scientific reports

OPEN

A target detection model HR-YOLO for advanced driver assistance systems in foggy conditions

Yao Zhang & Na Jia✉

To improve the accuracy and real-time performance of detection algorithms in Advanced Driver Assistance Systems (ADAS) under foggy conditions, this paper introduces HR-YOLO, an improved YOLO-based model specifically designed for vehicle and pedestrian detection. To enhance detection performance under complex meteorological conditions, several critical modules have been optimized. First, the Efficient High-Precision Defogging Network (EHPD-Net) is introduced to strengthen feature extraction. Inspired by the Global Attention Mechanism (GAM), the Enhanced Global-Spatial Attention (EGSA) module is incorporated to effectively improve the detection of small targets in foggy conditions. Second, the Depth-Normalized Defogging Network (DND-Net) is applied to enhance image quality. Additionally, the Dynamic Sample (DySample) module is integrated into the neck network, complemented by optimizations to the convolution and C2f modules, which significantly improve feature fusion efficiency. Furthermore, The Wise Intersection over Union (WIoU) loss function is introduced to enhance target localization accuracy. The robustness and accuracy of HR-YOLO were validated through experiments on two foggy weather datasets: RTTS and Foggy Cityscapes. The results indicate that HR-YOLO achieved a mean Average Precision (mAP) of 79.8% on the RTTS dataset, surpassing the baseline by 5.9%. On the Foggy Cityscapes dataset, it achieved an mAP of 49.5%, representing a 9.7% improvement over the baseline. This model serves as an effective solution for target detection tasks under foggy conditions and establishes a foundation for future advancements in this field.

Keywords Object detection, Advanced driver assistance systems (ADAS), Foggy conditions, Image enhancement, Dehazing networks

With the rapid development of autonomous driving, ADAS have become essential technologies for improving road safety and supporting drivers. ADAS employs various sensing technologies, including cameras and radars, to monitor the surrounding environment in real-time and assist drivers in identifying potential hazards[1]. The visual perception module, particularly camera-based object detection algorithms, plays a pivotal role in ADAS. By leveraging these algorithms, ADAS can identify vehicles, pedestrians, and other traffic obstacles from captured images, enabling timely responses in dynamic traffic scenarios[2].

However, existing object detection algorithms face significant limitations in terms of accuracy and robustness[3–5]. Their real-time applicability in complex traffic scenarios remains suboptimal, particularly under adverse environmental conditions such as fog[6–8]. Fog significantly reduces visibility and degrades image quality, hindering conventional object detection algorithms from accurately capturing target features. As a result, key targets, including vehicles and pedestrians, are often misidentified or missed[9–11]. In high-speed driving scenarios, such inaccuracies can directly compromise traffic safety and potentially lead to accidents[12,13]. Traditional approaches mainly rely on image defogging techniques to mitigate fog-induced image degradation[14]. However, these methods typically require substantial computational resources and often underperform in dynamically changing traffic environments[15].Therefore, developing an object detection algorithm that can effectively address adverse weather conditions while simultaneously achieving high detection accuracy and real-time performance has become a critical focus and ongoing challenge in ADAS research.

To tackle the challenges of object detection under foggy conditions, this paper presents HR-YOLO, an innovative model specifically designed for foggy-weather object detection. The model integrates a novel backbone network, EHPD-Net, and a lightweight unsupervised defogging network, DND-Net, significantly enhancing

College of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin 150040, China. ✉email: Jiana@nefu.edu.cn

detection performance in foggy environments. Compared to traditional YOLO-based algorithms, HR-YOLO achieves superior accuracy under adverse weather conditions while maintaining high detection efficiency.

The contributions of this paper are summarized as follows:

(1) We propose an innovative haze weather target detection backbone network EHPD-Net, which combines Swin Transformer, DWConv and EGSA modules. Compared with the current mainstream defog detection framework, this framework can still improve the feature extraction capability of the backbone network under low-quality pictures while maintaining a small computing overhead.

(2) Inspired by AOD-Net, we proposed a new lightweight unsupervised defogging network DND-Net and applied it in the preprocessing process. By efficiently integrating important features at different scales, we provide higher quality feature details to the modeling part of the network.

(3) We propose a more efficient neck detection network and introduce DySample module, DWConv module and C2 module. This combination method can better utilize multi-scale features when dealing with haze weather detection tasks, avoid interference from redundant information, and optimize identification efficiency.

(4) The WIoU loss function is introduced, combined with the dynamic non-monotonic focusing mechanism, to accurately deal with the target positioning problems in haze images, and significantly improve detection accuracy and robustness.

The rest of this paper is organized as follows. "Related Work" reviews the state-of-the-art object detection methods in foggy environments and discusses related advancements. "Methodology" provides a detailed description of the proposed HR-YOLO model. "Experiments and Analysis" validates the model's effectiveness and robustness through experimental evaluations. Finally, "Conclusion" concludes the work and outlines directions for future research.

Related work

Development of image defogging methods

Object detection and recognition in foggy environments have long been a focal point in visual measurement, particularly in both civilian and military applications[16–18]. Prior to the emergence of deep learning algorithms[19–27], much of the research focused on traditional image processing techniques, such as image enhancement and restoration methods, to recover fog-degraded images. Guo et al.[28] proposed a lightweight deep network that restores low-light images by adjusting their dynamic range. Zhang et al.[29] introduced a multi-level fusion module that utilizes complementary relationships between low-level and high-level features, enabling the restoration of fog-free images without the need for estimating atmospheric light intensity. Balla et al.[30] developed a 14-layer residual convolutional neural network that extracts features indicative of fog intensity to restore fog-free images. Xiao et al.[31] proposed a dehazing algorithm that directly learns the residual haze image through an end-to-end approach, effectively removing haze from blurred images. Zhang et al.[32], leveraging the concept of the dark channel prior, employed an approximate method to estimate atmospheric light and transmission, improving dehazing performance while significantly enhancing computational efficiency.

Although these methods can partially restore image clarity, they often require manual parameter adjustment for different scenarios. Moreover, their robustness is still insufficient, particularly in dynamic traffic environments with low visibility and strong background noise.

Advances in deep learning-based object detection algorithms

With the rapid advancement of deep learning, object detection algorithms have been categorized into two main types: two-stage and one-stage detection methods. Two-stage algorithms, exemplified by R-CNN[33], Fast R-CNN[34], and Faster R-CNN[35], extract object information by generating region proposals followed by precise classification and localization. These methods demonstrate high detection accuracy across diverse environments. However, their inherent complexity often results in slower processing speeds and higher computational requirements, posing significant challenges for real-time applications of ADAS.

To address these limitations, one-stage object detection algorithms emerged, with You Only Look Once (YOLO)[36]standing out as one of the most prominent approaches. YOLO redefines object detection as a regression task, simultaneously predicting object locations and categories in a single forward pass. This paradigm shift has led to substantial improvements in detection speed. Subsequent iterations of YOLO, including YOLOv2[37], YOLOv3[38], YOLOv4[39], and YOLOv5[40], have achieved notable advancements in real-time performance and computational efficiency. As a result, these methods have become widely adopted in fields such as video surveillance, intelligent transportation, and autonomous driving[41].

Despite these successes, applying YOLO to object detection tasks under foggy conditions has proven challenging. The presence of fog often obscures or blurs target features, compromising YOLO's ability to accurately detect objects. Consequently, its performance in such adverse weather conditions remains suboptimal, necessitating further advancements to improve robustness and accuracy in foggy environments.

Integrated methods combining dehazing and object detection

To overcome the challenges associated with object detection in foggy weather, researchers have increasingly explored integrating dehazing techniques with deep learning-based object detection algorithms. By incorporating an image dehazing module, these methods aim to enhance the accuracy and robustness of object detection under adverse weather conditions.Wang et al.[42] proposed a framework specifically tailored for YOLO, which integrates a Quasi-Translation Network (QTNet) and a Feature Calibration Network (FCNet). This framework adapts object detection models to adverse weather domains by bridging the gap between normal and adverse weather conditions. Zhang et al.[43] introduced a multi-class object detection approach that jointly trains visibility enhancement, object classification, and object localization tasks within a dehazing network. Unlike traditional standalone dehazing methods, this end-to-end learning paradigm significantly improves detection accuracy and demonstrates robust adaptability in complex traffic environments. Li et al.[44] developed a Stepwise Domain Adaptive YOLO (S-DAYOLO) framework, which bridges domain discrepancies through the construction of an auxiliary domain. This approach incrementally learns domain transitions, thereby enhancing the detection accuracy and robustness of YOLO under foggy conditions. Despite its effectiveness, this method imposes significant computational demands and requires highly curated datasets, which may hinder its practicality in real-world applications.

In contrast to these existing methods, our approach not only employs an enhanced dehazing module for preprocessing but also incorporates a highly efficient backbone network and redesigned neck network to further improve detection capabilities. Moreover, by leveraging both natural fog and synthetic fog datasets, our method achieves significant improvements in detection accuracy, offering robust performance across diverse foggy conditions.

Methodology

HR-YOLO network architecture

This study introduces HR-YOLO, a high-precision object detection network specifically tailored for driving scenarios under foggy conditions. Built upon YOLOv8[45], the HR-YOLO architecture is depicted in Fig. 1. The preprocessing stage is managed by DND-Net, which enhances the visual quality of foggy images by restoring them to a clarity comparable to fog-free conditions. These preprocessed images are then passed to the backbone network.

The backbone EHPD-Net, incorporates DWConv, EGSA, and Swin Transformer modules. This network is designed to extract multi-scale features efficiently, perform down-sampling, and adjust channels, ensuring robust feature representation. The extracted features are subsequently forwarded to the neck network for further refinement. In the neck network, feature fusion strategies are employed to emphasize relevant target features in foggy conditions. By integrating the DySample up-sampling module, the DWConv module, and the C2f module, the network significantly enhances its ability to extract features from small objects. Finally, the detection head generates the classification confidence scores and location regression results of the detected objects. Within ADAS, HR-YOLO facilitates real-time decision-making by generating actionable commands based on detection outcomes. This approach not only improves driving safety but also optimizes overall driving efficiency.

EHPD-Net design

In foggy weather, reduced visibility poses significant challenges to extracting background targets, necessitating a backbone network with robust feature extraction capabilities for object detection algorithms. For HR-YOLO, the backbone network is required to efficiently capture sufficient global information within a limited time frame to ensure accurate target identification under foggy conditions, while simultaneously retaining detailed local features for precise classification.

