexity of the Transformer limits real-time adaptability, and the generalization of semantic priors under complex dynamic lighting remains a challenge. Liu Xiao-hong et al. [30] introduced GridDehazeNet+, a model-free dehazing method that achieves haze removal and improved real-world scene adaptability through multi-module operations. While it enhances image quality via end-to-end multi-level feature optimization, the grid architecture incurs high computational overhead, limiting real-time performance under extreme haze. Wang et al. [31] further proposed Cycle-SNSPGAN, an unsupervised dehazing method leveraging cycle-consistent generative adversarial networks to demonstrate strong generalization and detail restoration. Nevertheless, its complex network architecture reduces computational efficiency, and performance under extreme haze requires further improvement. Aryan Mehra et al. [32] developed ReViewNet, a lightweight dehazing network achieving efficient haze removal through multi-component collaboration. While offering real-time potential and robust generalization, it requires further optimization in transmission rate stability, dynamic lighting adaptation, and computational overhead of multi-perspective feature fusion. He et al. [33] presented a weakly supervised dehazing approach using collaborative dense blocks and multi-dimensional attention to enhance imaging quality in complex media, but robustness under extreme conditions and real-time performance still need improvement. Liu et al. [34] proposed a foggy weather object detection method combining multi-modal feature fusion with YOLOv3 for improved detection efficiency and accuracy, yet it struggles with feature degradation under extreme fog, high computational cost, and generalization across diverse weather. Wei Xing et al. [35] introduced DADENet, a joint dehazing network enhancing dehazing accuracy and depth estimation via multi-module collaboration, though its cascaded design limits real-time adaptability, cross-weather generalization, and cross-dataset robustness. Zhanwen Liu et al. [36] proposed UFEM, an unsupervised feature enhancement method for robust performance in complex scenes, but convergence stability, computational complexity, and channel correlation generalization under extreme weather require further evaluation. Wei Zhou et al. [37] proposed RRPD/NRBP, a no-reference quality assessment method to capture image distortion, showing strong discriminative ability; however, cross-dataset generalization, extreme fog robustness, and real-time efficiency need validation. Finally, Xiongkuo Min et al. [38] developed DHQI, a no-reference dehazing quality metric with a large-scale multi-feature database. While providing strong discriminative capability and open-source deployment, high online computational complexity, feature adaptability under extreme haze, and assessment under varying illumination and dynamic weather remain challenges.

To address this problem and meet the practical requirements for high accuracy and real-time performance, this study proposes a novel YOLO-based network for autonomous driving target detection in adverse weather conditions, integrating a DHNet residual network for dehazing with a wavelet channel attention module. The proposed network achieves enhanced sensitivity to channel-specific features by sequentially integrating an innovative channel value attention mechanism with wavelet soft-threshold denoising technology, while maintaining only a minor increase in model complexity. Furthermore, it

incorporates a dehazing attention mechanism module that combines the DHNet residual network and the MixDehazeNet hybrid architecture for image dehazing in a step-wise manner, significantly improving haze removal capability. Additionally, to address the issue of dataset class imbalance, the system incorporates the Adaptive Threshold Focal Loss (ATFL), enabling more efficient and robust model training, thereby enhancing performance in target detection and classification tasks under adverse weather conditions. The main contributions of this study can be summarized as follows.

- (1) Innovative Proposal of WTCAM Module. The WTCAM (Wavelet Threshold Channel Attention Module) innovatively integrates wavelet soft-threshold denoising with channel attention mechanisms. By leveraging wavelet soft-threshold denoising for noise suppression and dynamically calibrating feature responses through channel attention, this module achieves effective noise removal, strengthens key information channels, and enhances critical feature extraction.
- (2) First Proposal of DHMD module. The DHMD (Dehazing-Halo Mitigation Detector) module is the first to establish a coupled system based on DHNet and MixDehazeNet, incorporating bidirectional residual feature propagation paths. This architecture effectively suppresses halo artifacts while restoring detailed road surface textures.
- (3) Proposed YOLO-DH Model. The YOLO-DH model integrates the WTCAM module and DHMD module by replacing all original C2f modules with C2f-WTCAM modules and embedding the DHMD dehazing module after the backbone network and the generation of P3/8 feature maps in the detection head. By incorporating Adaptive Threshold Focal Loss (ATFL) with dynamic category weighting coefficients and hard example mining mechanisms, the model effectively alleviates the class imbalance issue between small targets (e.g., pedestrians, animals) and background. Experimental results on three public datasets demonstrate its superior performance compared to YOLOv8.

2. Basic Theoretical Research

This section presents two novel modules: the Wavelet Transform-based Channel Attention Mechanism (WTCAM) and the Dehazing and Meteorological Disturbance Module (DHMD). The WTCAM enhances the model's ability to extract channel features and improve object detection performance through adaptive channel recalibration and multi-scale spatial-contextual modeling. The DHMD module focuses on improving dehazing performance and mitigating the effects of adverse meteorological conditions. In addition, we introduce the Adaptive Threshold Focal Loss (ATFL), which strengthens the model's attention to hard-to-classify samples and addresses class imbalance, thereby improving both training efficiency and detection robustness. Finally, we present an integrated system architecture that combines these components to achieve high-performance object detection under severe weather conditions.

2.1. Wavelet Transform-Based Channel Attention Mechanism (WTCAM)

Wavelet Soft Thresholding Denoising (WSTD) is an improved variant of the conventional Wavelet Thresholding Denoising (WTD) method, based on the principles of wavelet transform. As a widely adopted signal denoising technique, WTD effectively suppresses noise in raw data while minimizing signal distortion and information loss [39–41], thereby significantly enhancing the accuracy of time-series prediction models. The soft-thresholding function in WSTD, as expressed in Formula (1), preserves continuity within the range of

the estimated wavelet coefficients, overcoming the tendency of conventional thresholding methods to introduce oscillations during wavelet reconstruction.

$$\hat{W}_x = \begin{cases} \text{sign}(W_x)(|W_x| - \lambda), & |W_x| \geq \lambda \\ 0, & |W_x| \leq \lambda \end{cases} \quad (1)$$

where W_x represents the wavelet transform coefficient, λ is the threshold, $\text{sign}(\cdot)$ denotes the sign function, and $\hat{W}_x$ is the estimated wavelet coefficient after thresholding.

The Channel Attention Mechanism (CAM) models inter-channel relationships through internal feature dependencies, adaptively enhancing critical channel representations. As shown in Formula (2), the input feature map passes through two parallel MaxPool and AvgPool layers, reducing its spatial dimensions from $C \times H \times W$ to $C \times 1 \times 1$. Subsequently, it is processed by a Shared MLP module. Within this module, the channel dimension is first compressed to $1/r$ of the original (where $1/r$ is the reduction ratio), then expanded back to the original channel count. After applying the ReLU activation function, two activated outputs are generated. The two output results are first summed element-wise, then passed through a sigmoid activation function to generate the final Channel Attention output. This output is subsequently multiplied with the original feature map, restoring its dimensions to $C \times H \times W$.

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (2)$$

Among them, F is the input feature map, *AvgPool* and *MaxPool* represent global average pooling and global max pooling operations, respectively, MLP denotes a multi-layer perceptron, σ is the Sigmoid activation function, and $M_c(F)$ is the channel attention weight matrix.

