scientific reports

Check for updates

OPEN

DSNet enables feature fusion and detail restoration for accurate object detection in foggy conditions

Zhiyong Jing[1,3], Zhaobing Chen[2], Yucheng Shi[1], Lei Shi[2,4]✉, Lin Wei[2]✉ & Yufei Gao[2,4]

In real-world scenarios, adverse weather conditions can significantly degrade the performance of deep learning-based object detection models. Specifically, fog reduces visibility, complicating feature extraction and leading to detail loss, which impairs object localization and classification. Traditional approaches often apply image dehazing techniques before detection to enhance degraded images; however, these processed images often retain a rough appearance with a loss of detail. To address these challenges, we propose a novel network, DehazeSRNet(DSNet), which is designed to optimize feature transmission and restore lost image details. First, DSNet utilizes the dehaze fusion network (DFN) to learn dehazing features, applying differentiated processing weights to regions with light and dense fog. Second, to enhance feature transmission, DSNet introduces the MistClear Attention (MCA) module, which is based on a re-parameterized channel-shuffle attention mechanism and effectively optimizes feature information transfer and fusion. Finally, to restore image details, we design the hybrid pixel activation transformer (HPAT), which combines channel attention and window-based self-attention mechanisms to activate additional pixel regions. Experimental results on the Foggy Cityscapes, RTTS, DAWN, and rRain datasets demonstrate that DSNet significantly outperforms existing methods in accuracy and achieves exceptional real-time performance, reaching 78.1 frames per second (FPS), highlighting its potential for practical applications in dynamic environments. As a robust detection framework, DSNet offers theoretical insights and practical references for future research on object detection under adverse weather conditions.

Keywords Object detection, Adverse weather, Image processing, Feature fusion

In recent years, object detection methods based on deep learning[1–5] have demonstrated notable performance in various traffic scenarios. Currently, mainstream object detection algorithms are primarily benchmarked against standard datasets, such as MSCOCO[6], PASCAL-VOC[7], and Imagenet[8]. In real-world environments, images captured by cameras are frequently influenced by unavoidable environmental factors, such as fog, snow, and rain. These conditions can blur object contours in the images, significantly affecting the performance of detection systems.

The low visibility, image blur, and increased noise in foggy conditions present substantial challenges for object detection. Current solutions can be divided into three categories. The most common strategy is to preprocess the input image using established dehazing algorithms (such as AOD-Net[9], FFA-Net[10]) before inputting it into the detection network. However, images processed in this way exhibit limited generalization capability in practical applications because the restored images remain unclear and may lose important details. An alternative approach is to directly train the detection model on degraded images. This strategy often relies on the feature extraction capability of the detection model. Several studies[11–13], have optimized the entire model with a joint loss for restoration and detection. Moreover, several methods[14,15] have enhanced detection performance by incorporating domain adaptation techniques, enabling models trained under normal weather conditions to transfer effectively to adverse conditions, such as rain and fog. However, this approach faces challenges in feature extraction, as well

[1]School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001, Henan, China. [2]School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450003, Henan, China. [3]College of Software Engineering, Zhengzhou University of Light Industry, Zhengzhou 450001, Henan, China. [4]SongShan Laboratory, Zhengzhou 450001, Henan, China. ✉email: 2006090@email.zzuli.edu.cn; weilin@zzu.edu.cn

as in the transmission and fusion of feature information. These difficulties result in increased computational complexity and reduced inference speed, making it less suitable for resource-constrained environments.

To address this challenge, we introduce the DehazeSRNet (DSNet) model, comprising three core modules. First, the Dehaze Fusion Network (DFN) module learns dehazing features by combining channel attention and pixel attention mechanisms. By assigning differentiated processing weights to thin and dense haze regions, the DFN module significantly enhances dehazing performance. Next, the MistClear Attention (MCA) module improves feature transmission and fusion efficiency. Leveraging channel shuffling and structurally re-parameterized convolutions, the MCA module effectively optimizes feature information transmission and fusion accuracy. As depicted in Fig. 4, the structurally re-parameterized convolution employs a multi-branch architecture during training, which is transformed into a single-branch structure during inference, streamlining the inference process. This reduction in computational complexity enhances inference speed, which is crucial for real-time object detection in foggy conditions. Finally, the HPAT module combines channel attention and window self-attention mechanisms to activate more pixels and recover lost image details. The window self-attention mechanism applies adaptive weighting to local regions, focusing on critical areas to facilitate detail recovery and enhance object detection accuracy. As shown in Fig. 1, more details are retained after being processed by DSNet.

Our contributions can be summarized as follows:

- A novel DehazeSRNet (DSNet) model is introduced, integrating three innovative modules: the DFN, which employs channel and pixel attention mechanisms to optimize feature extraction by addressing uneven haze distribution; the MCA, which enhances feature fusion efficiency using channel-shuffling-based structural re-parameterized convolutions; and the HPAT, which combines channel attention and window self-attention mechanisms to strengthen cross-window feature interactions and recover image details.
- DSNet demonstrates exceptional performance on the Foggy Cityscapes dataset, achieving a highest detection precision of 37.8% mAP-a significant improvement over existing methods-while maintaining an inference speed of 78.1 FPS. Moreover, DSNet exhibits strong robustness and generalization capabilities on the RTTS dataset as well as on other adverse weather datasets, including DAWN and rRain.
- The model's modular design and optimized computational efficiency highlight its practical application potential, making DSNet particularly suitable for resource-constrained real-time scenarios, such as autonomous driving and intelligent surveillance under challenging adverse weather conditions.