Traditional Convolutional Neural Network (CNN) are inherently constrained in their ability to comprehensively extract contextual information, especially under adverse weather conditions[46]. To address these limitations, we designed EHPD-Net, a backbone network optimized for feature extraction in foggy environments. EHPD-Net leverages a combination of key modules to achieve its superior performance. The

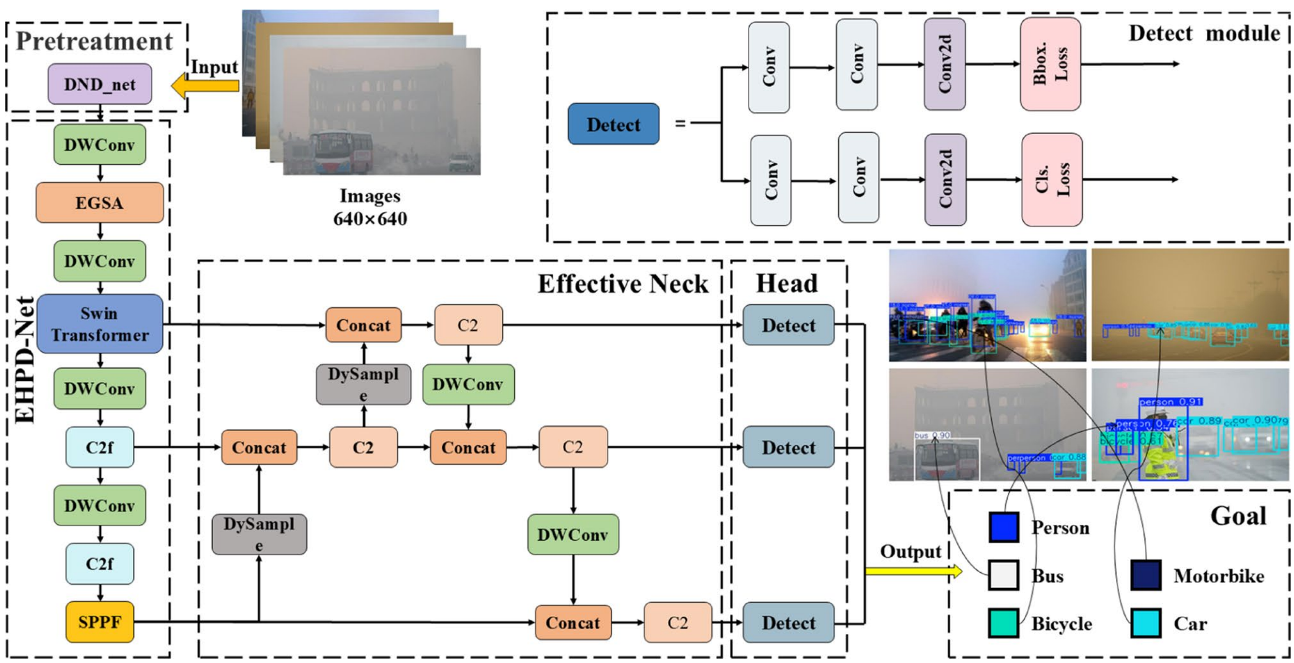


Fig. 1. HR-YOLO Architecture.

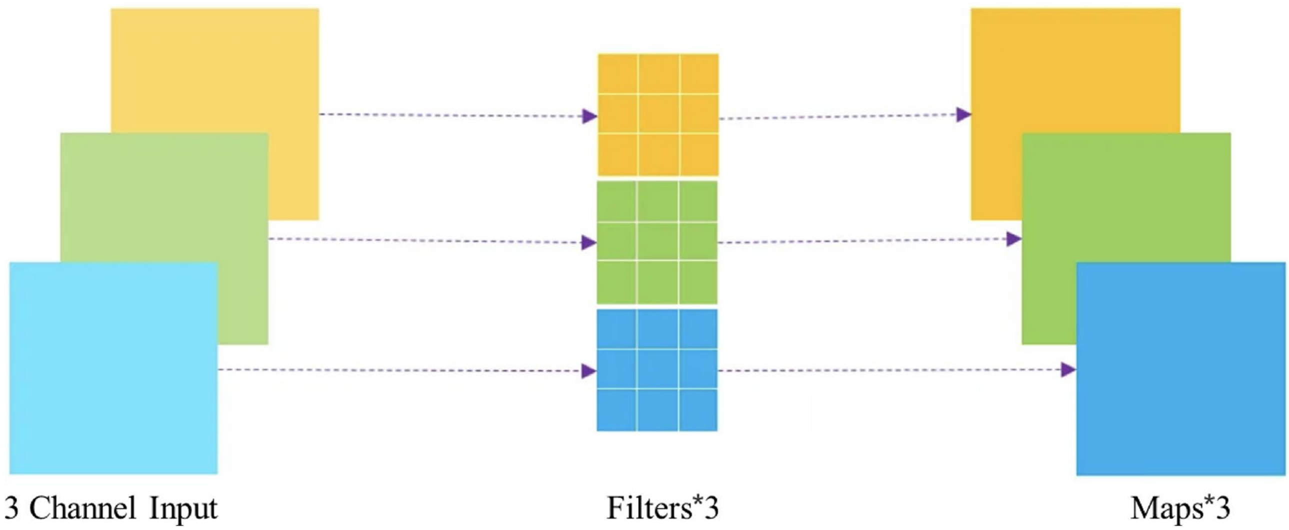


Fig. 2. Convolution kernel design of DWConv.

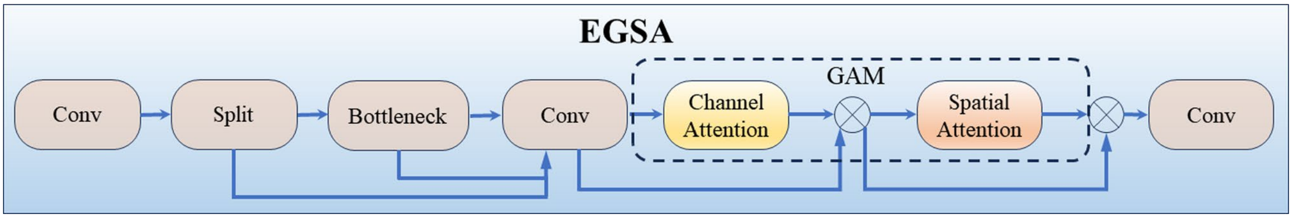


Fig. 3. Network architecture of the GAM attention mechanism.

DWConv module is introduced to enhance detection speed, while EGSA module significantly improves detection accuracy. Additionally, the Swin Transformer module is utilized to capture global key feature points, ensuring high-quality feature representations without increasing computational overhead. The synergistic integration of these modules enables EHPD-Net to deliver enhanced feature extraction capabilities compared to conventional backbone networks. This design demonstrates superior performance in foggy conditions, making it a robust solution for high-precision object detection tasks.

a. DWConv Block.

To accelerate the recognition speed of the backbone network, this study introduces an optimized and reallocated DWConv module. The primary advantage of DWConv lies in its unique convolution kernel design, as depicted in Fig. 2. For the three channels of an input feature map, each convolution kernel operates exclusively on a single channel. This design ensures that the number of output channels remains consistent with the input channels, thereby contributing to a streamlined computational pipeline.

The reallocation strategy of the DWConv module enhances feature extraction efficiency by minimizing interactions among redundant information across channels. Compared to traditional convolutional methods, DWConv significantly reduces both computational complexity and parameter count, achieving superior computational efficiency. Furthermore, this method maintains high recognition accuracy even under the challenging conditions of foggy environments, demonstrating its robustness and effectiveness.

b. EGSA Module.

In foggy weather, targets in images often exhibit low pixel intensities, which poses significant challenges for feature extraction. As EHPD-Net processes deeper layers, there is a risk of neglecting or losing the features of small-scale targets. Attention mechanisms effectively mitigate this issue by dynamically assigning higher weights to critical target features while suppressing irrelevant background information.

The Global Attention Mechanism (GAM) integrates channel and spatial attention strategies, enhancing the representational power of deep neural networks by emphasizing globally relevant features[47]. Inspired by GAM, we propose the EGSA module to replace the original C2f module, as illustrated in Fig. 3. The EGSA module's core innovation lies in iteratively refining feature weights to facilitate meaningful interaction between feature representations. Specifically, one-dimensional convolution (Conv 1D) operations are employed to update weights across feature channels and spatial dimensions.

The input feature map $F \in R^{C\times H\times W}$ is first fed into GAM, where feature interactions occur to produce an updated feature map with the same dimensions. This refined feature map is then forwarded to the next stage, which utilizes two-dimensional convolution (Conv 2D) for further processing. By minimizing information

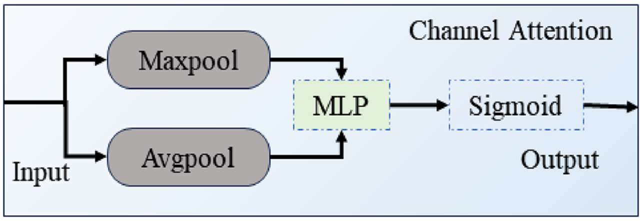


Fig. 4. Channel attention submodule.

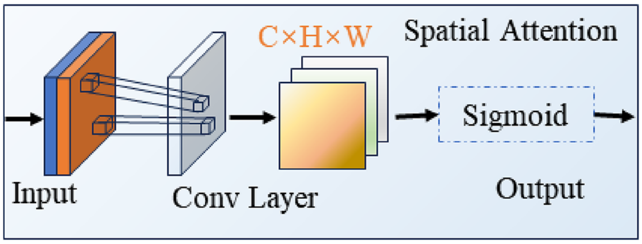


Fig. 5. Spatial Attention Submodule.

loss and enhancing global feature interactions, the GAM attention mechanism enables the model to prioritize channel-specific information more effectively. The output of the GAM attention mechanism, given an input feature map, is formally defined by Eqs. (1) and (2).

$$F_c = W_c(F) \otimes F \tag{1}$$

$$F_s = W_s(F_c) \otimes F_c \tag{2}$$

Where $\otimes$ denotes element-wise multiplication, $\mathrm{F_c}$ represents the output of the channel attention submodule, and $\mathrm{F_s}$ represents the output of the spatial attention submodule.

As shown in Fig. 4, the channel attention submodule preserves the three-dimensional structure of the features through a carefully arranged three-dimensional representation. Initially, the feature map $F \in \mathrm{R}^{C\times H\times W}$ undergoes a dimensional transformation, resulting in a new feature map with dimensions [C, H, W], where C, H, and W denote the number of channels, height, and width, respectively. This transformed feature map is subsequently processed by a Multi-Layer Perceptron (MLP). The MLP employs a sigmoid activation function to calculate the channel-wise attention weights, thereby refining the feature representation and enhancing the discriminative power of the relevant channels.

As illustrated in Fig. 5, the spatial attention submodule is designed to emphasize spatial information by employing two convolutional layers for spatial feature fusion. First, a convolution operation with a kernel size of 7 is applied to the input feature map, reducing computational complexity and transforming its dimensions from [C, H, W] to [$\frac{\mathrm{C}}{\mathrm{r}}$, H, W]. Next, an additional convolution operation increases the number of channels, ensuring that the output channel dimensions remain consistent with the input. Finally, the refined feature map undergoes a sigmoid activation function, which computes spatial attention weights to prioritize spatially significant regions in the feature map.