This study integrates wavelet soft-threshold denoising with channel attention mechanisms to construct the WTCAM module. As shown in Figure 1, the noisy signal (adverse weather image) first undergoes wavelet transformation, decomposing it into wavelet coefficients of different scales and frequencies. For instance, after multi-layer wavelet decomposition, the image will generate low-frequency approximation coefficients and high-frequency detail coefficients. By setting appropriate thresholds for these wavelet coefficients and applying a soft-thresholding function, coefficients with absolute values smaller than the threshold are set to zero, while those exceeding the threshold undergo shrinkage. After removing noise-corrupted coefficients, the denoised signal is reconstructed into the original signal space through inverse wavelet transform, yielding a preliminary noise-reduced signal. The signal is then fed into a convolutional neural network (CNN) to extract feature maps from different channels. Global average pooling is applied to compress each channel's feature map into a scalar value. Subsequently, a fully connected layer or a simple network with activation functions calculates the attention weight for each channel. Finally, these weights are multiplied with the original feature maps of corresponding channels to achieve feature recalibration through channel-wise reweighting. To address the degradation issues of channel values in autonomous driving images under adverse weather conditions (including brightness/contrast imbalance, color inaccuracy, and noise interference), this paper integrates the WTCAM module with the C2f's convolutional layer to enhance the model's sensitivity to channel-specific feature representations.

Unlike traditional channel attention modules such as SE (Squeeze-and-Excitation) and CBAM (Convolutional Block Attention Module), the proposed Channel Value Attention mechanism introduces a wavelet-domain calibration process that directly captures the frequency-domain correlations of channel values rather than relying solely on spatial pooling operations. Specifically, by integrating wavelet soft-thresholding, the WTCAM module dynamically suppresses noise components in high-frequency sub-bands and am-

plifies discriminative low-frequency responses before channel reweighting. This two-stage denoising–recalibration design allows for the network to enhance contrast-sensitive channel information degraded by adverse weather. Consequently, the proposed mechanism achieves a more physically interpretable and weather-adaptive attention process, fundamentally differing from the purely statistical feature weighting used in SE or CBAM.

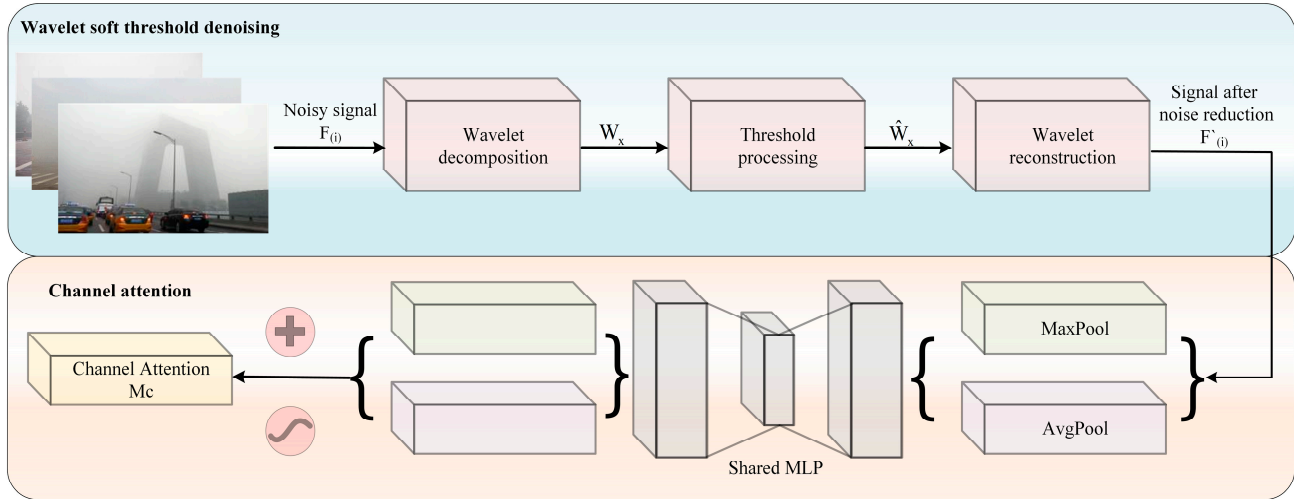


Figure 1. WTCAM (Wavelet Threshold Channel Attention Module) structure diagram.

2.2. Dehazing Module Based on DHNet Residual Network (DHMD)

To address the challenges of low contrast, reduced saturation, and color distortion in UAV aerial images of forested areas under haze conditions, Lu et al. [42] proposed the DHNet residual network module, as formulated in Equations (3) and (4). This architecture incorporates information transfer modules at each level of the backbone network, transforming feature maps into weight maps for selective filtering and inter-layer transmission. Thresholding at the receiving end mitigates the adverse effects of redundant information. Through dense connections, the network enhances global dehazing performance and improves haze removal quality across both high-frequency and low-frequency regions in forest UAV images.

The Global Feature Fusion (GFF) module selectively filters and extracts features from other layers while facilitating inter-layer feature transmission. By employing a gating mechanism to evaluate the usefulness of information, it effectively reduces redundancy. The weighting formula for quantifying feature vector importance is presented in Equation (3).

$$\tilde{X}_n = (1 + G_n) \cdot X_n + (1 - G_n) \cdot \sum_{i=1, i \neq n}^N G_i \cdot X_i \quad (3)$$

After feature transmission, the incorporation of a Densely Fused Network (DFN) module with dense connections can enhance information density, thereby optimizing the overall structure, chromaticity, and luminance performance in image dehazing. Following the output of the Gated Fully Fusion (GFF) module in the backbone network, a Dense Feature Pyramid (DFP) module is employed to encode the output information, capturing global contextual features.

$$y_i = H_i(y_n, \tilde{X}_1, \dots, \tilde{X}_{i-1}) \quad (4)$$

In the formula, y_n represents the output of the n th layer in the backbone network, $\tilde{X}_n$ denotes the output of the i th module in the Global Feature Fusion (GFF), and H_i refers to a single convolutional layer. The role of the Dynamic Feature Network (DFN) is to

combine the outputs from GFF with the convolutional layers of the backbone network, thereby further improving the overall quality of the dehazed image.

The use of large convolutional kernels often overlooks the multi-scale characteristics inherent in images, while the conventional serial arrangement of attention modules is insufficient for addressing the inhomogeneous distribution of fog. To overcome these limitations, Chan et al. [43] proposed MixDehazeNet, which introduces a Multi-Scale Parallel Large Convolution Kernel Module (MSPLCK, defined in Formula (3)) and an Enhanced Parallel Attention Module (EPA, described in Formulas (5)–(10)). Compared with single large convolution kernels, multi-scale parallel large convolution kernels better preserve texture details during dehazing. Additionally, the EPA module leverages parallel connections of attention mechanisms, achieving improved handling of uneven haze distribution [44]. Extensive experiments on three benchmark datasets demonstrate the effectiveness of this approach in enhancing dehazing performance.

$$\begin{aligned} x_1 &= PWConv(\hat{x}), \\ x_2 &= Conv(x_1), \\ x_3 &= Concat(DWDCConv19(x_2), \\ &\quad DWDCConv13(x_2), \\ &\quad DWDCConv7(x_2)) \end{aligned} \quad (5)$$

Here, PWConv denotes pointwise convolution. Conv represents standard convolution with a kernel size of 5. DWDCConv19 indicates a depthwise dilated convolution with an effective kernel size of 19 (equivalent to a 7×7 kernel with a dilation rate of 3). DWDCConv13 corresponds to a depthwise dilated convolution with an effective kernel size of 13 (equivalent to a 5×5 kernel with a dilation rate of 3). DWDCConv7 refers to a depthwise dilated convolution with an effective kernel size of 7 (equivalent to a 3×3 kernel with a dilation rate of 3). Finally, Concat denotes channel concatenation of features.

$$y = x + PWConv(GELU(PWConv(x_3))) \quad (6)$$

Next, x_3 is fed into a multilayer perceptron (MLP) that transforms its feature dimensions to match those of x . This MLP comprises two pointwise convolution layers and utilizes the GELU activation function. The MLP output is then added to the identity shortcut connection xxx . We hypothesize that this MLP not only facilitates the fusion of three distinct feature types, but also plays a crucial role in capturing and fitting dehazing-related features [44].

$$\begin{aligned} PF_s &= Conv(PWConv(\hat{x})) \\ PA_s &= Sigmoid(PWConv(\hat{x})), \\ F_s &= PF_s \otimes PA_s \end{aligned} \quad (7)$$

Among them, the feature extraction branch (PF_s) and the pixel gating branch (PA_s) are designed to work synergistically. Here, PA_s serves as the pixel-wise gating signal for PF_s , where $PWConv$ denotes element-wise multiplication, CONV represents pointwise convolution (1×1 convolution), and $CONV3 \times 3CONV3 \times 3$ indicates standard convolution with a kernel size of 3. The pixel attention module incorporates PA_p branches, enabling the extraction of global pixel-wise gating features.

$$\begin{aligned} PA_p &= Sigmoid(PWConv(GELU(PWConv(\hat{x})))), \\ F_p &= \hat{x} \otimes PA_p. \end{aligned} \quad (8)$$

The network employs a PWConv-GELU-PWConv structure to fit feature representations. The Sigmoid function is utilized to extract global pixel-wise gating features, then

PA_p is applied as global pixel gating signals to modulate feature maps of $\hat{x}$ through element-wise multiplication.