The rest of this paper is organized as follows. In Sect. 2, we review the related work on dehazing methods for object detection and foggy day object detection. In Sect. 3, we present our proposed method in detail. In Sect. 4, we provide experimental results and analysis. Finally, in Sect. 5, we summarize our work.

Related work

Image dehazing

Image dehazing is an important research direction in the field of computer vision, aiming to enhance visibility by removing haze effects from images. In recent years, many researchers have proposed various methods to address this issue, which can be primarily categorized into traditional methods and deep learning-based approaches.

Prior-based dehazing methods aim to utilize well-validated priors during the dehazing process. Ju et al.[16] and Wang et al.[17], drawing upon the principles of hazy imaging , typically employ an atmospheric scattering model to simulate the image generation process. This model can be expressed as:

$$I(x) = J(x)t(x) + A(1 - t(x)), \tag{1}$$

where $I(x)$ is the hazy image captured by the camera; $J(x)$ is the clear image; A is the global atmospheric light; and $t(x)$ represents the medium transmission. Therefore, the image dehazing process can be expressed as:

Fig. 1. **a** shows the image processed by traditional dehazing methods, **b** presents the image processed by our proposed DSNet method. To provide a more intuitive comparison of the two methods, specific areas are highlighted with red boxes and enlarged in the images.

$$J(x) = \frac{(I(x) - At(x))}{t(x)} + A, \tag{2}$$

According to Eq. (2), traditional methods can utilize prior knowledge for image dehazing. He et al.[18] proposed the Dark Channel Prior, which relies on the observation that certain regions in natural outdoor images demonstrate extremely low pixel values in specific channels. Zhu et al.[19] introduced color attenuation prior to effectively recovering depth information to estimate transmission. Although these methods have demonstrated effectiveness in image dehazing, their performance is often limited because manually designed priors may not generalize well across diverse hazy images.

Learning-based methods aim to leverage the powerful feature extraction capabilities of Convolutional Neural Networks (CNNs) in combination with large-scale paired data for image dehazing[20]. Ren et al.[21] conducted pioneering research utilizing CNNs to tackle the haze removal challenge. Qin et al.[10] proposed an end-to-end Feature Fusion Attention Network (FFA-Net) to directly recover haze-free images, in which the Feature Attention (FA) module integrates channel attention and pixel attention mechanisms. Li et al.[9] introduced the AOD-Net method for image dehazing utilizing CNNs. Chen et al.[22] proposed a Detail Enhancement Attention Block (DEAB) comprising Detail Enhancement Convolution (DEConv) and Content-guided Attention (CGA) to enhance feature learning, thereby improving dehazing performance. Although these methods can improve image quality to some extent, they may inadequately emphasize detailed information, impacting dehazing performance.

Object detection in foggy weather

Object detection in hazy weather faces numerous challenges due to the decline in image quality. Haze can reduce visibility, blur object boundaries, and exacerbate scattering effects, thereby decreasing detection accuracy. In recent years, researchers have proposed various solutions through image enhancement techniques, domain adaptation, as well as models specifically tailored for object detection under hazy conditions. Liu et al.[23] employed a Differentiable Image Processing (DIP) module to enhance hazy images prior to detection. Subsequently, Kalwar et al.[24] designed a Gated Differentiable Image Processing (GDIP) module, achieving progressive image enhancement through a multi-stage guidance scheme. Zhang et al.[11] introduced a CPA-Enhancer chain-of-thought prompting mechanism, which adapts to unknown degradation conditions by incorporating a chain-of-thought prompting mechanism for image enhancement. Wang et al.[14] proposed the R-YOLO framework, consisting of an Image Quasi-Translation Network (QTNet) and a Feature Calibration Network (FCNet) to progressively adapt from clear weather domains to adverse weather conditions. To address domain adaptation for vehicle detection in heavy fog, Hu et al.[15] introduced an algorithm called DAGL-Faster, which handles domain differences from three perspectives: local image level, global image level, and instance level. Additionally, it incorporates consistency regularization to facilitate simultaneous alignment at both image and instance levels, optimizing overall alignment effects. Zhang et al.[25] introduced the MSFFA-YOLO network, which combines YOLOv7 with a multi-scale feature fusion attention mechanism to enhance object localization and classification accuracy in hazy conditions, while simultaneously improving image visibility via a recovery subnet. Zhong et al.[26] proposed DR-YOLO, integrating the atmospheric scattering model and co-occurrence relationship graph into an end-to-end detection framework. This approach enhances dehazing feature extraction and object detection performance through a recovery subnet and relationship reasoning module, while an adaptive feature fusion module further improves detection effectiveness. Wang et al.[27] proposed RDMNet, a restoration-enhanced object detection network for adverse weather scenarios. It uses a dual-branch structure with a restoration branch and degradation modeling to capture multi-scale degradation representations, improving adaptability to various weather conditions. A multi-scale bidirectional feature fusion module and restoration-weight decay strategy enable collaborative optimization of detection and restoration tasks.