To visually demonstrate the effectiveness of the EGSA module, we generated heatmaps on Foggy Cityspaces dataset, as depicted in Fig. 6. Figure 6(b) presents the heatmap of the C2f module's output, which reflects the network's initial focus on the target features. In contrast, Fig. 6(c) displays the heatmap of the EGSA module's output, where red regions denote areas of highestf significance, highlighting the regions where the model concentrates most during decision-making. Notably, after incorporating the EGSA module, critical target information becomes more pronounced in the heatmap. This observation underscores the EGSA module's ability to effectively enhance the network's attention to essential features. By integrating the EGSA module, the EHPD-Net achieves a more precise focus on feature map information across various channels and spatial dimensions, thereby significantly improving the detection accuracy of small objects under foggy weather conditions.

c. Swin Transformer Block.

Under haze weather conditions, the traditional CNN architecture adopted by YOLOv8 expands the receptive field by stacking convolutional layers, but because it relies on local operations, redundant features (such as background noise) are often extracted, and key features such as target edges are insufficiently captured, resulting in an increase in the false detection rate. To this end, HR-YOLO introduces a Transformer structure to enhance global modeling capabilities. Swin Transformer[48] has become an ideal choice because its sliding window mechanism reduces the computational complexity while strengthening target context information extraction through interaction between windows. We replaced the Transformer part of EHPD-Net with Swin

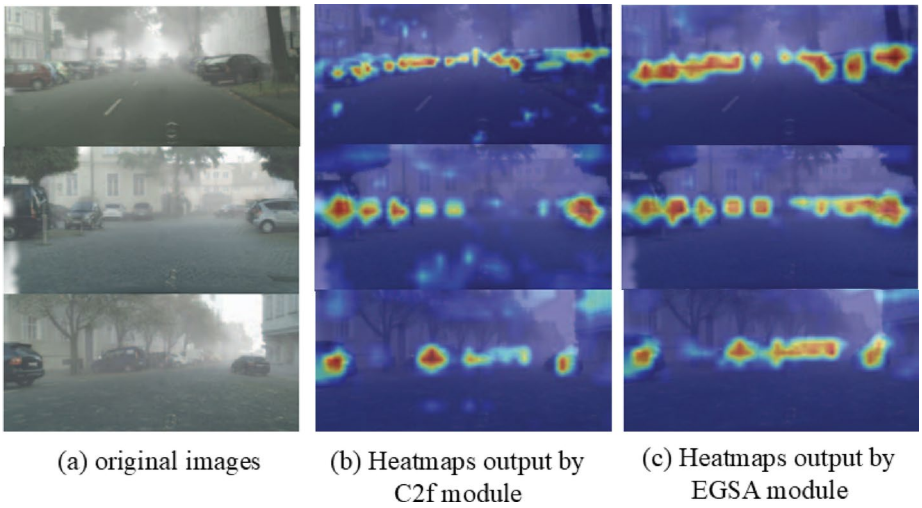


Fig. 6. Heatmaps of outputs from the C2f and EGSA modules.

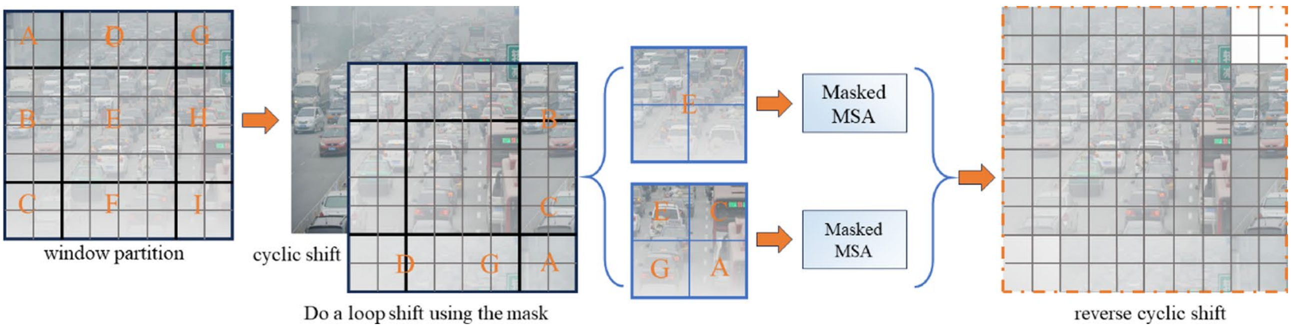


Fig. 7. Swin Transformer architecture with sliding window attention.

Transformer to achieve more efficient and accurate feature extraction in haze scenarios, significantly improving the robustness of target detection.

Swin Transformer adopts a hierarchical pyramid architecture, and its core innovation lies in the ability to model dynamic feature interactions based on the self-attention mechanism of local windows and cross-windows interactions. As shown in Fig. 7, the model dynamically generates an attention weight matrix when performing standard self-attention calculations within each window to adaptively adjust feature response intensity according to semantic content. This method of attention calculation in local window breaks the limitations of traditional Transformers, and significantly improves computing efficiency while maintaining global modeling capabilities. In order to further break the window boundary limitation, the model introduces a shift window division strategy. By alternately using the SW-MSA module of the regular window and the offset window, a two-layer cross-window feature interaction mechanism is built, so that the feature information of adjacent windows can be dynamically flowed and semantic fusion deep. In terms of hierarchical structure, Swin Transformer performs multi-scale aggregation of shallow local features and deep global semantics, ultimately forming a hierarchical feature representation that takes into account both fine-grained details and high-level semantics.

Based on this sliding window strategy, the construction of the Swin Transformer can be formalized as Eqs. (3)–(6):

$$\hat{z}^l = W - SMA\left(LN\left(z^{l-1}\right)\right) + z^{l-1} \tag{3}$$

$$z^l = MLP\left(LN\left(\hat{z}^l\right)\right) + \hat{z}^l \tag{4}$$

$$\hat{z}^{l+1} = W - SMA\left(LN\left(z^l\right)\right) + z^l \tag{5}$$

$$z^{l+1} = MLP\left(LN\left(\hat{z}^{l+1}\right)\right) + \hat{z}^{l+1} \tag{6}$$

Here, $\hat{z}^l$and z^ldenotes the output of the Multi-Head Self-Attention (MSA) module, and MLP refers to the output of the MLP module.

This sliding window mechanism captures the feature information of adjacent windows and indirectly constructs and characterizes global features, thus making up for the limitations of information in local windows, effectively solving the long-distance information attenuation problem caused by traditional CNNs due to the number of stacked layers. During the gradual merger of features in each stage, the receptive field continues to expand, and ultimately achieves the effect of approximate global attention. In haze weather, this slip mechanism can obtain higher quality feature points in the same time compared to the baseline model, while using less computational overhead, significantly enhancing its performance in haze weather.

Depth-Normalized defogging module design

Under foggy weather conditions, raw images are frequently compromised by substantial noise, significantly obstructing the extraction of critical features by object detection algorithms. To overcome this challenge, we propose the incorporation of a defogging module as a preprocessing step to restore high-quality, clear images.

Following a comprehensive evaluation of various techniques, including image denoising[49–51], shadow remova[52–55], and image restoration[56–58], we developed the Depth-Normalized Defogging Network (DND-Net) as the preprocessing module for HR-YOLO. Building upon the atmospheric scattering model and inspired by the lightweight, end-to-end trainable defogging model AOD-Net[59], DND-Net introduces enhancements that enable the generation of high-quality fog-free images in a single forward pass, thereby achieving superior processing efficiency.

The atmospheric scattering model, which forms the theoretical foundation of DND-Net, is represented in Eq. (7):

$$I\left(x\right)\;=\;J\left(x\right)t\left(x\right)\;+\;A\left(1\;-\;t\left(x\right)\right) \tag{7}$$

In this model, x denotes the pixel intensity in the input feature map, $\mathrm{I}\left(\mathrm{x}\right)$ represents the observed pixel value, $\mathrm{J}\left(\mathrm{x}\right)$corresponds to the true scene radiance, A signifies the global atmospheric light intensity, and $\mathrm{t}\left(\mathrm{x}\right)$ denotes the transmission matrix. Within the framework of the atmospheric scattering model, accurately computing the transmission rate and global atmospheric light intensity enables the estimation of the original scene's radiance and illumination conditions, thereby significantly alleviating the degradation effects caused by fog in the image.

DND-Net consists of three primary modules: the Efficient Estimation Module, the Positional Normalization Module (PONO)[60], and the Clear Image Generation Module. The overall architecture of DND-Net is depicted in Fig. 8. Among these, the Efficient Estimation Module serves as the core component, tasked with estimating image depth and relative fog density. This module comprises an input layer, multiple Depthwise separable convolution (DSConv)[61] layers operating at different scales, and intermediate connections. The input layer receives foggy images and forwards them for processing. The DSConv layers perform the color mapping of the input image $F\in\;\mathrm{C}^{R\times\;G\times\;B}$, as mathematically expressed in Eqs. (8)–(11):

$$F_{1a}^{C}\;=W_1\;*\;I^{C}\;+\;B_1,\,C\in[R,G,B] \tag{8}$$

$$F_{1b}\;=W_2\;*\;F_{1a}\;+\;B_2 \tag{9}$$

$$F_{1a}\;=\;[F_{1a}^{R},F_{1a}^{G},F_{1a}^{B}] \tag{10}$$

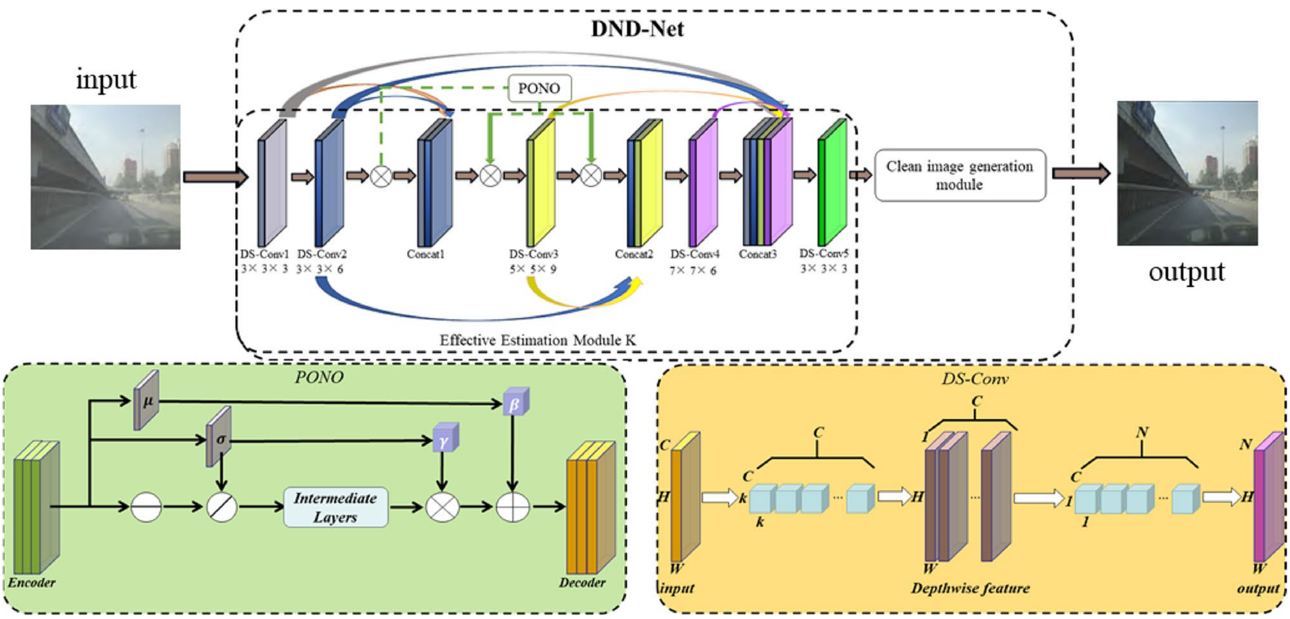


Fig. 8. Architecture of the DND-Net defogging network.