Channel attention mechanisms can effectively capture global information and adjust the feature representation across channel dimensions. They typically incorporate a CA_c branch to extract holistic channel characteristics.

$$CA_c = \text{Sigmoid}(PWConv(GELU(PWConv(GAP(\hat{x}))))), \quad (9)$$

$$F_c = \hat{x} \otimes CA_c.$$

The global channel gating features are extracted by sequentially applying Global Average Pooling (GAP), PWConv-GELU-PWConv layers, and Sigmoid functions, then CA_c is utilized as global channel gating signals of $\hat{x}$ for feature recalibration.

$$F = \text{Concat}(F_s, F_c, F_p)$$

$$y = x + PWConv(GELU(PWConv(F))) \quad (10)$$

First, concatenate the outputs of three distinct attention gating mechanisms along the channel dimension. Subsequently, apply a multi-layer perceptron (MLP) composed of PWConv-GELU-PWConv layers to reduce the concatenated feature's channel dimension to match the input x . Finally, the output of the MLP is combined with the identity shortcut connection $\hat{x}$ through summation.

As shown in Figure 2, this study integrates the dense residual network DHNet with the dehazing module to learn the non-linear mapping between degraded and clear images. This approach aims to enhance the network's adaptability to non-uniform haze distribution, thereby improving image quality. By sequentially combining the DHNet module with the MixDehazeNet module, the system effectively fuses their complementary advantages. The DHNet module is specifically designed for images with channel-value blurring characteristics under adverse weather conditions. Its Gated Feature Fusion (GFF) module (as shown in Formula (5)) effectively transfers and filters multi-level features through a gated mechanism, enhancing image details and color restoration. Additionally, the Deep Feature Network (DFN) module (as shown in Formula (6)) optimizes the overall image quality by refining feature interactions and contextual modeling. The MixDehazeNet module, with its hybrid structure blocks, enhances feature extraction capabilities and demonstrates superior adaptability to complex fog conditions.

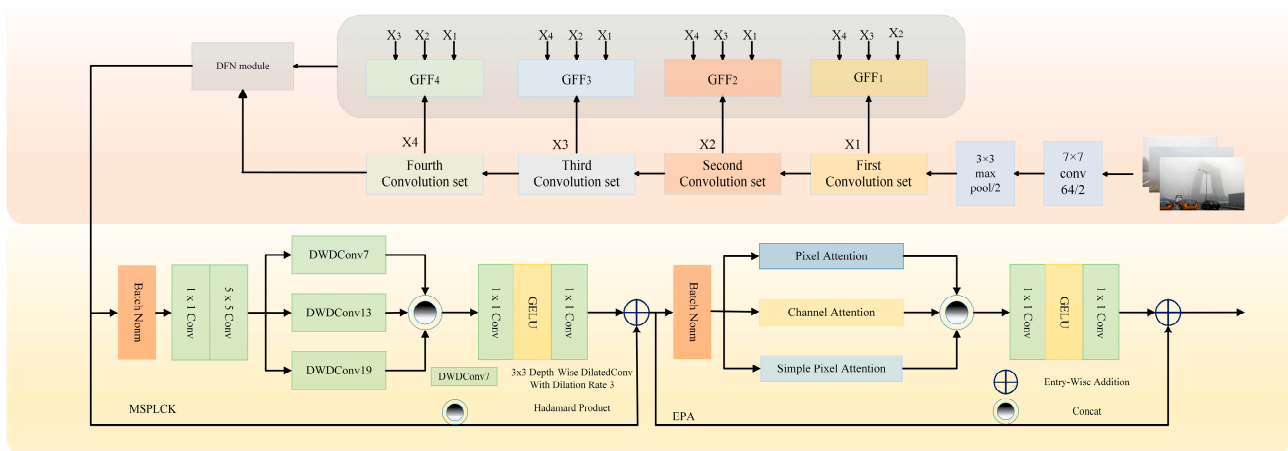


Figure 2. DHMD structure diagram.

The integration of DHNet with MixDehazeNet is not arbitrary, but is theoretically motivated by their complementary feature-learning paradigms. DHNet, as a residual-based

architecture, emphasizes local structural recovery through gated feature fusion and dense connections, effectively restoring fine-grained textures. In contrast, MixDehazeNet focuses on global contextual perception via multi-scale large-kernel convolutions and enhanced attention modules. By coupling these two models sequentially, DHMD enables hierarchical learning—DHNet first reconstructs high-frequency spatial details, while MixDehazeNet subsequently optimizes low-frequency global luminance and color consistency. This design ensures that both local and global degradations in hazy images are simultaneously corrected, providing a theoretically grounded synergy rather than an ad hoc combination.

By integrating the DHNet module (specialized in handling blurry scenes) with MixDehazeNet's hybrid architecture, this combination leverages the scene-adaptive properties of DHNet while capitalizing on the mixed structure blocks of MixDehazeNet to improve processing across diverse fog scenarios [44]. This integration is expected to significantly enhance dehazing performance, achieving superior results in tasks involving varying fog densities and scene complexities. It optimizes both visual quality and detail preservation in dehazed images.

2.3. Adaptive Threshold Focal Loss (ATFL)

In autonomous driving datasets for adverse weather conditions, the application of Adaptive Threshold Focal Loss (ATFL) is widespread. These datasets suffer from class imbalance issues, where background samples significantly outnumber foreground target samples. Under harsh weather scenarios, irrelevant background information dominates, while critical target samples (e.g., vehicles, pedestrians) are sparse. The Adaptive Threshold Focal Loss (ATFL) optimizes traditional cross-entropy loss by adaptively adjusting thresholds based on sample-related features, enabling precise identification of hard-to-classify samples and dynamic allocation of loss weights. Compared to conventional Focal Loss, ATFL demonstrates superior adaptability to diverse sample distributions across datasets. In this study, we introduce the Adaptive Threshold Focal Loss function to enhance model training efficiency and robustness, significantly improving model performance in challenging tasks such as target detection and classification under adverse weather conditions.