Although the aforementioned methods have made significant contributions to the field, they still tend to lose image details after processing. When the visual features extracted by the model are insufficiently influenced by haze, it can negatively impact the performance of the detector.

Methodology

The method proposed in this paper is DehazeSRNet (DSNet). In target detection under foggy conditions, challenges primarily stem from reduced visibility, which complicates feature extraction and results in the loss of image details. Therefore, our proposed DSNet consists of three components, as shown in Fig. 2: the Dehaze Fusion Network (DFN) facilitates the detector's learning of dehazing features, followed by the MistClear Attention (MCA) module, which improves both the efficiency and accuracy of feature extraction. Lastly, the Hybrid Pixel Activation Transformer (HPAT) module further refines the image to restore lost details. The remainder of this section will provide a detailed overview of DFN, MCA, and HPAT.

Dehaze fusion network module

Hazy environments significantly impair image quality, with uneven distributions of thin and dense fog regions leading to varying degrees of detail loss across different areas. Traditional CNN-based image dehazing networks often treat channel and pixel features uniformly, disregarding the non-uniformity of haze distribution and lacking targeted optimization. To overcome this limitation, we propose the Dehaze Fusion Network (DFN) module, which incorporates Channel Attention and Pixel Attention mechanisms to enhance flexibility in processing diverse types of information. By focusing on pixels within dense fog regions and prioritizing critical channel information, the DFN module achieves targeted dehazing optimization. The Group Architecture of the

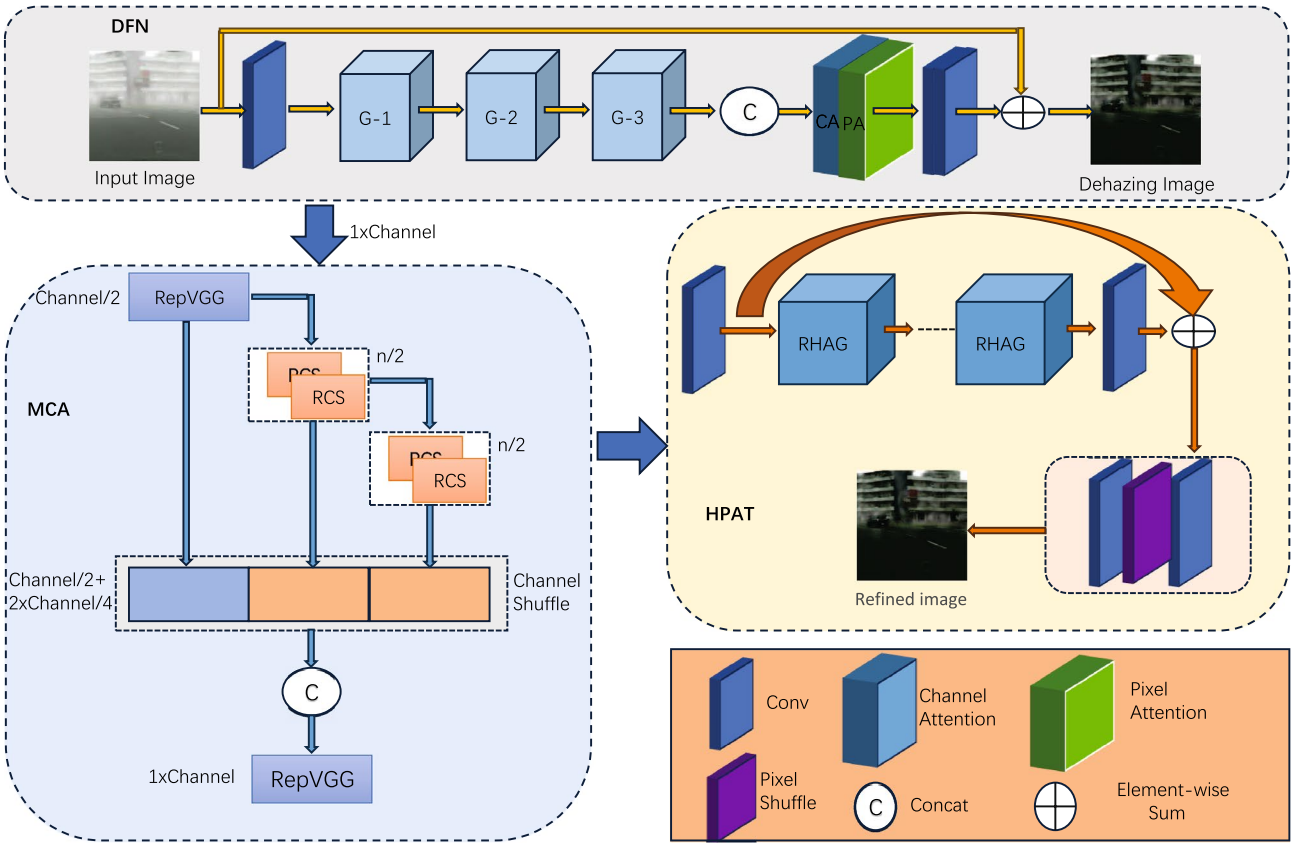


Fig. 2. The overall structure of DehazeSRNet (DSNet). First, the DFN module extracts dehazing features using the Group Architecture (G-n) to address uneven haze distribution. These features are then processed by the MCA module to enhance feature extraction and fusion. Finally, the HPAT module refines the image, progressively recovering lost details. Within the HPAT module, Residual Hybrid Attention Groups (RHAG) are employed to restore intricate image details, ensuring the model's effectiveness in adverse weather conditions.