$$F_1 = max\,(0\,,\,F_{1b}\,) \tag{11}$$

Initially, the input feature map undergoes a primary convolution operation to compute the weight coefficient matrix (W_1) and bias matrix (B_1), where (I^{C}) represents the input image's color matrix, and W_1 and B_1 denote the weight and bias parameters of the convolution layer, respectively. This step is followed by fusing color channels to produce the pixel value matrix ($\mathrm{F}_{1\mathrm{a}}^{\mathrm{C}}$). Subsequently, a secondary convolution operation is performed using k -filters of size 1×1, which yields updated weight and bias matrices (W_2) and (B_2). Finally, a modified ReLU activation function is applied in the activation layer to introduce non-linear regression, enabling the extraction of the final output feature map.

The Efficient Estimation Module leverages five DSConv layers and integrates filters of varying sizes to construct multi-scale feature representations. This design enables the network to effectively capture salient image features across different scales. Additionally, the use of intermediate connections compensates for potential information loss during convolution, thereby maximizing the suppression of fog-induced image degradation.

The core principle of the Normalization Module, also referred to as the PONO, is to adapt to feature transformations during the decoding phase. It achieves this by utilizing Moment Shortcut (MS) to combine the extracted mean and standard deviation, generating the normalization parameters β and γ In the encoding phase, PONO applies a normalization strategy to regularize post-convolution features, ensuring consistency in feature representation. To emphasize channel-wise feature extraction, the PONO module is strategically placed after the Conv 2D layer. Following two convolutional layers with kernel size 3×3 the extracted feature information is fed into the PONO module. Here, the module employs a unified normalization approach, formalized in Eqs. (12)–(14), to transform the features into a standardized form, facilitating robust processing and enhanced feature interpretability.

$$\mu_{\,B,H,W} = \frac{1}{c}\sum{}_{C=1}^{c} X_{B,C,H,W} \tag{12}$$

$$\sigma_{\,B,H,W} = \sqrt{\frac{1}{c}\sum{}_{C=1}^{c}\left(X_{B,C,H,W} - \mu_{\,B,H,W}\right)^2 + \epsilon} \tag{13}$$

$$X'_{B,C,H,W} = \gamma\left(\frac{X_{B,C,H,W}-\mu}{\sigma}\right) + \beta \tag{14}$$

Here, μ and σ denote the mean and standard deviation of the feature map, while β and γ represent the updated mean and standard deviation obtained through normalization using the PONO. The normalized feature information generated by the PONO module interacts with the feature map both preceding and succeeding the 5×5 convolutional layer. As a result, the module significantly improves the model's robustness to variations in input conditions.

In foggy weather scenarios, this mechanism allows the DND-Net to effectively adapt to images with varying levels of fog density. Consequently, it enhances the overall defogging model's capability to restore image quality under adverse environmental conditions.

The Clear Image Generation Module comprises element-wise multiplication layers and multiple element-wise addition layers. These layers collaboratively generate restored images by iteratively refining feature representations, as formalized in Eqs. (15) and (16):

$$J\,(x) = K\,(x)\,I\,(x) - K\,(x) + b \tag{15}$$

$$K\,(x) = \frac{\frac{1}{t(x)}\,(I\,(x) - A) + (A - b)}{I\,(x) - 1} \tag{16}$$

Where b represents a constant bias term with a default value of 1. This approach effectively minimizes the reconstruction error between the output image and the ground truth clear image. By incorporating joint estimation of the haze map, it enhances the defogging process, enabling the DND-Net to more accurately restore the lighting conditions and structural details of degraded images. By employing a highly efficient DSConv using only three filters, alongside pose normalization techniques, DND-Net significantly reduces computational complexity while maintaining outstanding performance in restoring fog-free images. This design makes it particularly well-suited for handling foggy weather conditions.

To rigorously evaluate the quality capabilities of DND-NET, we compared its performance with several state-of-the-art delayering methods, including Dark Channel Prior (DCP)[62], Boundary Constraints and Context Regularization (BCCR)[63], Color Decay Prior (CAP)[64], Multi-Scale Convolutional Neural Network (MSCNN)[65], Dehazenet[66], and Originals AOD-Net. Comparative experiments were performed on the HSTS dataset under the RESIED[67] dataset, and the comparison results shown in Fig. 9 clearly demonstrate the excellent performance of DND-NET. Compared with the original foggy images and existing methods, DND-NET exhibits significant clarity and improves recovery of structural details, highlighting its effectiveness in restoring image features and mitigating fog-induced degradation.

A more efficient neck detection network

The original YOLOv8 employs the Path Aggregation Network - Feature Pyramid Network (PAN-FPN) as its neck network to enhance multi-scale feature representation. However, under foggy weather conditions, an increase in the number of targets leads to the growth of irrelevant features, significantly increasing the computational burden. To address this challenge, we propose a more efficient design for the neck network.

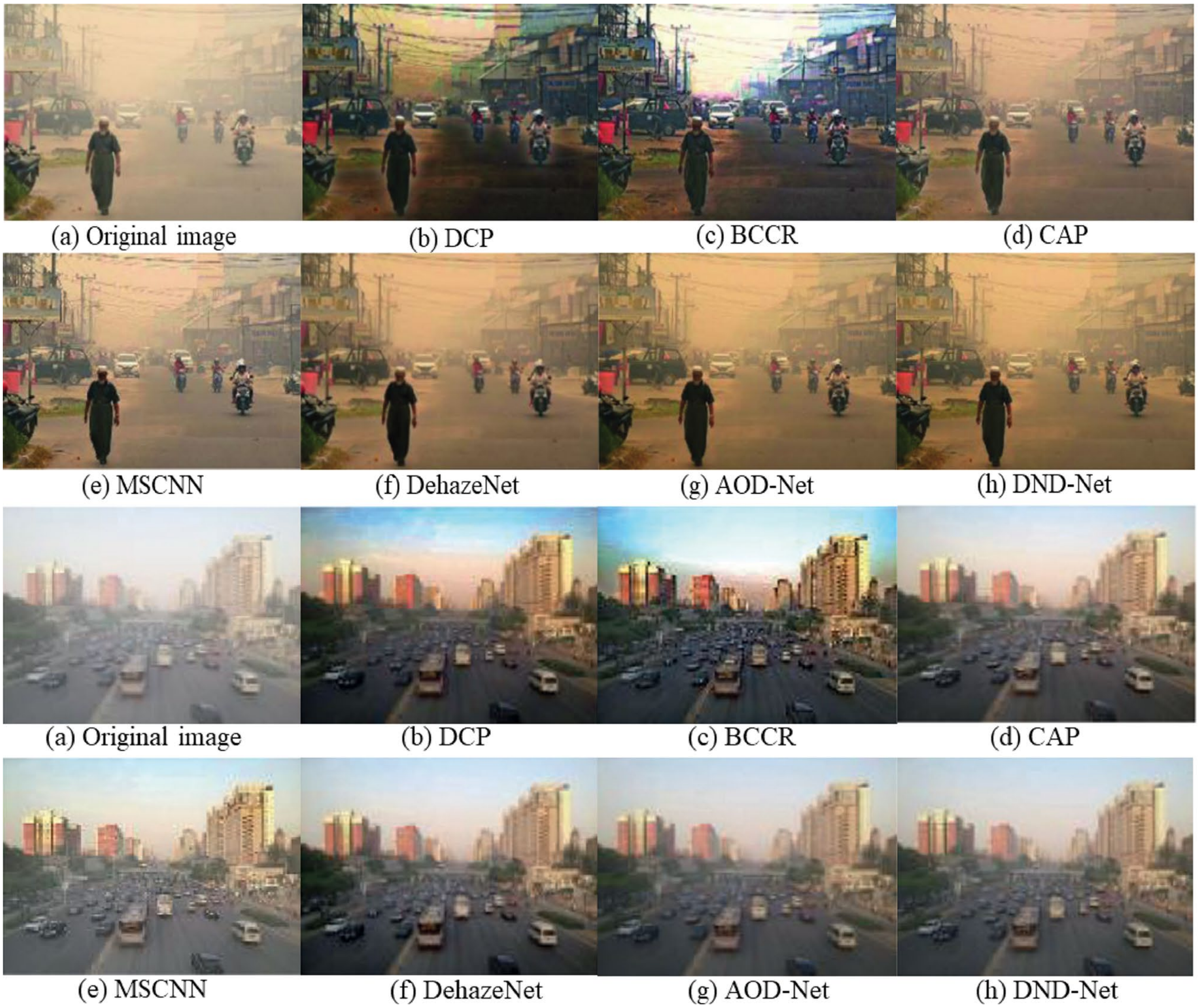


Fig. 9. Comparison of DND-Net with other algorithms.

Specifically, the up-sampling component is replaced with a dynamic sampling-based method[68], as illustrated in Fig. 10. DySample leverages dynamic perceptron to iteratively learn the sampling point coordinates in the input feature map by generating dynamic range factors. This approach produces content-aware sampling points, enabling adaptive resampling of the feature map. Unlike traditional methods that rely on computationally intensive dynamic convolution kernels, Dydymply adopts a more efficient and flexible sampling strategy. In haze weather, this process avoids the complex computing model that relies on dynamic convolution kernels in traditional methods, and adopts a more efficient and flexible sampling strategy that can effectively deal with the numerous redundant and irrelevant information present in haze images. This greatly reduces the inference latency and parameter counting, while achieving a significant improvement in detection accuracy.