For easy-to-classify samples, we expect the loss value to decrease as the predicted probability p_t increases, thereby further reducing the contribution of such samples to the total loss. At the beginning of training, even the predicted probabilities of easy samples are relatively low, but as training progresses, these probabilities gradually increase. Consequently, their corresponding loss values should asymptotically approach zero. The predicted probability values of true targets $\hat{p}_c$ can be mathematically modeled to dynamically guide the training process.

$$\hat{p}_c = 0.05 \times \frac{1}{t-1} \sum_{i=0}^{t-1} \bar{p}_i + 0.95 \times p_t \quad (11)$$

In the formula, $\hat{p}_c$ represents the predicted value for the next training epoch, p_t denotes the current average predicted probability value, and $\bar{p}_i$ indicates the average predicted probability value for each training epoch. Based on Shannon's information theory, the adaptive modulation factor Y can be expressed as

$$Y = -\ln(\hat{p}_c) \quad (12)$$

However, during the later stages of network training, the expected probability values become excessively large, which reduces the proportion of hard samples. We denote η as

$$\eta = -\ln(p_t) \quad (13)$$

Therefore, the mathematical expression of the Adaptive Threshold Focal Loss (ATFL) can be derived as

$$ATFL = \begin{cases} -(\lambda - p_t)^{-\ln(p_t)} \log(p_t) p_t & \leq 0.5 \\ -(1 - p_t)^{-\ln(\hat{p}_c)} \log(p_t) p_t & > 0.5 \end{cases} \quad (14)$$

2.4. Overall Network

As shown in Figure 3, the proposed YOLO-DH is an improved version of YOLOv8. The model input is a three-channel image resized to $640 \times 640 \times 3$.

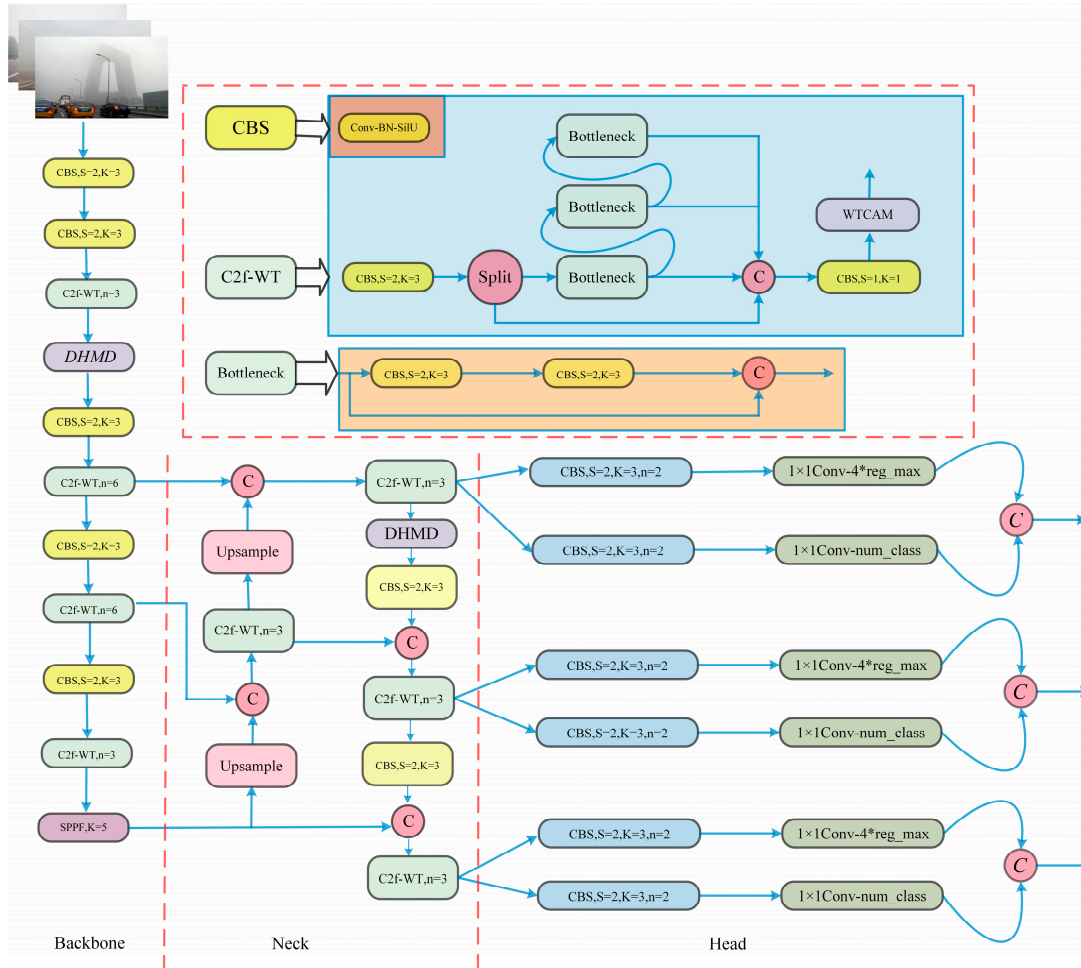


Figure 3. YOLO-DH architecture.

In the traditional YOLOv8 architecture, the C2F module serves as the fundamental building block. Under low-light conditions (e.g., nighttime, tunnels) or dense fog scenarios, the overall brightness of images significantly decreases, causing pixel information reflected by channel values to become dim and blurred. To enhance the model's sensitivity to channel-wise features, all original C2F modules are replaced with the C2f-WTCAM module.

The backbone network serves as the core of the model, responsible for extracting features from images. It primarily consists of multiple convolutional layers and downsampling layers, with the introduction of residual connections and bottleneck structures to reduce model complexity while enhancing performance. In this study, the DHMD dehazing module is integrated into the backbone network. By removing fog interference before the image enters deeper feature extraction layers, subsequent convolutional layers can more effectively capture the intrinsic features of the image. This design prevents haze from

degrading feature extraction and significantly improves the overall feature representation capability of the model.

The head module is responsible for the final target detection and classification. In YOLOv8, there are three detection heads designed to detect targets of three different sizes. In this study, we introduce the DHMD Dehazing Module after the generation of the P3/8 feature map to enhance the feature representation of small targets in foggy environments. Since the P3/8 feature map primarily handles small target detection, the dehazing operation improves the clarity of small targets, thereby boosting their detection accuracy.

The YOLOv8 loss function consists of two components: bounding box regression loss (CIoU + DFL) and class classification loss (VFL). In adverse weather conditions, the blurred features of target objects exacerbate the positive-negative sample imbalance issue. For example, large background areas obscured by fog or rain are prone to be misclassified as negative samples, while true target objects (positive samples) are significantly fewer in comparison. Focal Loss addresses class imbalance by assigning varying weights to samples of different difficulty levels, prioritizing hard-to-classify samples while reducing the influence of easily classified ones. This mechanism enables models to focus on accurately identifying blurred or ambiguous targets in adverse weather conditions (e.g., fog, rain), thereby minimizing classification errors. During training, we employ the Adaptive Threshold Focal Loss (ATFL) to further enhance classification accuracy.

3. Experimental Results and Discussion

This section first introduces the construction process of the autonomous driving image dataset, experimental environment setup, and evaluation metrics for experimental results. Subsequently, through extensive experiments, the proposed target detection network (YOLO-DH) is comprehensively compared with existing mainstream object detection networks, including YOLOv5, YOLOv6, YOLOv8, YOLOv10, and YOLOv11.

3.1. Dataset

This paper introduces three datasets: COCO2017, RTTS, and KITTI. Table 1 provides a comprehensive overview of these datasets, detailing the following characteristics for each: training-test dataset configurations, annotation types, resolution, image sources, and dataset scale.

Table 1. Detailed descriptions of COCO2017, RTTS, and KITTI datasets.

Datasets	COCO2017	RTTS	KITTI
Number of Classes	80	5	8
Training Set Size	118,287	3025	7481
Test Set Size	5000	1297	7518
Annotation Type	Object Detection	Object Detection	Object Detection
Resolution	640 × 640	640 × 640	1280 × 384
Image Source	General Object Detection	Vehicle Vision and Autonomous Driving	Vehicle Vision and Autonomous Driving
Dataset Size	25 GB	1 G	12 GB

Prior to training, all images from the COCO2017, RTTS, and KITTI datasets were preprocessed and augmented to ensure consistency and enhance model robustness. Each image was resized to 640 × 640 pixels, normalized to the [0, 1] range, and converted to RGB format. Basic contrast-limited adaptive histogram equalization (CLAHE) was applied to mitigate low-visibility effects in hazy scenes.