DFN module, illustrated in Fig. 3, highlights its advantages in effectively addressing the challenges posed by non-uniform haze distributions.

The Channel Attention (CA) module primarily focuses on the importance variations of features across different channels. First, the global spatial information of each channel is converted into a channel descriptor through a global average pooling operation:

$$G_c = h_p(f_c) = \frac{1}{E \times F} \sum_{i=1}^{E} \sum_{j=1}^{F} X_c(i,j), \tag{3}$$

where $X_c(i,j)$ represents the value of the c-th channel X_c at position (i,j), and h_p denotes the global pooling function. After pooling, the feature map's shape changes from $C \times E \times F$ to $C \times 1 \times 1$, effectively capturing the global information of each channel. Next, the channel descriptor undergoes two convolution layers, followed by ReLU and sigmoid activation functions to generate attention weights for each channel:

$$BA_c = \sigma\left(\mathrm{Conv}\left(\delta\left(\mathrm{Conv}\left(G_c\right)\right)\right)\right), \tag{4}$$

where σ is the sigmoid activation function, and δ is the ReLU activation function. Finally, element-wise multiplication is performed between the input feature F_c and the channel attention weight BA_c to obtain the weighted channel feature W_c^*, which emphasizes the channel information in thick haze regions (Fig. 4):

$$W_c^* = BA_c \otimes W_c. \tag{5}$$

The Pixel Attention (PA) module addresses the uneven distribution of haze in the image, ensuring that the network assigns higher weights to regions with thick haze and areas containing high-frequency information. To achieve this, the channel-weighted output W^* is fed into the PA module. The PA module processes this input through two convolution layers, followed by ReLU and sigmoid activation functions to generate pixel-level attention weights PA, changing the shape from $C \times H \times W$ to $1 \times H \times W$:

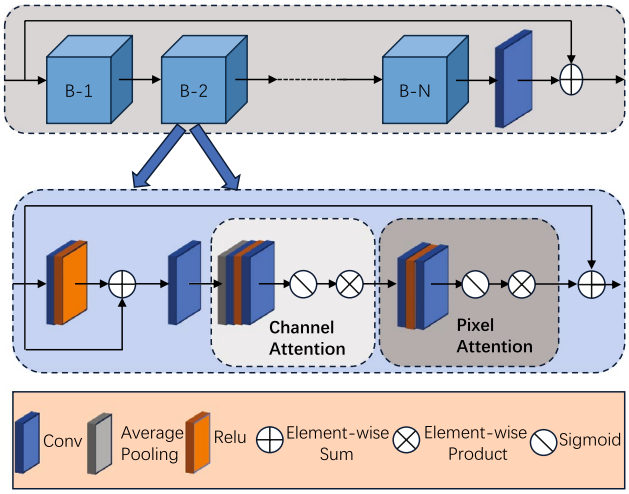


Fig. 3. The group architecture structure within the DFN module.

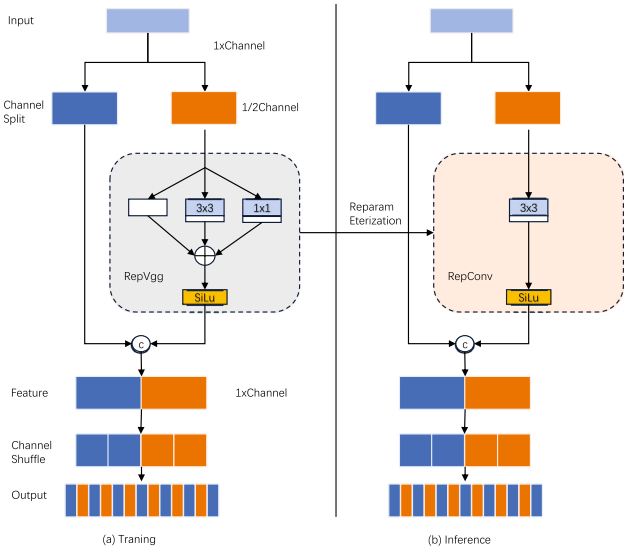


Fig. 4. The composition of the RCS module comprises: **a** the RepVGG structure utilized during the training phase and **b** the RepConv structure employed during model inference or deployment. In these structures, rectangles with black borders represent specific module operations performed on the tensors, while rectangles with gradient-filled shading indicate the properties of the tensors, with the width of the rectangle reflecting the number of channels in the tensor.

$$PA = \sigma(\mathrm{Conv}(\delta(\mathrm{Conv}(W^*)))), \tag{6}$$

where σ is the sigmoid activation function, and δ is the ReLU activation function. Finally, element-wise multiplication is performed between W^* and PA to obtain the final output of the Future Attention (FA) module, denoted as $\widetilde{W}$:

$$\widetilde{W} = W^* \otimes PA. \tag{7}$$

To visually demonstrate the effectiveness of the PA module, pixel-level feature weight maps are presented in Fig. 5. As shown in Fig. 6, even in environments with high-concentration haze, the DFN module effectively removes the haze, preserving key scene details' clarity and significantly enhancing the image's visibility. Experimental results indicate that the DFN demonstrates strong adaptability to varying haze densities, maintaining consistent dehazing performance and generating clear, information-rich images across different haze levels.