In the convolutional part, we continue to continue the lightweight design idea in EHPD-Net and use deep convolution (DWConv) for lightweight design. By decomposing the standard convolution into a cascade operation of deep convolution and point-by-point convolution, the number of parameters and calculation complexity are significantly reduced while ensuring the unchanged receptive field, so as to improve the deployment efficiency of the model on the mobile terminal. For the C2f module, we did not use the original module of YOLOv8, but instead adopted the built-in C2 module of YOLO. By eliminating jump connections and parallel computing paths, it effectively suppressed the activation of irrelevant features caused by haze interference. At the same time, it introduced a serial processing mechanism, so that the network can adaptively focus on discriminant texture information, effectively improving the detection speed while reducing the amount of calculation. In haze weather, the combination of these modules significantly improves the target recognition speed of HR-YOLO, and the detection accuracy of low-virtuality targets is particularly significant.

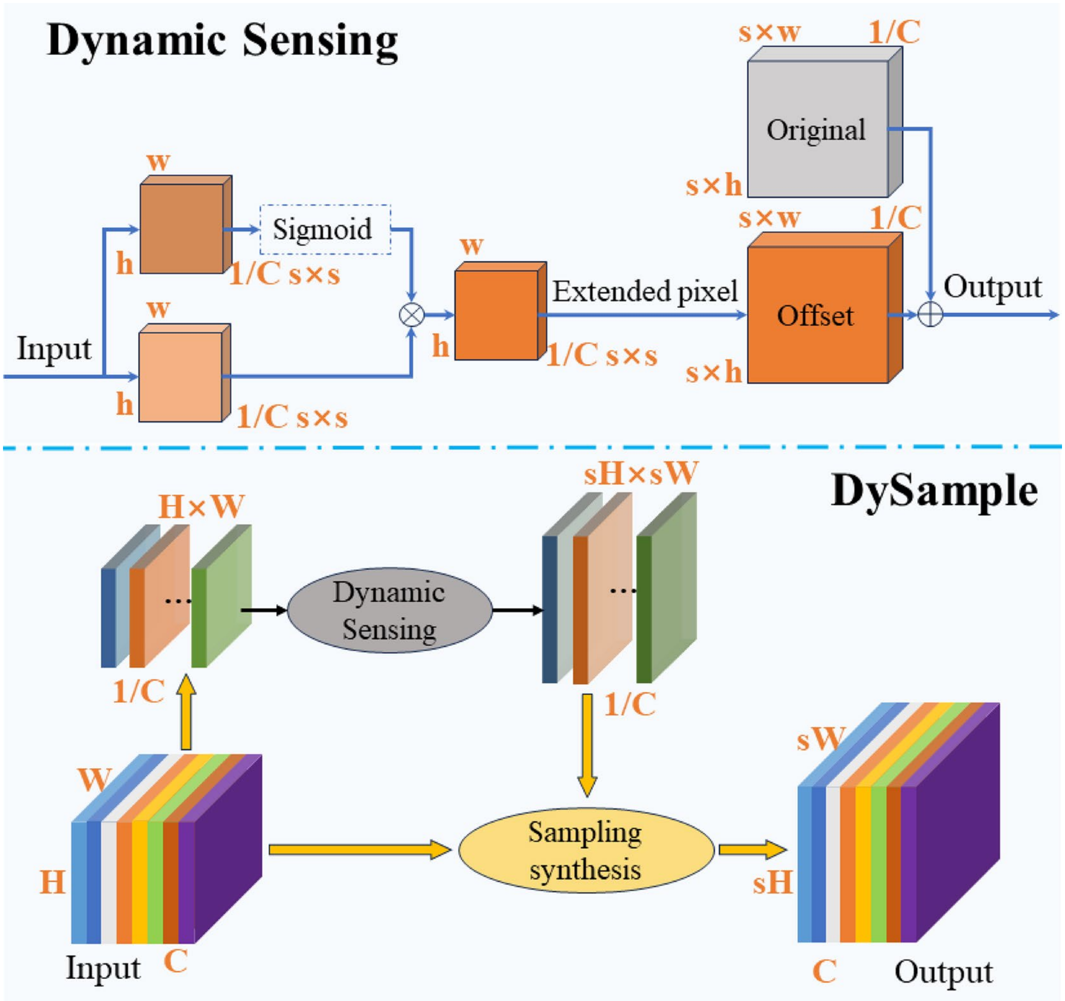


Fig. 10. DySample Network Architecture.

Loss function optimization

In YOLOv8, the bounding box loss employs Complete Intersection over Union (CIoU)[69]. Combining IoU, center point distance and width and height ratios to optimize the matching of the prediction box and the real box. Formulas for CIoU, please see (17)-(19):

$$CIoU = IoU - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha v \tag{17}$$

$$\alpha = \frac{v}{1 - IoU + v} \tag{18}$$

$$v = \frac{4}{\pi^2}\left(arctan\frac{W^{gt}}{h^{gt}} - arctan\frac{W}{h}\right) \tag{19}$$

where $\rho^2(b, b^{gt})$ represents the Euclidean distance between the predicted frame b and the center of the actual frame b^{gt}, and c represents the diagonal distance that just contains the smallest closed area of the predicted frame and the real frame. W^{gt} and h^{gt}represent the width and height of the actual frame of the ground, respectively. This loss increases the matching degree between the prediction box and the real box by increasing the length and width loss, so as to improve the overlap between the prediction frame and the actual frame in the inference stage. However, this method does not consider the sample difficulty balance problem. In smog weather, CIoU may over-punish low-quality samples, reducing model generalization capabilities. To this end, this paper introduces the Wise-IoU (WIoU) loss function[70], which incorporates a dynamic non-monotonic focusing mechanism. WIoU replaces the traditional IoU metric with the “outlier degree” and enhances it with an additional focusing mechanism. This study employs the state-of-the-art WIoU v3, with the corresponding formulas provided in Eqs. (20)–(23).

$$L_{WIoU\ v3} = \tau R_{WIoU} L_{IoU} \tag{20}$$

$$R_{WIoU} = exp\left[\frac{\left(x - x_{gt}\right)^{2} + \left(y - y_{gt}\right)^{2}}{\left(W_{g}^{2} + H_{g}^{2}\right)^{*}}\right],\ R_{WIoU} \in \left[1, e\right) \tag{21}$$

$$\tau = \frac{\beta}{\delta\,\alpha^{\,\beta\,-\delta}} \tag{22}$$

$$\beta = \frac{L_{IoU}^{*}}{\overline{L_{IoU}}} \in \left[0, +\infty\right),\ L_{IoU} \in \left[0,1\right] \tag{23}$$

Here, x_{gt} and y_{gt} represent the center coordinates of the ground truth bounding box, while W_g and H_g denote the width and height of the smallest enclosing box. $R_{WIoU}$is the scaling factor for regular-quality anchor boxes L_{IoU}, and β represents the outlier degree of anchor box quality. Finally, τ is the non-monotonic focusing coefficient, aimed at mitigating the adverse effects of low-quality samples by suppressing harmful gradient contributions. WIoU v3 modulates the relationship between outlier degree (β) and gradient gain (τ) using hyperparameters α and δ. During the later training stages, it allocates smaller gradient gains to low-quality anchor boxes, mitigating their adverse influence while prioritizing normal-quality anchor boxes to enhance the model's localization performance.

Experiments and analysis
Experimental setup
This paper experiment is based on the Ubuntu22.04 operating system, and uses i5-9400 F CPU and NVIDIA RTX 1080 Ti GPU for model training and testing. The HR-YOLO model is developed under the Pytorch framework. The deep learning framework used is Pytorch 12.1 and the programming language is Python 3.9. During the training process, to ensure fairness of the experiment, we set the size of the input image to 640 ×640. It should be noted that although this configuration can ensure the repeatability of the experiment, it may not be an optimal parameter in practical applications. In terms of optimizers, we chose YOLOv8's default Adam optimizer to ensure that it is universal compared with other methods. For details of the hyperparameter configuration, please refer to Table 1.

Evaluation metrics
In ADAS, the accuracy of the object detection algorithm is directly related to its security performance, and mAP is a key indicator for measuring the overall performance and reliability of the object detection algorithm[71]. Therefore, in order to comprehensively evaluate the validity of our proposed model, this study adopted mean accuracy (mAP) as the main evaluation indicator. To gain a deeper understanding of the concept of mAP, we introduced two basic indicators of recall and precision, which represent the proportion of correctly identified positive samples and the proportion of correctly identified correct instances among all identified positive instances. Their calculation formulas are shown in formulas (24)-(25):

$$Recall = \frac{TP}{TP+FN} \tag{24}$$

$$Precision = \frac{TP}{TP+FP} \tag{25}$$

Among them, TP (True Positives) is the number of bounding boxes correctly detected, FP (False Positives) is the number of bounding boxes incorrectly detected, and FN (False Negatives) is the number of bounding boxes missing during the detection process of the model. Furthermore, the average accuracy (AP) represents the tradeoff between the accuracy rate and the accuracy rate, which is quantified as the area under the accuracy rate-seeking rate curve. See formula for AP calculation (26):

$$AP = \frac{1}{m}\sum{}_{i}^{m} P_{i} = \int P\left(R\right)dR \tag{26}$$

The mAP is the average of AP values across all classes, and its calculation is expressed in Eq. (27):

Parameter name	Parameter value
Initial Learning rate	0.01
Batchsize	16
Momentum	0.937
Training Epoch	300
Decay strategy	Cosine annealing decay
Final learning rate	0.0001
Data augmentation	Mosaic data augmentation
Closing Mosaic epochs	10
Weight decay	0.0005

Table 1. Experimental parameters.

$$mAP = \frac{1}{N} \sum_{j}^{N} AP_j \tag{27}$$

Here, N denotes the total number of categories in the object recognition task. In this study, mAP calculations specifically refer to mAP50. To further evaluate the model, Parameters (the number of model parameters) and Giga Floating-Point Operations per Second (GFLOPs) are employed to measure the model size and computational efficiency. Additionally, Frames Per Second (FPS) is utilized to quantify the inference speed.

Dataset configuration

To assess the generalization ability of the proposed method, experiments were conducted on both real-world and synthetic foggy datasets. These experiments aim to evaluate the model's performance and effectiveness across different environmental conditions.

Real-World Task-Driven testing set

In 2019, Li et al. introduced the RESIDE dataset, comprising the real-world foggy scene object detection dataset RTTS and synthetic datasets. RESIDE is notable for being the only real-world foggy scene dataset with multi-class detection labels. As shown in Fig. 11, in this study, the RTTS dataset, containing 4322 grayscale images spanning diverse visible light conditions such as daytime, cloudy, and foggy weather, was employed to validate the proposed method. Although RTTS lacks ground truth annotations, it is widely used for evaluating dehazing methods. Its application in this study ensures a fair evaluation and establishes a robust benchmark for comparison.