Data augmentation techniques, including random horizontal flipping ($p = 0.5$), random rotation ($\pm 10^\circ$), brightness and contrast adjustments ($\pm 15\%$), and Gaussian noise injection, were employed during training to improve generalization under diverse weather and illumination conditions.

For experiments involving both synthetic (RTTS) and real-world (COCO2017, KITTI) datasets, a 3:2 ratio between real and synthetic samples was maintained in each training batch. Domain discrepancies were alleviated through normalization of pixel intensity distributions and adaptive fine-tuning of the early convolutional layers.

These preprocessing and augmentation procedures ensure that the proposed model is trained on a diverse and balanced dataset, improving robustness to variations in visibility, illumination, and domain characteristics.

3.2. Experimental Setup

During the model training process, all experiments were conducted under the software environment of Windows 11, Python 3.8, and PyTorch 1.12. The training configuration includes the following settings: learning rate of 0.001, batch size of 16, 350 training epochs, and the use of the Adam optimizer. The hardware platform is configured as follows: CPU: Intel Core i9-10900K (10 cores/20 threads, base frequency 3.7 GHz, max turbo frequency 5.3 GHz, 14 nm process, TDP 125W)145, GPU: NVIDIA GeForce GTX 3090 (24 GB GDDR6X VRAM, 35,580 GFLOPS at FP16 precision), Acceleration Libraries: CUDA 11.1, cuDNN 8. The evaluation metrics include Precision, Recall, mean Average Precision at 50% IoU (mAP@0.5), Floating Point Operations (FLOPs), Frame Rate (FPS), and Model Parameter Count (Params).

To ensure full reproducibility, the evaluation protocol is explicitly defined. For each dataset, the samples are randomly divided into training, validation, and test sets with ratios of 70%, 15%, and 15%, respectively. The model is trained for 350 epochs with a batch size of 16 using the Adam optimizer and an initial learning rate of 0.001. To guarantee experimental consistency, the random seed is fixed at 42 across all runs. During evaluation, the model performance is reported as the average of three independent runs to minimize the impact of stochastic variations.

3.3. Experimental Results

3.3.1. Ablation Experiment

To investigate the contributions of different components in the proposed method, an ablation experiment was conducted on the RTTS dataset. The experimental results are shown in Table 1. Method (a) is the original YOLOv8n model. In method (b), only the Adaptive Threshold Focal Loss (ATFL) is applied to the original YOLOv8n. From the results, the FPS of the model is slightly improved, and both mAP@50 and mAP@50-95 are increased by 0.7%. The number of model parameters changes little, which verifies that ATFL can achieve relatively excellent results for the data imbalance problem in the autonomous driving dataset. In method (c), after adding ATFL, the proposed Dehazing Module (DHMD) is added before the P3/8 layer of the backbone network and after the generation of the P3/8 feature maps of the detection head. Improvements of 0.9%, 2.3%, 2.9%, and 2.4% are achieved in the P, R, mAP@50, and mAP@50-95 metrics, respectively. Only 875,594 bits of parameters are added, but the GFLOPS parameter increases by 12 points. This is because more complex structures or operations are introduced to improve the performance such as detection accuracy. It is a trade-off between sacrificing a small part of the inference speed and hardware performance requirements to improve the accuracy. Method (d) adds the proposed Weighted Transformer Channel Attention Module (WTCAM) on the basis of method (b). The P and R metrics are slightly improved compared with method (b).

Although the mAP@50 and mAP@50-95 metrics change little, it is sufficient to show that this module can slightly improve the channel feature extraction ability of the model. It can be seen that the approach of blindly pursuing the data extraction ability of the model for images in bad weather without dehazing treatment is not advisable, and the benefits are not optimistic. Method (e) is the proposed YOLO-DH model, as shown in Figure 4 and Table 1. Compared to the original algorithm, with a small loss in GFLOPS, FPS, and the number of parameters, the P, R, mAP@50, and mAP@50-95 are increased by 2.8%, 2.3%, 3%, and 2.7%, respectively. For more information, please refer to Figures 4 and 5, Tables 2 and 3.

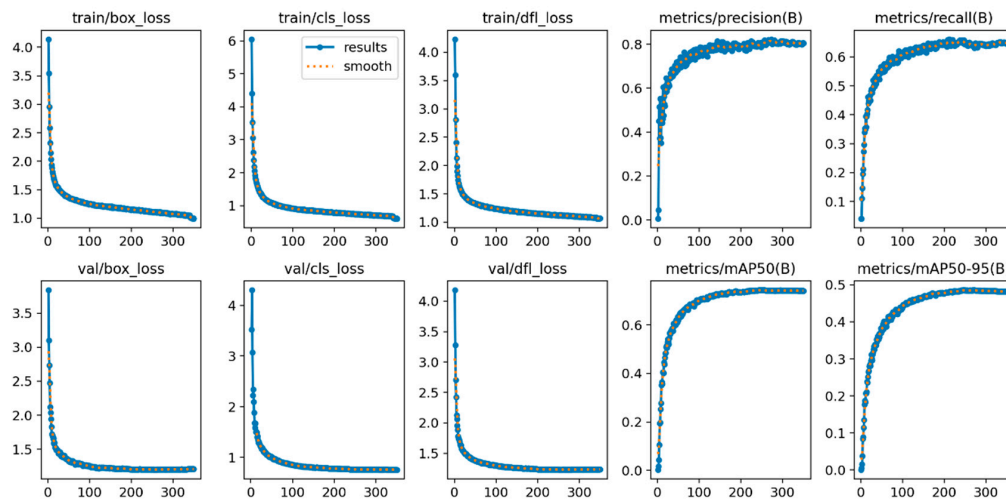


Figure 4. Training results of YOLOv8-DH on the RTTS dataset.