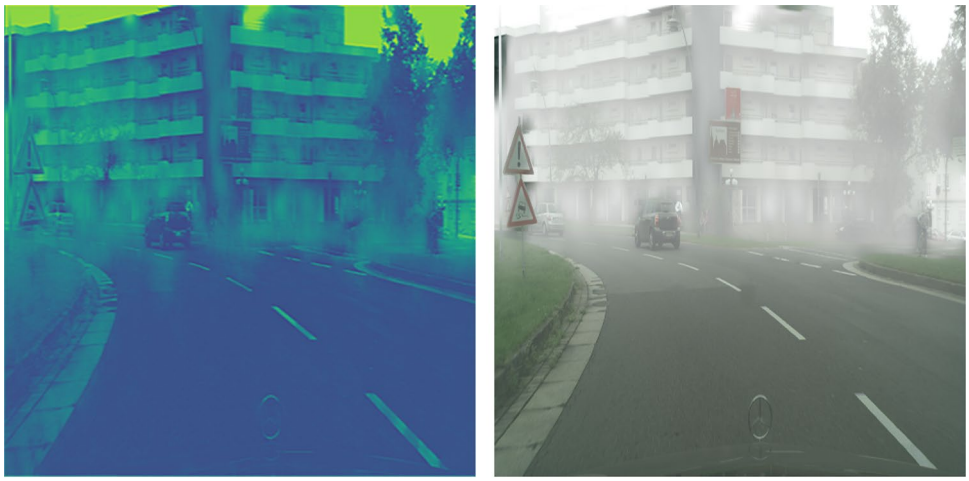


Fig. 5. PA attention map.

Fig. 6. After processing through the DFN, images with varying fog densities are displayed. The first row presents foggy images from the Foggy Cityscapes dataset, each with simulated attenuation coefficients of 0.005, 0.01, and 0.02, respectively. The second row displays the corresponding dehazed images.

MistClear attention module

Although the DFN module effectively extracts both global and local feature information during dehazing, feature extraction in hazy conditions remains challenging, particularly in terms of inference speed. To overcome this limitation, we propose the MistClear Attention (MCA) module, which accelerates the inference process while enhancing feature fusion capabilities. The specific structure of the MCA module is depicted in Fig. 2, with its core component, the RCS module, illustrated in Fig. 4. Inspired by ShuffleNet, the RCS module integrates the strengths of RepVGG[28] and RepConv. By utilizing channel shuffling-based structurally re-parameterized convolutions, the RCS module significantly increases the information density of feature extraction and substantially reduces inference time. Consequently, the MCA module plays a critical role in object detection under hazy conditions, enhancing both the computational efficiency and real-time performance of the network.

Under foggy conditions, image visibility is significantly reduced, and the loss of fine details complicates feature extraction, thereby posing greater challenges for object detection algorithms. Specifically, the low contrast and blurring induced by fog exacerbate inter-channel information redundancy within the image. Traditional convolutional neural networks (CNNs) often rely on local features from specific channels, overlooking potential information from other channels, which leads to inefficiency and an increased computational burden. To address this issue, the RCS module effectively mitigates inter-channel information redundancy through a channel-shuffling mechanism, optimizing both information flow and feature extraction efficiency. The channel-shuffling process disrupts the original channel order and rearranges it, allowing features from different channels to be combined more effectively. This operation fosters broader interaction between channels, reducing the

accumulation of redundant features and thus enhancing feature extraction efficiency. Particularly under foggy conditions, this mechanism facilitates the extraction of more discriminative features from blurry images, ultimately improving object recognition accuracy.

The RCS module integrates the advantages of RepVGG/RepConv with ShuffleNet, leveraging structural reparameterization and channel-shuffling operations. During the training phase, it captures rich feature information through a multi-branch structure, and during inference, it is transformed into a simplified single-branch structure, thereby reducing memory consumption and accelerating the inference process. This design is particularly well-suited for object detection tasks in low-visibility environments, such as foggy conditions, thereby enhancing the model's overall performance.