Semantic understanding image dataset for urban street scenes

Most existing foggy image datasets are primarily derived from the domain of image enhancement, yet they often lack detailed annotations and authentic real-world scene data. To overcome these limitations, this study utilized the Foggy Cityscapes dataset[72], which builds upon the Cityscapes dataset and is generated through a combination of depth information and synthetic techniques. As shown in Fig. 12, the dataset includes multiple versions with varying fog densities, distinguished by a constant attenuation coefficient that defines the corresponding visibility range. This dataset is specifically designed to evaluate semantic understanding capabilities in urban street scenes. In this study, we utilized the highest fog density version (constant attenuation coefficient of 0.02) to thoroughly evaluate the dehazing model's performance in complex urban environments.

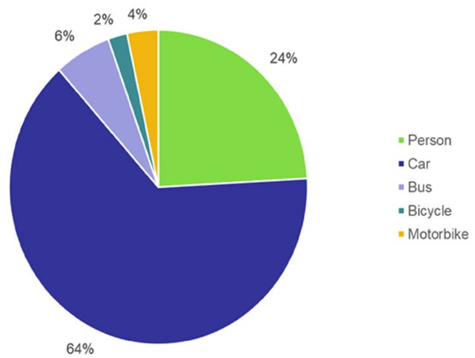
Experimental analysis

Comparative experiments

To assess the performance of the proposed enhanced YOLOv8 model, we performed comparative experiments against several state-of-the-art object detection models, including YOLOv11n, YOLOv10n, and Faster R-CNN. All experiments were conducted on the previously mentioned foggy datasets under identical experimental conditions, with the results summarized in Tables 2 and 3.

Experimental results on the RTTS dataset reveal that the proposed method attained a peak accuracy of 79.8% in object detection tasks under foggy weather conditions. Compared to other YOLO methods, such as YOLOv10[73], the proposed model demonstrated an improvement in recognition accuracy ranging from 4.7 to 8.9% points. Compared with other YOLO methods such as YOLOv10[67], the recognition accuracy of this model is improved by 4.7 to 8.9% points. This is because EPHD-Net's feature extraction ability of dense fog images is significantly better than that of other network architectures. At the same time, DND-Net restores foggy images in the preprocessing stage, and the combination of other modules achieves the optimal matching in HR-YOLO. The proposed model outperformed several benchmark methods, with performance improvements of 12.9% over faster R-CNN, 12.0% over SSD300[74], 14.4% over retinanet[75], 10.0% over CenterNet[76], 12.3% over DETR[77], 11.8% over efficientdet[78], 12.5% over Rtmdet[79],0.7% over MSFFA-YOLO, and 3.2% over R-YOLO. The proposed model

(a)The RTTS dataset used for this study



(b) The various labels contained in the Foggy Cityspaces dataset are displayed in the form of a pie chart with corresponding percentages

Fig. 11. Analysis of the RTTS dataset.

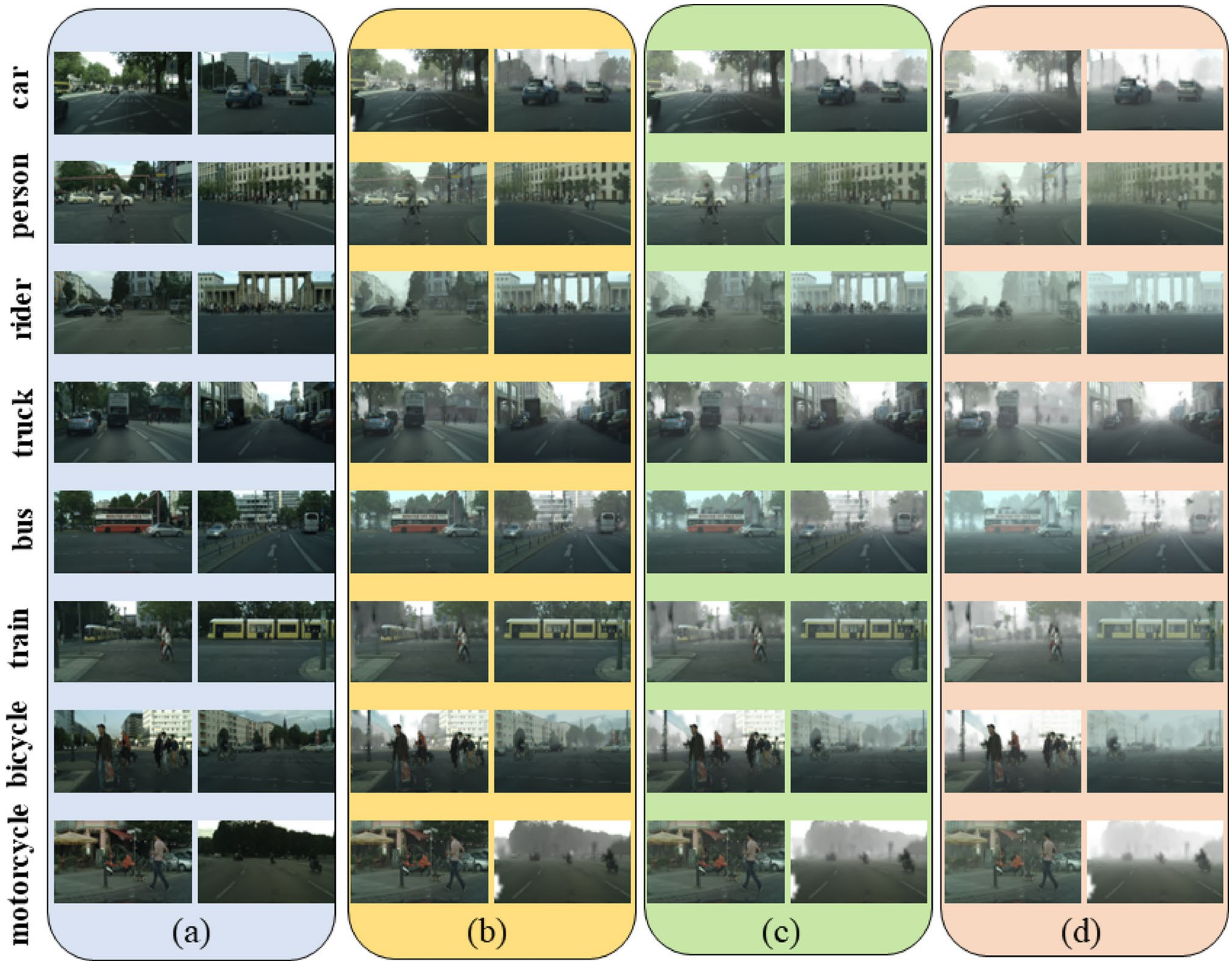


Fig. 12. Visualization of artificial fog effects in the Foggy Cityscapes dataset: (**a**) original image, (**b**) synthesized foggy image with a constant attenuation coefficient of 0.005, (**c**) synthesized foggy image with a constant attenuation coefficient of 0.01, and (**d**) synthesized foggy image with a constant attenuation coefficient of 0.02.

outperformed several benchmark methods, with performance improvements of 12.9% over faster R-CNN, 12.0% over SSD300, 14.4% over retinanet, 10.0% over CenterNet, 12.3% over DETR, 11.8% over efficientdet, and 12.5% over Rtmdet. Although not the fastest, the FPS of the detection module in the mainstream on-board ADAS system is usually between 10–60 frames, and HR-YOLO Far exceeds the industry standard, which proves that it can fully Meet the needs of haze weather target detection tasks in actual applications.

Furthermore, the HR-YOLO model excels in detecting small objects (e.g., pedestrians and vehicles) under foggy weather conditions, a capability critical for object recognition in complex environments. As presented in Table 3, experimental results on the Foggy Cityscapes dataset show that the proposed HR-YOLO model achieved the highest mAP50 value of 49.5%. This represents a substantial improvement of 8.5 to 15.2% points over other YOLO series methods. Specifically, the mAP increased by 16.4% compared to Faster R-CNN, 14.9% compared to SSD300, 18.0% compared to RetinaNet, 25.3% compared to CenterNet, 22.1% compared to DETR, 23.1% compared to EfficientDet, 15.7% compared to Rtmdet, 6.8% compared to MSFFA-YOLO, and 1.2% compared to R-YOLO.

We present inference map comparisons of these models in Fig. 13, and these results emphasize the robustness and accuracy of HR-YOLO in handling foggy scenes, especially for identifying small objects in urban environments, which is often a challenge to existing methods.

Ablation study

To rigorously assess the contributions of each proposed module, we conducted comprehensive ablation experiments on the RTTS dataset.

In these experiments, the YOLOv8 model was incrementally enhanced with different module configurations, including YOLOv8-A (integrating EPHD-Net), YOLOv8-B (incorporating both EPHD-Net and DND-Net), along with YOLOv8-C, YOLOv8-D, YOLOv8-E, and the complete HR-YOLO model. These configurations

Methods	Precision	Recall	mAP50	Person	Car	Bus	Bicycle	Motorbike	FPS
Faster RCNN	0.701	0.623	0.669	0.753	0.697	0.642	0.634	0.619	7.83
SSD300	0.704	0.619	0.678	0.744	0.738	0.646	0.625	0.637	20.84
RetinaNet	0.693	0.631	0.654	0.729	0.704	0.635	0.593	0.608	10.67
CenterNet	0.711	0.632	0.698	0.768	0.754	0.652	0.664	0.653	39.72
DETR	0.706	0.635	0.675	0.637	0.778	0.625	0.693	0.642	24
EfficentDet	0.714	0.618	0.684	0.769	0.756	0.643	0.616	0.636	36
Rtmdet	0.717	0.632	0.673	0.654	0.769	0.643	0.616	0.636	131
YOLOv5-L	0.786	0.633	0.717	0.805	0.813	0.634	0.625	0.708	67
YOLOv5-X	0.773	0.639	0.726	0.801	0.902	0.628	0.627	0.672	81
YOLOX-Tiny	0.793	0.640	0.709	0.803	0.806	0.629	0.613	0.694	74
YOLOv7-s	0.802	0.627	0.718	0.817	0.826	0.643	0.627	0.677	93
YOLOv8-n	0.828	0.637	0.737	0.825	0.87	0.667	0.638	0.695	264
YOLOv8-s	0.836	0.643	0.751	0.841	0.883	0.671	0.656	0.704	196
YOLOv10-n	0.801	0.624	0.717	0.81	0.858	0.616	0.623	0.68	283
YOLOv11-n	0.79	0.637	0.728	0.82	0.87	0.659	0.641	0.649	**316**
MSFFA-YOLO	0.825	0.671	0.777	0.853	0.746	**0.824**	0.693	**0.769**	199
R-YOLO	0.801	0.656	0.722	0.795	0.738	0.642	0.677	0.758	240
HR-YOLO	**0.832**	**0.698**	**0.798**	**0.872**	**0.913**	0.720	**0.740**	0.745	223

Table 2. Performance comparison on the RTTS dataset.