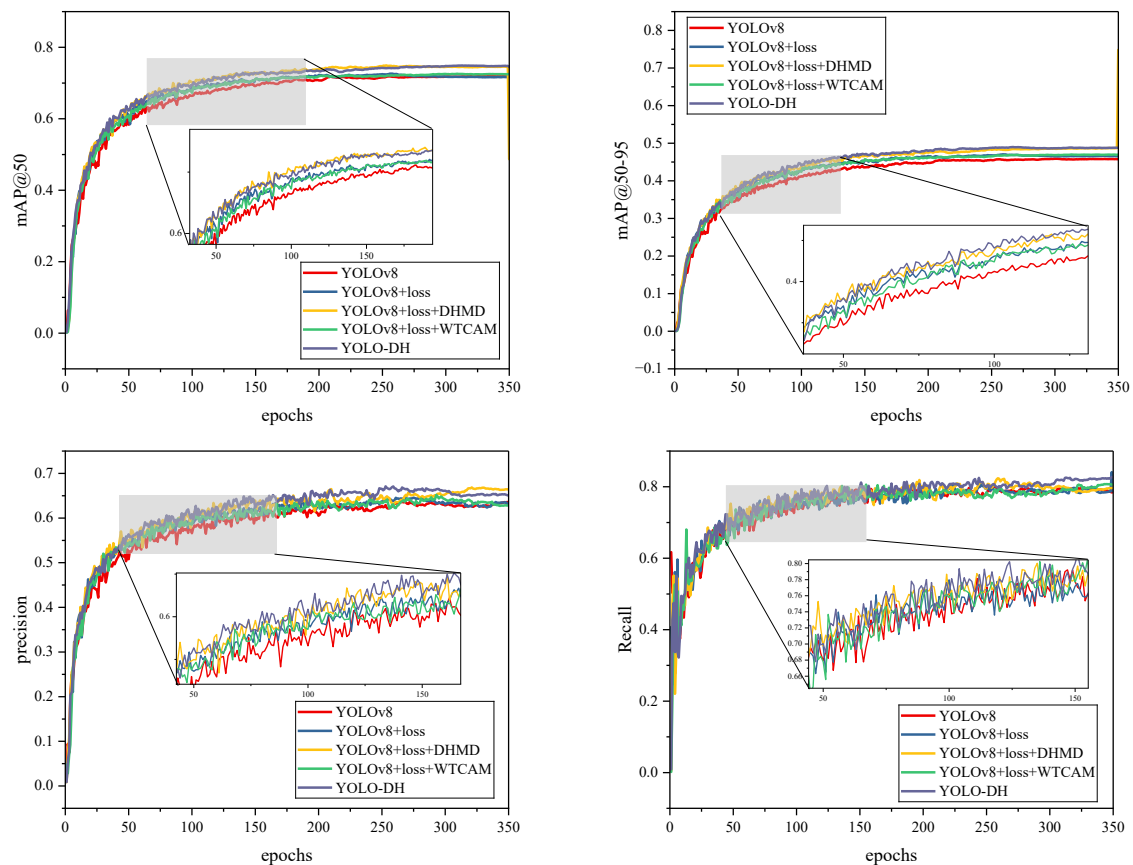


Figure 5. Ablation experimental results.

Table 2. Ablation study results of the proposed method on the RTTS dataset.

Method	Loss	DHMD	WTCAM	P	R	mAP@50	mAP@50-95	Number of Parameters	GFLOPS	Frame Rate
a				0.784	0.64	0.718	0.463	3,011,823	8.2	59.49
b	✓			0.79	0.64	0.725	0.47	3,011,823	8.86	61.23
c	✓	✓		0.793	0.663	0.747	0.487	3,887,417	20	44.94
d	✓		✓	0.8	0.641	0.727	0.477	3,079,831	8.4	59
e	✓	✓	✓	0.812	0.663	0.748	0.49	3,955,041	20.2	43.53

Table 3. Comparison of mAP@50 for specific categories in the model ablation study on the RTTS dataset.

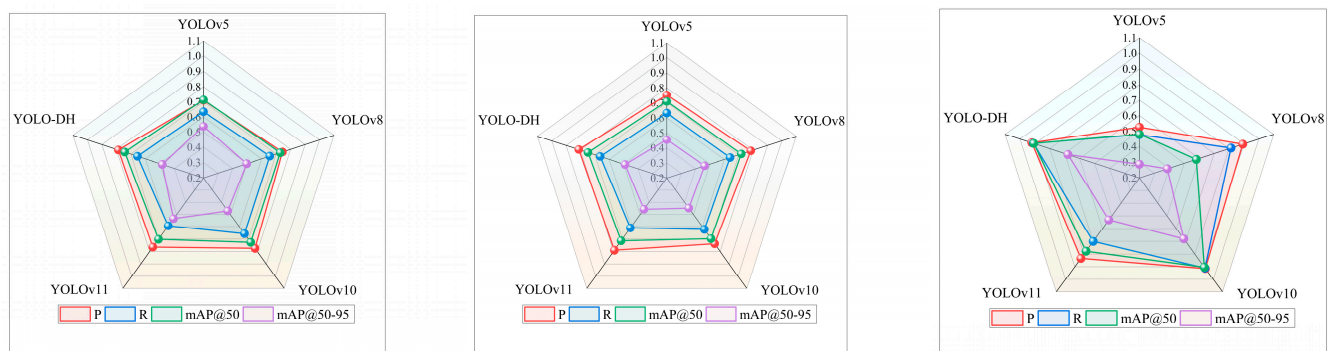
Method	All	Person	Bus	Car	Motorcycle	Bicycle
a	0.718	0.789	0.634	0.853	0.678	0.644
b	0.725	0.787	0.643	0.855	0.683	0.658
c	0.747	0.814	0.673	0.868	0.707	0.675
d	0.727	0.798	0.628	0.851	0.695	0.654
e	0.748	0.818	0.647	0.868	0.718	0.687

3.3.2. Comparative Analysis with Other Methods

Table 3 presents the quantitative comparison results between the proposed YOLO-DH and other detection models. Overall, our model outperforms other methods across most metrics on three datasets. On the COCO2017 dataset, YOLO-DH achieves the best performance in Precision (P) and mAP@50, with significant improvements of 2.2% and 1.4% in P and mAP@50 metrics compared to the runner-up model. On the RTTS adverse weather dataset, YOLO-DH outperforms all other comparison models across all metrics, achieving improvements of 2.2% in Precision (P), 2.3% in Recall (R), 3% in mAP@50, and 2.7% in mAP@50-95 compared to the second-best model. On the KITTI dataset, the proposed model remains the top performer, surpassing all competitors. For more details, please refer to Table 4, Figures 6 and 7.

Table 4. Comparative study of the proposed model with state-of-the-art methods on three publicly available datasets.

Method	COCO2017				RTTS				KITTI			
	P	R	mAP@50	mAP@50-95	P	R	mAP@50	mAP@50-95	P	R	mAP@50	mAP@50-95
YOLOv5	0.709	0.634	0.713	0.537	0.752	0.632	0.709	0.455	0.524	0.475	0.478	0.287
YOLOv8	0.744	0.657	0.729	0.498	0.784	0.64	0.718	0.463	0.892	0.813	0.581	0.387
YOLOv10	0.775	0.657	0.727	0.47	0.738	0.622	0.697	0.447	0.914	0.91	0.903	0.681
YOLOv11	0.765	0.593	0.702	0.534	0.79	0.61	0.714	0.457	0.834	0.701	0.778	0.531
YOLO-DH	0.787	0.655	0.743	0.484	0.812	0.663	0.748	0.49	0.921	0.91	0.908	0.679