Additionally, the MCA module enhances feature reuse and information flow between different channels by stacking multiple RCS modules across various layers of the network, all while maintaining low memory consumption and computational complexity. The MCA module also incorporates the concept of path aggregation, aligning feature maps of different sizes through a combination of upsampling and downsampling operations, thereby facilitating information exchange across multiple prediction layers and ensuring both fast and high-accuracy inference. To further optimize computational efficiency, the MCA module employs a multi-scale feature fusion strategy. By reducing the number of detection heads and optimizing anchor generation, it alleviates the computational burden and significantly shortens the computation time of post-processing steps, such as Non-Maximum Suppression (NMS). Due to its superior computational efficiency and accuracy, the MCA module is particularly well-suited for resource-constrained real-time object detection tasks, such as autonomous driving and video surveillance systems, where high detection accuracy and speed are critical, even under foggy conditions.

Hybrid pixel activation transformer module

Although the DFN and MCA modules provide effective solutions for feature extraction and inference efficiency, dehazed images may still appear coarse, falling short of the requirements for downstream tasks such as object detection. To further enhance image detail quality, we propose the Hybrid Pixel Activation Transformer (HPAT) module, which refines and strengthens image features. The HPAT module integrates channel attention and self-attention mechanisms, incorporating the Integrated Overlap Attention Block (IOAB) to facilitate efficient interactions between adjacent window features. This design activates more pixels and enhances image details, thereby significantly improving object detection accuracy under hazy conditions. Collaborating seamlessly with the previously introduced DFN and MCA modules, the HPAT module forms part of a comprehensive framework, encompassing feature extraction, detail refinement, and recovery, ensuring robust support for efficient and accurate object detection in complex weather scenarios.

As illustrated in Fig. 7, the entire network comprises three components: shallow feature extraction, deep feature extraction, and image reconstruction. Each RHAG consists of multiple Dynamic Attention Integration Blocks (DAIB), one Integrated Overlap Attention Block (IOAB), and a 3×3 convolution layer with a residual connection. The reconstruction module utilizes a pixel shuffling method[29] to upsample the fused features.

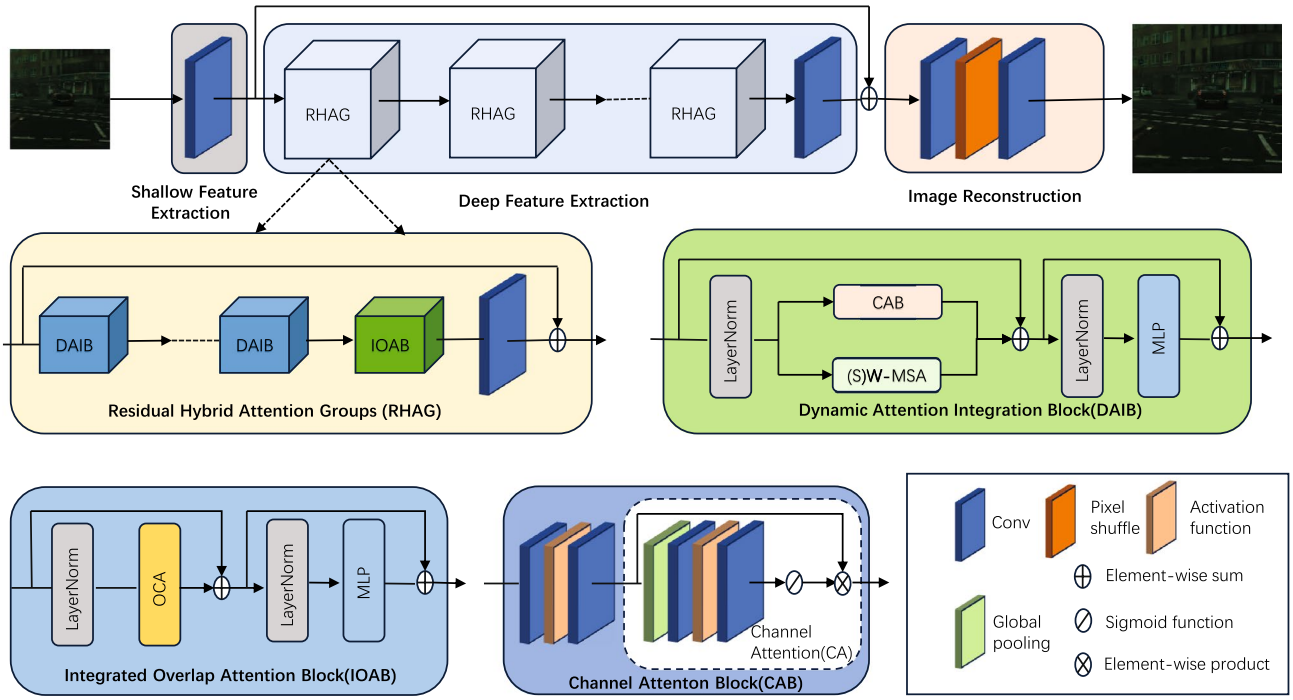


Fig. 7. The overall architecture of the HPAT, along with the structures of the RHAG, DAIB, and IOAB.