Method	mAP50	Person	Rider	Car	Truck	Bus	Train	Motorbike	Bicycle	FPS
Faster RCNN	0.331	0.272	0.336	0.431	0.163	0.258	0.091	0.117	0.269	9.4
SSD300	0.346	0.334	0.379	0.425	0.284	0.366	0.378	0.255	0.314	19
RetinaNet	0.315	0.291	0.397	0.429	0.208	0.374	0.241	0.265	0.299	11
CenterNet	0.242	0.275	0.366	0.377	0.131	0.286	0.027	0.178	0.294	23
DETR	0.276	0.248	0.307	0.415	0.236	0.338	0.197	0.213	0.265	26
EfficentDet	0.264	0.279	0.362	0.352	0.16	0.283	0.102	0.246	0.325	29
Rtmdet	0.338	0.336	0.379	0.485	0.265	0.387	0.236	0.28	0.336	27
YOLOv5-L	0.343	0.299	0.433	0.435	0.235	0.36	0.328	0.301	0.352	45
YOLOv5-X	0.41	0.432	0.478	0.586	0.238	0.457	0.392	0.315	0.382	30
YOLOX-Tiny	0.401	0.399	0.473	0.513	0.279	0.411	0.352	0.363	0.418	45
YOLOv7-s	0.379	**0.471**	0.497	0.539	0.252	0.389	0.159	0.311	0.414	47
YOLOv8-n	0.398	0.332	0.475	0.479	0.316	0.474	0.409	0.323	0.371	110
YOLOv8-s	0.442	0.44	0.439	0.603	0.316	0.504	**0.515**	0.317	0.406	81
YOLOv10-n	0.399	0.389	0.499	0.543	0.259	0.483	0.337	0.286	0.396	91
YOLOv11-n	0.408	0.358	0.408	0.595	0.332	0.493	0.427	0.274	0.377	**128**
MSFFA-YOLO	0.468	0.448	0.433	0.642	0.398	0.551	0.501	0.383	0.388	108
R-YOLO	0.489	0.473	0.495	**0.666**	0.391	0.558	**0.522**	0.409	0.444	98
HR-YOLO	**0.495**	0.45	**0.516**	0.609	**0.424**	**0.571**	0.515	**0.417**	**0.456**	108

Table 3. Comparative experimental results on the foggy cityscapes dataset.

allowed us to systematically evaluate the impact of the DySample module, DWConv module, C2 module, and WIoU loss function on object detection performance. The results are detailed in Table 4.

On the RTTS dataset, the baseline YOLOv8 model achieved a mAP of 73.9%. By introducing EPHD-Net, YOLOv8-A demonstrated a 2.8% increase in mAP, validating EPHD-Net's role in enhancing the backbone network's feature extraction capabilities. When DND-Net was added alongside EPHD-Net, YOLOv8-B further improved by 4.2%, reaching an mAP of 78.1%.

These results highlight the complementary strengths of EPHD-Net and DND-Net: the former significantly enhances image feature extraction, while the latter boosts defogging performance. Together, these modules effectively address the challenges of object detection under foggy weather conditions, enabling substantial accuracy improvements.

The results for YOLOv8-C, YOLOv8-D, YOLOv8-E, and HR-YOLO highlight that, with the integration of EPHD-Net and DND-Net, the incorporation and combination of all enhanced modules in the neck network yielded an additional 4.6–5.9% improvement in recognition accuracy. These findings further substantiate the individual and collective contributions of each module to the overall enhancement of object detection

Fig. 13. Comparative inference performance of HR-YOLO and other algorithms on the RTTS and Foggy Cityscapes datasets, highlighting the model's superior accuracy and robustness under challenging weather conditions.

Figure 13. (continued)

Model	EPHD-Net	DND-Net	DySample	DWConv	C2	Wiou	mAP50	Parameters	GFLOPs	FPS
YOLOv8	×	×	×	×	×	×	0.739	**3.0M**	**10.2**	**264**
YOLOv8-A	√	×	×	×	×	×	0.767	5.1 M	16.2	182
YOLOv8-B	√	√	×	×	×	×	0.781	6.0 M	16.4	168
YOLOv8-C	√	√	√	×	×	×	0.782	5.7 M	15.5	180
YOLOv8-D	√	√	√	√	×	×	0.785	4.8 M	13.3	192
YOLOv8-E	√	√	√	√	√	×	0.788	4.6 M	13.1	210
HR-YOLO	√	√	√	√	√	√	**0.798**	4.6 M	13.0	223

Table 4. Results of the ablation study on the RTTS dataset.

performance. Specifically, the proposed EPHD-Net and DND-Net exhibited remarkable efficacy in improving object detection under foggy weather conditions, driven by their precise feature extraction capabilities and efficient defogging processes. In particular, when using DySample, we found that it has not improved much in mAP, but DySample's unique processing mechanism has significantly reduced the GFLOPs and parameter volume, and at the same time, the FPS has also significantly improved, which is more obvious in the case of thick fog.

In summary, the ablation study underscores the pivotal role of each proposed module in augmenting object detection performance. The synergistic combination of all methods delivered the most substantial improvement, thereby affirming the robustness and effectiveness of the proposed enhancements.

Conclusion

This paper presents HR-YOLO, an improved YOLOv8-based object detection method specifically optimized for traffic scenarios in foggy weather. The proposed model introduces several enhancements: (1) EPHD-Net as the backbone network to strengthen image feature extraction; (2) a lightweight unsupervised dehazing network, DND-Net, evolved from AOD-Net, to enhance defogging performance; (3) the integration of DySample, DWConv, and C2 modules into the neck network, boosting detection efficiency and the utilization of multi-scale features; and (4) the adoption of the WIoU loss function to improve target localization accuracy.

Experiments conducted on the RTTS and Foggy Cityscapes datasets reveal that HR-YOLO achieves mAP improvements of 5.9% and 9.7%, respectively, compared to the baseline, demonstrating its high efficacy and precision under adverse meteorological conditions, all while maintaining low computational overhead. Notably, these results highlight the model's ability to balance accuracy and efficiency, making it suitable for deployment in real-world applications.

Future work will focus on further refining the network architecture to enhance detection speed without compromising accuracy. Such advancements aim to extend the applicability of HR-YOLO to edge computing devices and mobile platforms, enabling broader deployment in practical scenarios and promoting its integration into resource-constrained environments.

Data availability

The datasets used in this study are publicly available. RTTS dataset is available on the official website: https://sites.google.com/view/reside-dehaze-datasets/reside-%CE%B2/(accessed on 10 July 2024). Foggy cityspaces dataset is available on the official website: https://www.cityscapes-dataset.com/(accessed on 28 August 2024).

Received: 22 January 2025; Accepted: 10 April 2025
Published online: 16 April 2025

References

1. Birch, R. S., Bolwell, A. & Maloney, J. J. *Development of a high performance, memory based, relational database system using Ada*, in *The Third International IEEE Conference on Ada Applications and Environments* 39, 40, 41, 42, 43, 44 – 39, 40, 41, 42, 43, 44 (1988).
2. Ma, C. & Xue, F. A review of vehicle detection methods based on computer vision. *J. Intell. Connected Veh.* **7**, 1–18 (2024).
3. Wang, S., Hu, X., Sun, J. & Liu, J. Hyperspectral anomaly detection using ensemble and robust collaborative representation. *Inf. Sci.* **624**, 748–760 (2023).
4. Huang, Q. et al. A novel image-to-knowledge inference approach for automatically diagnosing tumors. *Expert Syst. Appl.* **229**, 120450 (2023).
5. Michaelis, C. et al. Benchmarking robustness in object detection: Autonomous driving when winter is coming. *arXiv preprint arXiv:07484* (2019). (1907).
6. Xiao, J., Long, H., Li, R. & Li, F. *Research on Methods of Improving Robustness of Deep Learning Algorithms in Autonomous Driving*, in *IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)* (IEEE, 2022). pp. 644–647. (2022).
7. Chandrakar, R. et al. Enhanced the moving object detection and object tracking for traffic surveillance using RBF-FDLNN and CBF algorithm. *Expert Syst. Appl.* **191**, 116306 (2022).
8. Kejriwal, R., Ritika, H. & Arora, A. *Vehicle detection and counting using deep learning basedYOLO and deep SORT algorithm for urban traffic management system*, in *First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)* (IEEE, 2022). pp. 1–6. (2022).
9. Ding, L. *SlimEMA-YOLOv8: Enhanced Traffic Sign Recognition for Autonomous Driving Using EMA and Slim-Neck in YOLOv8*, in *6th International Conference on Internet of Things, Automation and Artificial Intelligence (IoTAAI)* (IEEE, 2024). pp. 723–726. (2024).
10. Ćorović, A., Ilić, V., Đurić, S., Marijan, M. & Pavković, B. *The real-time detection of traffic participants using YOLO algorithm*, in *26th Telecommunications Forum (TELFOR)* (IEEE, 2018). pp. 1–4. (2018).
11. Kumar, D. & Muhammad, N. Object detection in adverse weather for autonomous driving through data merging and YOLOv8. *Sensors* **23**, 8471 (2023).
12. Pandharipande, A. et al. Sensing and machine learning for automotive perception: A review. *IEEE Sens. J.* **23**, 11097–11115 (2023).
13. Chen, G. et al. *YOLOv8-MixFaster: A Lightweight Traffic Sign Recognition Algorithm*, in. *4th International Conference on Computer, Big Data and Artificial Intelligence (ICCBD + AI)* (IEEE, 2023). pp. 748–753. (2023).
14. Xi, Y. et al. FiFoNet: fine-grained target focusing network for object detection in UAV images. *Remote Sens.* **14**, 3919 (2022).
15. Dong, J. & Pan, J. *Physics-based feature dehazing networks*, in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28*, Proceedings, Part XXX 16; (Springer, 2020). pp. 188–204. (2020).
16. He, K., Sun, J. & Tang, X. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**, 2341–2353 (2010).
17. Liu, X., Ma, Y., Shi, Z. & Chen, J. *Griddehazenet: Attention-based multi-scale network for image dehazing*, in *Proceedings of the IEEE/CVF international conference on computer vision* 7314–7323 (2019).
18. Dong, H. et al. *Multi-scale boosted dehazing network with dense feature fusion*, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* pp. 2157–2167. (2020).
19. Peng, C., Li, X. & Wang, Y. Td-yoloa: an efficient Yolo network with attention mechanism for tire defect detection. *IEEE Trans. Instrum. Meas.* **72,** 1-11. (2023).
20. Cheng, S., Zhu, Y. & Wu, S. Deep learning based efficient ship detection from drone-captured images for maritime surveillance. *Ocean Eng.* **285**, 115440 (2023).
21. He, Y. & Li, J. TSRes-YOLO: an accurate and fast cascaded detector for waste collection and transportation supervision. *Eng. Appl. Artif. Intell.* **126**, 106997 (2023).