**Figure 6.** Experimental results radar chart for three publicly available datasets.

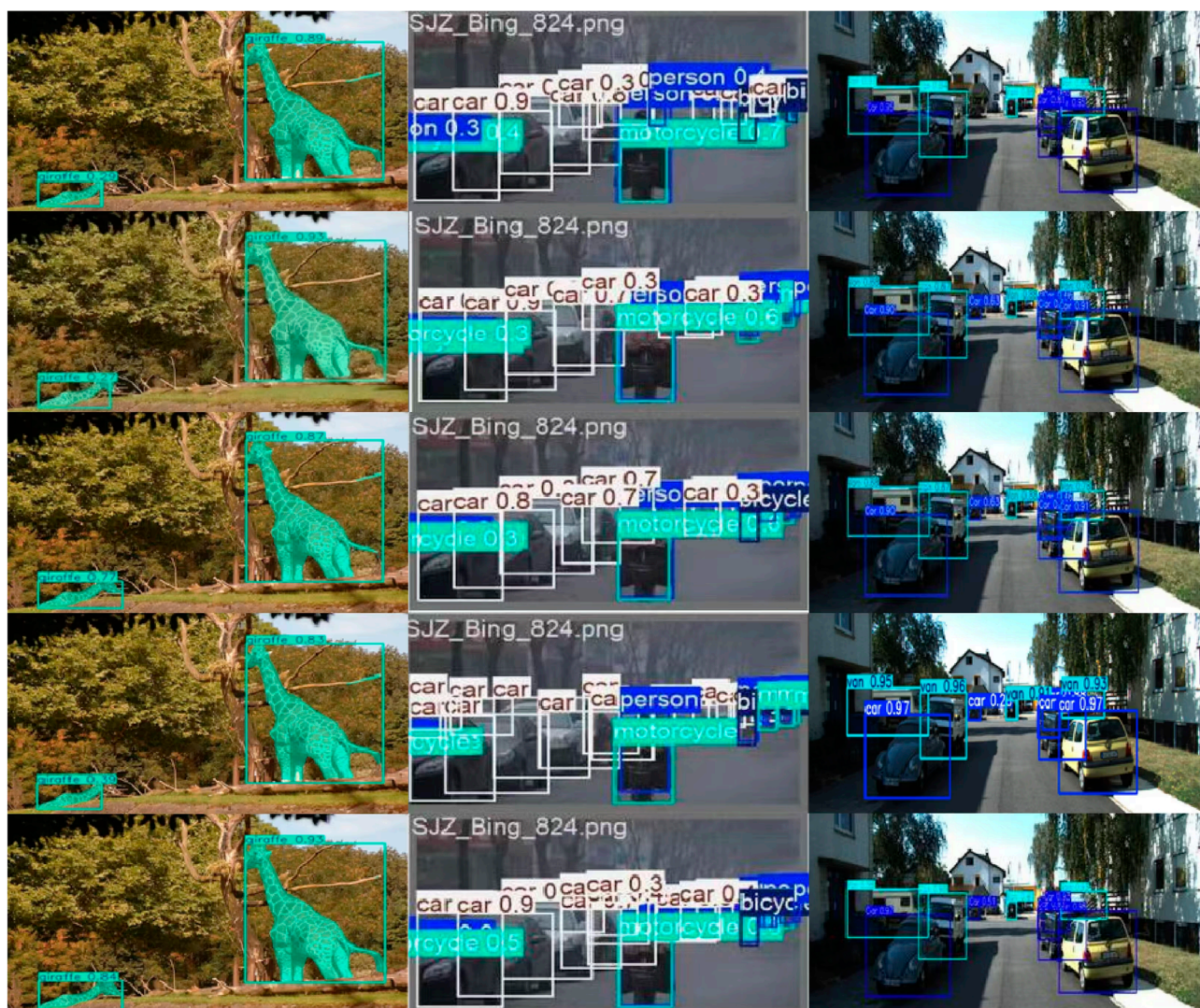


Figure 7. Performance comparison of three publicly available datasets.

We added qualitative visualizations to illustrate the perceptual improvements achieved by the proposed model. Specifically, for selected scenes under adverse weather conditions, we present images before and after dehazing, alongside corresponding detection results. These visual examples clearly demonstrate enhanced image clarity and improved detection of target objects such as vehicles and pedestrians. The figures complement the quantitative metrics reported in the experiments, allowing readers to visually appreciate the effectiveness of our preprocessing and detection pipeline under challenging weather scenarios.

4. Conclusions and Future Work

In this paper, we propose a novel Weather-aware Threshold Channel Attention Module (WTCAM) to enhance sensitivity to channel-wise features in adverse weather images. Additionally, we introduce the DeHazing Module for Detection (DHMD) for the first time, which is integrated into both the backbone network and detection head of YOLOv8, significantly improving dehazing performance. The resulting model, named YOLO-DH. Extensive experimental results demonstrate that our proposed method exhibits strong generalization capabilities across three publicly available datasets. Although it introduces a slight increase in model complexity and hardware requirements, the approach achieves a significant improvement in accuracy. In future work, we aim to reduce the loss in

model complexity while maintaining the substantial accuracy gains, thereby enhancing its practicality for real-world deployment scenarios.

Despite the promising results achieved, this study still has several limitations. First, the experimental evaluation is primarily conducted under daytime and moderate-visibility conditions; thus, the model's performance under nighttime scenarios or extreme low-visibility environments (e.g., heavy fog, blizzards, or glare interference) has not been extensively validated. Second, while the proposed YOLO-DH framework demonstrates robustness on benchmark datasets, real-world deployment scenarios may involve domain shifts such as sensor variations and dynamic lighting conditions that could affect generalization. Finally, the current model introduces a moderate increase in computational cost, which may limit its direct application in embedded or real-time edge devices. These limitations will be addressed in future work through dataset expansion, domain adaptation strategies, and lightweight model optimization.

Author Contributions: L.J.: Supervision, Writing—original draft, Investigation, Project administration. G.M.: Resources, Software, Validation, Writing—review and editing. W.G.: Conceptualization, Methodology, Visualization. Y.S.: Writing—review and editing, Funding acquisition, Data curation. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Research Foundation of Liaoning Province (No. LJKQZ20222436).

Data Availability Statement: Derived data supporting the findings of this study are available from the corresponding author on request.

Conflicts of Interest: The authors affirm that they have no financial interests or personal relationships that could potentially bias the work reported in this paper.

References

1. Kolar, Z.; Chen, H.; Luo, X. Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images. *Autom. Constr.* **2018**, *89*, 58–70. [\[CrossRef\]](#)
2. Wang, S.; Jiang, F.; Zhang, B.; Ma, R.; Hao, Q. Development of UAV-Based Target Tracking and Recognition Systems. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 3409–3422. [\[CrossRef\]](#)
3. Qayyum, A.; Usama, M.; Qadir, J.; Al-Fuqaha, A. Securing Connected & Autonomous Vehicles: Challenges Posed by Adversarial Machine Learning and the Way Forward. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 998–1026.
4. Wang, Y.; Cai, P.; Lu, G. Cooperative autonomous traffic organization method for connected automated vehicles in multi-intersection road networks. *Transp. Res. Part C Emerg. Technol.* **2020**, *111*, 458–476. [\[CrossRef\]](#)
5. Tuncali, C.E.; Fainekos, G.; Prokhorov, D.; Ito, H.; Kapinski, J. Requirements-Driven Test Generation for Autonomous Vehicles with Machine Learning Components. *IEEE Trans. Intell. Veh.* **2020**, *5*, 265–280. [\[CrossRef\]](#)
6. Bie, M.; Liu, Y.; Li, G.; Hong, J.; Li, J. Real-time vehicle detection algorithm based on a lightweight You-Only-Look-Once (YOLOv5n-L) approach. *Expert Syst. Appl.* **2023**, *213*, 119108. [\[CrossRef\]](#)
7. Han, Y.; Suo, J.; Zhang, B.; Zhao, J.; Cheng, Y.; Dai, Q. Real-Time Sketching of Harshly Lit Driving Environment Perception by Neuromorphic Sensing. *IEEE Trans. Intell. Veh.* **2024**, *9*, 7885–7899. [\[CrossRef\]](#)
8. Hnewa, M.; Radha, H. Object Detection Under Rainy Conditions for Autonomous Vehicles: A Review of State-of-the-Art and Emerging Techniques. *IEEE Signal Process. Mag.* **2021**, *38*, 53–67. [\[CrossRef\]](#)
9. Hsu, W.-Y.; Lin, W.-Y. Ratio-and-Scale-Aware YOLO for Pedestrian Detection. *IEEE Trans. Image Process.* **2021**, *30*, 934–947. [\[CrossRef\]](#)
10. Ghosh, R. A Faster R-CNN and recurrent neural network based approach of gait recognition with and without carried objects. *Expert Syst. Appl.* **2022**, *205*, 117730. [\[CrossRef\]](#)
11. Yu, J.; Ye, X.; Tu, Q. Traffic Sign Detection and Recognition in Multiimages Using a Fusion Model with YOLO and VGG Network. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 16632–16642. [\[CrossRef\]](#)
12. Zheng, X.; Li, S.; Li, Y.; Duan, D.; Yang, L.; Cheng, X. Confidence Evaluation for Machine Learning Schemes in Vehicular Sensor Networks. *IEEE Trans. Wirel. Commun.* **2023**, *22*, 2833–2846. [\[CrossRef\]](#)
13. Khalvati, K.; Kiani, R.; Rao, R.P. Bayesian inference with incomplete knowledge explains perceptual confidence and its deviations from accuracy. *Nat. Commun.* **2021**, *12*, 5704. [\[CrossRef\]](#)

14. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6848–6856.
15. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861. [\[CrossRef\]](#)
16. Rocky, A.; Wu, Q.J.; Zhang, W. Review of accident detection methods using dashcam videos for autonomous driving vehicles. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 8356–8374. [\[CrossRef\]](#)
17. Smith, T.M.; Lakshmanan, V.; Stumpf, G.J.; Ortega, K.L.; Hondl, K.; Cooper, K.; Calhoun, K.M.; Kingfield, D.M.; Manross, K.L.; Toomey, R.; et al. Multi-Radar Multi-Sensor (MRMS) Severe Weather and Aviation Products: Initial Operating Capabilities. *Bull. Am. Meteorol. Soc.* **2016**, *97*, 1617–1630. [\[CrossRef\]](#)
18. Gerstmair, M.; Melzer, A.; Onic, A.; Huemer, M. On the Safe Road Toward Autonomous Driving: Phase Noise Monitoring in Radar Sensors for Functional Safety Compliance. *IEEE Signal Process. Mag.* **2019**, *36*, 60–70. [\[CrossRef\]](#)
19. Li, Y.; Xing, L.; Wang, W.; Wang, H.; Dong, C.; Liu, S. Evaluating impacts of different longitudinal driver assistance systems on reducing multi-vehicle rear-end crashes during small-scale inclement weather. *Accid. Anal. Prev.* **2017**, *107*, 63–76. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Liu, Z.; Cai, Y.; Wang, H.; Chen, L.; Gao, H.; Jia, Y.; Li, Y. Robust Target Recognition and Tracking of Self-Driving Cars with Radar and Camera Information Fusion Under Severe Weather Conditions. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 6640–6653. [\[CrossRef\]](#)
21. Wang, H.; Chen, Y.; Cai, Y.; Chen, L.; Li, Y.; Sotelo, M.A.; Li, Z. SFNet-N: An Improved SFNet Algorithm for Semantic Segmentation of Low-Light Autonomous Driving Road Scenes. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 21405–21417. [\[CrossRef\]](#)
22. Yang, T.; Lv, C. A Secure Sensor Fusion Framework for Connected and Automated Vehicles Under Sensor Attacks. *IEEE Internet Things J.* **2022**, *9*, 22357–22365. [\[CrossRef\]](#)
23. Ren, C.; He, X.; Pu, Y.; Nguyen, T.Q. Learning Image Profile Enhancement and Denoising Statistics Priors for Single-Image Super-Resolution. *IEEE Trans. Cybern.* **2021**, *51*, 3535–3548. [\[CrossRef\]](#)
24. Oude Lohuis, M.N.; Pie, J.L.; Marchesi, P.; Montijn, J.S.; de Kock, C.P.; Pennartz, C.M.; Olcese, U. Multisensory task demands temporally extend the causal requirement for visual cortex in perception. *Nat. Commun.* **2022**, *13*, 2864. [\[CrossRef\]](#)
25. Karthikeyan, R.; Mehta, R.K. Towards a closed-loop neurostimulation platform for augmenting operator vigilance. In Proceedings of the 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Toronto, ON, Canada, 11–14 October 2020; pp. 3976–3983.
26. Wang, H.-F.; Jiao, Y.-M.; Hao, T.; Shan, Y.-H.; Song, S.-Z.; Huang, H. Low-Visibility Vehicle-Road Environment Perception Based on the Multi-Modal Visual Features Fusion of Polarization and Infrared. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 11997–12013. [\[CrossRef\]](#)
27. Rizani, F. Image quality improvement using image processing method image brightness contrast and image sharpening. *Multica Sci. Technol. (ACCREDITED-SINTA 5)* **2021**, *1*, 6–12. [\[CrossRef\]](#)
28. Oliveira, M.; Cerqueira, R.; Pinto, J.R.; Fonseca, J.; Teixeira, L.F. Multimodal PointPillars for efficient object detection in autonomous vehicles. *IEEE Trans. Intell. Veh.* **2024**, *10*, 81–91. [\[CrossRef\]](#)
29. Wang, Y.; Xiong, J.; Yan, X.; Wei, M. USCFormer: Unified Transformer with Semantically Contrastive Learning for Image Dehazing. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 11321–11333. [\[CrossRef\]](#)
30. Liu, X.; Shi, Z.; Wu, Z.; Chen, J.; Zhai, G. GridDehazeNet+: An Enhanced Multi-Scale Network with Intra-Task Knowledge Transfer for Single Image Dehazing. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 870–884. [\[CrossRef\]](#)
31. Wang, Y.; Yan, X.; Guan, D.; Wei, M.; Chen, Y.; Zhang, X.-P.; Li, J. Cycle-SNSPGAN: Towards Real-World Image Dehazing via Cycle Spectral Normalized Soft Likelihood Estimation Patch GAN. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 20368–20382. [\[CrossRef\]](#)
32. Mehra, A.; Mandal, M.; Narang, P.; Chamola, V. ReViewNet: A Fast and Resource Optimized Network for Enabling Safe Autonomous Driving in Hazy Weather Conditions. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 4256–4266. [\[CrossRef\]](#)
33. He, Y.; Liu, Z. A Feature Fusion Method to Improve the Driving Obstacle Detection Under Foggy Weather. *IEEE Trans. Transp. Electr.* **2021**, *7*, 2505–2515. [\[CrossRef\]](#)
34. Liu, P.; Zhang, C.; Qi, H.; Wang, G.; Zheng, H. Multi-Attention DenseNet: A Scattering Medium Imaging Optimization Framework for Visual Data Pre-Processing of Autonomous Driving Systems. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 25396–25407. [\[CrossRef\]](#)
35. Wei, X.; Ye, X.; Mei, X.; Wang, J.; Ma, H. Enforcing high frequency enhancement in deep networks for simultaneous depth estimation and dehazing. *Appl. Soft Comput.* **2024**, *163*, 111873. [\[CrossRef\]](#)
36. Liu, Z.; Li, Y.; Wang, Y.; Gao, B.; An, Y.; Zhao, X. Boosting visual recognition in real-world degradations via unsupervised feature enhancement module with deep channel prior. *IEEE Trans. Intell. Veh.* **2024**, *9*, 7208–7221. [\[CrossRef\]](#)
37. Zhou, W.; Zhang, R.; Li, L.; Yue, G.; Gong, J.; Chen, H.; Liu, H. Dehazed image quality evaluation: From partial discrepancy to blind perception. *IEEE Trans. Intell. Veh.* **2024**, *9*, 3843–3858. [\[CrossRef\]](#)

38. Min, X.; Zhai, G.; Gu, K.; Yang, X.; Guan, X. Objective Quality Evaluation of Dehazed Images. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 2879–2892. [[CrossRef](#)]
39. Donoho, D.L.; Johnstone, I.M. Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **1994**, *81*, 425–455. [[CrossRef](#)]
40. Donoho, D.L.; Johnstone, I.M. Adapting to Unknown Smoothness via Wavelet Shrinkage. *J. Am. Stat. Assoc.* **1995**, *90*, 1200–1224. [[CrossRef](#)]
41. Chen, Z.; Lei, X.; Yuan, Q.; Qi, Y.; Ma, Z.; Qian, S.; Lyu, X. Key technologies for autonomous fruit-and vegetable-picking robots: A review. *Agronomy* **2024**, *14*, 2233. [[CrossRef](#)]
42. Lu, L.; Xiong, Q.; Xu, B.; Chu, D. Mixdehazenet: Mix structure block for image dehazing network. In Proceedings of the 2024 International Joint Conference on Neural Networks (IJCNN), Yokohama, Japan, 30 June–5 July 2024; pp. 1–10.
43. Chan, K.H.; So, S.K. Using admittance spectroscopy to quantify transport properties of P3HT thin films. *J. Photonics Energy* **2011**, *1*, 011112. [[CrossRef](#)]
44. Guo, L.; Zhou, X.; Zhao, Y.; Wu, W. Improved YOLOv7 algorithm incorporating InceptionNeXt and attention mechanism for vehicle detection under adverse lighting conditions. *Signal Image Video Process.* **2025**, *19*, 299. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.