The DAIB enhances the network's representational power by integrating a convolution block based on channel attention into the standard Transformer architecture. Notably, shifted window-based self-attention (SW-MSA) is applied intermittently within consecutive DAIB layers, as demonstrated in[30,31]. To prevent conflicts between the Channel Attention Block (CAB) and Multi-Head Self-Attention (MSA) during optimization and visual representation, a small constant, α, is added to the CAB output. For a given input feature X, the DAIB computation process is as follows: Feature Preprocessing:

$$X_E = LE(X), \tag{8}$$

Feature Fusion:

$$X_F = (S)W\text{-}MSA(X_E) + \alpha CAB(X_E) + X, \tag{9}$$

Output Calculation:

$$Y = MLP(LE(X_E)) + X_F, \tag{10}$$

Here, X_E and X_F represent intermediate features, and Y is the output of the DAIB. Specifically, each pixel is treated as an embedded token, and MLP refers to a multi-layer perceptron. In the self-attention module, the input feature X has dimensions $H \times W \times C$ and is divided into $\frac{HW}{M^2}$ local windows of size $M \times M$. For the local window features $XW \in \mathbb{R}^{M^2 \times C}$, the query, key, and value matrices are derived through linear mappings to compute Q, K, and V. The formula for calculating window self-attention is:

$$\text{Attention}(Q, K, V) = \text{SoftMax}\left(\frac{QK^T}{\sqrt{d}} + B\right)V, \tag{11}$$

where d denotes the dimension of the query/key, and B represents the relative position encoding.

The CAB consists of two convolutional layers and a Channel Attention (CA) module. Transformers often require a large number of channels for token embedding, which can be computationally expensive. To address this, we reduce the number of channels in the convolutional layers using a constant β, which decreases the output channel count from C to $\frac{C}{\beta}$. We then expand it back to C channels in the second layer. Finally, the CA module is applied to recalibrate the channel features, improving object detection performance in foggy conditions.

The IOAB consists of an Overlapping Cross Attention (OCA) layer and a Multi-Layer Perceptron (MLP) layer, similar to the standard Swin Transformer block[31]. In the OCA, as shown in Fig. 8, we use different window sizes to partition the projected features. Specifically, for input features X, the queries, keys, and values X_Q, X_K, and $X_V \in \mathbb{R}^{H \times W \times C}$ are defined as follows: X_Q is divided into $\frac{HW}{M^2}$ non-overlapping windows of size $M \times M$, while X_K and X_V are split into $\frac{HW}{M^2}$ overlapping windows of size $M_o \times M_o$, where

$$M_o = (1 + \gamma) \times M, \tag{12}$$

and γ controls the overlap size. The overlapping partitioning can be viewed as a sliding window with kernel size M_o and stride M. To ensure consistent window sizes, zero padding of $\frac{\gamma M}{2}$ is applied. The attention matrix is calculated using the same procedure as in Eq. (11), with a relative position bias $B \in \mathbb{R}^{M \times M_o}$. Unlike Window Self-Attention (WSA), OCA computes keys and values over a broader range to capture more relevant information for the queries.

To comprehensively evaluate the dehazing performance of the proposed DSNet model in real-world environments, we conducted qualitative experiments on the RTTS dataset, which consists of real-world images captured under foggy conditions. As illustrated in Fig. 9, DSNet effectively removes haze from the input images, significantly enhancing visual clarity and visibility. The processed images reveal more useful information, restore important target regions previously obscured by fog, and preserve greater scene detail and structural

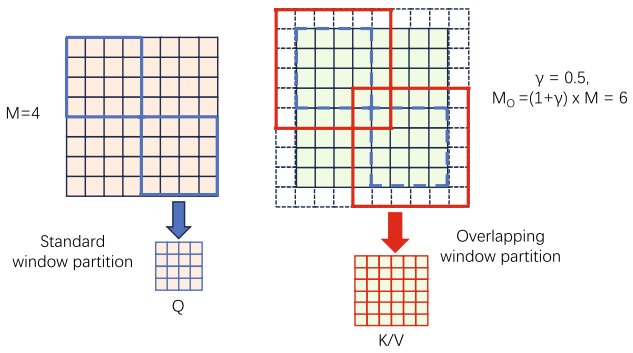


Fig. 8. The overlapping window partition for OCA.

Fig. 9. The images processed by the proposed DSNet model are presented, where the first row shows the original foggy images, and the second row displays the corresponding dehazed results.

content. These results demonstrate that DSNet maintains excellent dehazing performance in practical scenarios, showcasing its robustness and potential for real-world applications.

Experiment and analysis

This section begins with an introduction to the datasets and evaluation metrics utilized in our experiments. Subsequently, we provide a detailed account of the implementation of DSNet on these datasets. Specifically, we assess our method on both synthetic and real-world datasets to compare its performance with that of state-of-the-art (SOTA) methods. Additionally, an ablation study is conducted to further validate the effectiveness of our network.

Evaluation metrics and datasets

Evaluation metrics: To quantitatively evaluate the performance of the object detector, we utilize mean Average Precision (mAP), a widely recognized metric in object detection. mAP assesses the model's overall performance in multi-class detection tasks by averaging precision values across different recall levels, providing a comprehensive measure of both accuracy and stability. Another critical evaluation metric is Frames Per Second (FPS), which indicates the number of images the model can process per second. This metric is particularly important in resource-constrained scenarios, as it reflects the real-time processing capability of the model. Additionally, detection speed is evaluated through inference time per image, where shorter inference times signify faster detection speeds.