22. Guo, C. et al. *Zero-reference deep curve estimation for low-light image enhancement*, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* 1780–1789 (2020).
23. Gomaa, A. & Saad, O. M. Residual Channel-attention (RCA) network for remote sensing image scene classification. *Multimedia Tools Appl.* 1–25 (2025).
24. Gomaa, A. *Advanced domain adaptation technique for object detection leveraging semi-automated dataset construction and enhanced yolov8*, in *6th Novel Intelligent and Leading Emerging Sciences Conference (NILES)* (IEEE, 2024). pp. 211–214. (2024).
25. Gomaa, A., Abdelwahab, M. M. & Abo-Zahhad, M. Efficient vehicle detection and tracking strategy in aerial videos by employing morphological operations and feature points motion analysis. *Multimedia Tools Appl.* **79**, 26023–26043 (2020).
26. Salem, M., Gomaa, A. & Tsurusaki, N. *Detection of earthquake-induced building damages using remote sensing data and deep learning: A case study of mashiki town, Japan*, in *IGARSS 2023–2023 IEEE International Geoscience and Remote Sensing Symposium*IEEE, pp. 2350–2353. (2023).
27. Gomaa, A. & Abdalrazik, A. Novel deep learning domain adaptation approach for object detection using semi-self Building dataset and modified yolov4. *World Electr. Veh. J.* **15**, 255 (2024).
28. Zhao, H., Jin, J., Liu, Y., Guo, Y. & Shen, Y. FSDF: A high-performance fire detection framework. *Expert Syst. Appl.* **238**, 121665 (2024).
29. Zhang, X., Wang, T., Luo, W. & Huang, P. Multi-level fusion and attention-guided CNN for image dehazing. *IEEE Trans. Circuits Syst. Video Technol.* **31**, 4162–4173 (2020).
30. Balla, P. K., Kumar, A. & Pandey, R. A 4-channelled hazy image input generation and deep learning-based single image dehazing. *J. Vis. Commun. Image Represent.* **100**, 104099 (2024).
31. Xiao, J. et al. Single image dehazing based on learning of haze layers. *Neurocomputing* **389**, 108–122 (2020).
32. Zhang, B. & Zhao, J. Hardware implementation for real-time haze removal. *IEEE Trans. Very Large Scale Integr. VLSI Syst.* **25**, 1188–1192 (2016).
33. Girshick, R., Donahue, J., Darrell, T. & Malik, J. Rich *feature hierarchies for accurate object detection and semantic segmentation*, in *Proceedings of the IEEE conference on computer vision and pattern recognition* pp. 580–587. (2014).
34. Girshick, R. Fast r-cnn. *arXiv preprint arXiv:1504.08083* (2015).
35. Ren, S., He, K., Girshick, R., Sun, J. & Faster, R-C-N-N. Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1137–1149 (2016).
36. Redmon, J. *You only look once: Unified, real-time object detection*, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016).
37. Redmon, J. & Farhadi, A. *YOLO9000: better, faster, stronger*, in *Proceedings of the IEEE conference on computer vision and pattern recognition* pp. 7263–7271. (2017).
38. Redmon, J. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* (2018).
39. Bochkovskiy, A., Wang, C. Y. & Liao, H. Y. M. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:10934* (2020). (2004).
40. Zhu, X., Lyu, S., Wang, X. & Zhao, Q. *TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios*, in *Proceedings of the IEEE/CVF international conference on computer vision* pp. 2778–2788. (2021).
41. Cai, Y. et al. Pedestrian motion trajectory prediction in intelligent driving from Far shot first-person perspective video. *IEEE Trans. Intell. Transp. Syst.* **23**, 5298–5313 (2021).
42. Wang, L., Qin, H., Zhou, X., Lu, X. & Zhang, F. R-YOLO: A robust object detector in adverse weather. *IEEE Trans. Instrum. Meas.* **72**, 1–11 (2022).
43. Zhang, Q., Hu, X. M. S. F. F. A. Y. O. L. O. & Network Multi-Class object detection for traffic investigations in foggy weather. *IEEE Trans. Instrum. Meas.* **72,** 1-12 (2023).
44. Li, G., Ji, Z., Qu, X., Zhou, R. & Cao, D. Cross-domain object detection for autonomous driving: A Stepwise domain adaptative YOLO approach. *IEEE Trans. Intell. Veh.* **7**, 603–615 (2022).
45. Touvron, H. et al. PMLR,. *Training data-efficient image transformers & distillation through attention*, in *International conference on machine learning* pp. 10347–10357. (2021).
46. He, Y. et al. AEGLR-Net: attention enhanced global–local refined network for accurate detection of car body surface defects. *Robot. Comput. Integr. Manuf.* **90**, 102806 (2024).
47. Liu, Y., Shao, Z. & Hoffmann, N. Global attention mechanism: Retain information to enhance channel-spatial interactions. *arXiv preprint arXiv:2112.05561* (2021).
48. Liu, Z. et al. *Swin transformer: Hierarchical vision transformer using shifted windows*, in *Proceedings of the IEEE/CVF international conference on computer vision* pp. 10012–10022. (2021).
49. Wang, Z. et al. *Uformer: A general u-shaped transformer for image restoration*, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* pp. 17683–17693. (2022).
50. Fu, X. et al. *Removing rain from single images via a deep detail network*, in *Proceedings of the IEEE conference on computer vision and pattern recognition* pp. 3855–3863. (2017).
51. Zamir, S. W. et al. *Multi-stage progressive image restoration*, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* pp. 14821–14831. (2021).
52. Cun, X., Pun, C. M. & Shi, C. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34 10680–10687 (2020).
53. Wang, J., Li, X. & Yang, J. *Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal*, in *Proceedings of the IEEE conference on computer vision and pattern recognition* pp. 1788–1797. (2018).
54. Le, H. & Samaras, D. *Shadow removal via shadow image decomposition*, in *Proceedings of the IEEE/CVF International Conference on Computer Vision* pp. 8578–8587. (2019).
55. Hughes, M. J. & Hayes, D. J. Automated detection of cloud and cloud shadow in single-date Landsat imagery using neural networks and Spatial post-processing. *Remote Sens.* **6**, 4907–4926 (2014).
56. Wang, J., Lu, K., Xue, J., He, N. & Shao, L. Single image dehazing based on the physical model and MSRCR algorithm. *IEEE Trans. Circuits Syst. Video Technol.* **28**, 2190–2199 (2017).
57. Tarel, J. P. & Hautiere, N. Fast *visibility restoration from a single color or gray level image*, in *2009 IEEE 12th international conference on computer vision*IEEE, pp. 2201–2208. (2009).
58. Rasti, B., Scheunders, P., Ghamisi, P., Licciardi, G. & Chanussot, J. Noise reduction in hyperspectral imagery: overview and application. *Remote Sens.* **10**, 482 (2018).
59. Li, B., Peng, X., Wang, Z., Xu, J. & Feng, D. *Aod-net: All-in-one dehazing network*, in *Proceedings of the IEEE international conference on computer vision* pp. 4770–4778. (2017).
60. Ju, M. et al. Image dehazing and exposure using an enhanced atmospheric scattering model. *IEEE Trans. Image Process. 30*. **IDE**, 2180–2192 (2021).
61. Li, B., Wu, F., Weinberger, K. Q. & Belongie, S. Positional normalization. *Adv. Neural. Inf. Process. Syst.* **32** (2019).
62. Chollet, F. *Xception: Deep learning with depthwise separable convolutions*, in *Proceedings of the IEEE conference on computer vision and pattern recognition* pp. 1251–1258. (2017).
63. Meng, G., Wang, Y., Duan, J., Xiang, S. & Pan, C. *Efficient image dehazing with boundary constraint and contextual regularization*, in *Proceedings of the IEEE international conference on computer vision* pp. 617–624. (2013).

64. Zhu, Q., Mai, J. & Shao, L. A fast single image haze removal algorithm using color Attenuation prior. *IEEE Trans. Image Process.* **24**, 3522–3533 (2015).
65. Ren, W. et al. *Single image dehazing via multi-scale convolutional neural networks*, in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14*, Proceedings, Part II 14 (Springer, 2016). pp. 154–169. (2016).
66. Cai, B., Xu, X., Jia, K., Qing, C. & Tao, D. Dehazenet: an end-to-end system for single image haze removal. *IEEE Trans. Image Process.* **25**, 5187–5198 (2016).
67. Li, B. et al. Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process.* **28**, 492–505 (2018).
68. Liu, X., Ma, Y., Shi, Z. & Chen, J. *Griddehazenet: Attention-based multi-scale network for image dehazing*, in *Proceedings of the IEEE/CVF international conference on computer vision* pp. 7314–7323. (2019).
69. Zheng, Z. et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. *IEEE Trans. Cybernetics*. **52**, 8574–8586 (2021).
70. Tong, Z., Chen, Y., Xu, Z. & Yu, R. Wise-IoU: bounding box regression loss with dynamic focusing mechanism. *arXiv preprint arXiv:2301.10051* (2023).
71. Zaidi, S. S. A. et al. A survey of modern deep learning based object detection models. *Digit. Signal Proc.* **126**, 103514 (2022).
72. Sakaridis, C., Dai, D. & Van Gool, L. Semantic foggy scene Understanding with synthetic data. *Int. J. Comput. Vision*. **126**, 973–992 (2018).
73. Wang, A. et al. Yolov10: Real-time end-to-end object detection. *arXiv preprint arXiv:2405.14458* (2024).
74. Irie, K. & Nishikawa, K. *Detection Method from 4K Images Using SSD300 without Retraining*, in *2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*IEEE, pp. 877–883. (2022).
75. Wang, Y., Wang, C., Zhang, H., Dong, Y. & Wei, S. Automatic ship detection based on retinanet using multi-resolution Gaofen-3 imagery. *Remote Sens.* **11**, 531 (2019).
76. Duan, K. et al. *Centernet: Keypoint triplets for object detection*, in *Proceedings of the IEEE/CVF international conference on computer vision* pp. 6569–6578. (2019).
77. Zhu, X. et al. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159* (2020).
78. Tan, M., Pang, R. & Le, Q. V. *Efficientdet: Scalable and efficient object detection*, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* pp. 10781–10790. (2020).
79. Lyu, C. et al. Rtmdet: An empirical study of designing real-time object detectors. *arXiv preprint arXiv:2212.07784* (2022).

Acknowledgements
We appreciate the critical and constructive comments and suggestions from the reviewers that helped improve the quality of this manuscript. We also would like to offer our sincere thanks to those who participated in the data processing and provided constructive comments for this study.

Author contributions
Y.Z. contributed to the study conception, software and visualization. Validation and methodology were performed by Y.Z. and N.J. The first draft of the manuscript was written by Y.Z. All authors reviewed the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to N.J.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.