Datasets: Given the limited availability of publicly accessible datasets for detecting adverse weather conditions in real-world scenarios, we selected the Foggy Cityscapes, RTTS, DAWN, rRain, and KITTI datasets to comprehensively evaluate and compare the performance of our proposed DSNet with other detection methods under challenging weather conditions. The Foggy Cityscapes dataset offers significant advantages in foggy environments, as it accurately simulates the effects of fog on images using a physics-based model, providing high-quality synthetic fog images that facilitate effective evaluation of the model's performance in simulated foggy scenarios. The RTTS dataset further strengthens the practical relevance of the evaluation, as it includes foggy images captured from real-world traffic scenes, encompassing a variety of traffic-related objects such as motorcycles, bicycles, and pedestrians, thereby enabling the assessment of the model's performance in complex and dynamic real-world traffic environments. Meanwhile, the DAWN and rRain datasets serve as test sets to validate the robustness of our method under diverse weather conditions. KITTI is a dataset for clear weather conditions. A detailed description of the datasets used is provided in Table 1.

The Foggy Cityscapes dataset[38] is a synthetic dataset designed to simulate foggy environments with high realism. It integrates a physics-based optical model, accurately estimated depth maps, and precise atmospheric

Dataset	Number of Images
Foggy Cityscapes	3450
RTTS	4322
DAWN	1027
rRain	1900
KITTI	7481

Table 1. Number of images in each dataset.

light simulation to faithfully replicate the effects of fog on images. Each foggy image is generated by blending a clear image with depth maps sourced from the Cityscapes dataset using a rendering technique. Consequently, the annotations and data segmentation of the foggy images adhere to the standards of the original Cityscapes dataset. For our training process, we utilized the most challenging version of the foggy scenes, with a simulated attenuation coefficient set to $\beta = 0.02$, ensuring more representative foggy environmental conditions.

The Real-world Task-driven Testing Set (RTTS)[39] is a subset of the extensive RESIDE dataset, comprising both synthetic and real-world hazy images. RTTS includes 4,322 annotated foggy images captured under real-world conditions. The dataset encompasses five primary categories of traffic-related objects: motorcycles, bicycles, pedestrians, buses, and cars. The majority of these images originate from authentic traffic and driving scenarios, showcasing a diverse range of scenes.

The DAWN dataset[40] is a substantial image dataset focused on vehicle detection under adverse weather conditions, designed to provide researchers with a comprehensive and realistic platform for assessing and enhancing the performance of vehicle detection systems in challenging weather scenarios. This dataset comprises 1,027 images captured in real traffic environments, encompassing four types of severe weather conditions: fog, snow, rain, and dust storms. Each image is annotated with professional object bounding boxes that clearly indicate the location and size of vehicles, thereby enabling researchers to utilize this dataset for training and testing vehicle detection algorithms.

The rRain dataset[41] is a specially curated collection of real driving images captured under rainy conditions, comprising 1900 natural rain images taken at various locations and times. These images are annotated with five categories of traffic-related objects: pedestrians, bicycles, motorcycles, cars, and buses.

The KITTI dataset[42], captured under clear weather conditions, contains real-world image data from various scenes, including urban, rural, and highway environments. Each image includes up to 15 vehicles and 30 pedestrians, with varying degrees of occlusion and truncation. For our purposes, we have retained annotations for pedestrians, bicycles, cars, and trucks.

Implementation details

The training of DSNet was performed using input images with a resolution of 640×640 and a batch size of 16, over a total of 100 epochs. To enhance training stability and mitigate overfitting, Mosaic augmentation was disabled during the final 10 epochs. The initial learning rate was set to 1×10^{-2}, and a weight decay parameter of 5×10^{-4} was applied to prevent gradient explosion and promote efficient model convergence. We set the weighting factor in HPAT (α), the squeeze factor between two convolutions in CAB (β), and the overlapping ratio of OCA (γ) as 0.01, 3, and 0.5. During training, a Cosine Annealing Decay strategy was employed to dynamically adjust the learning rate, enabling finer adjustments with smaller learning rates in later stages, thereby improving the model's generalization capability. The experiments were implemented using the PyTorch framework and conducted on an NVIDIA GeForce RTX 3090 GPU with 24 GB of memory.

Performance of detectors on fog weather

Considering that DSNet prioritizes model accuracy and real-time performance, and that YOLOv8 excels in both areas, we selected YOLOv8 as the backbone of our baseline model. To ensure fairness, our approach was compared with state-of-the-art models within the YOLO series. In the "Separate" method, we evaluated AODNet[9], FFANet[10], and CPAEnhancer[11]; in the "Domain Adaptation" category, we compared DAGL-Faster[15], SWDA[33], LODS[34], and R-YOLO[14]; and for the "Union" method, we selected CF-YOLO[35], CDNet[36], DR-YOLO[26],RDMNet[27] and TogetherNer[37] as benchmarks for comparison.

Table 2 presents a performance comparison of our method against existing state-of-the-art techniques on the Foggy Cityspaces dataset. The table lists the names of the various methods, the types of training data utilized, and their mean Average Precision (mAP). The baseline method, YOLOv8, trained exclusively on foggy images, achieved a mAP of 32.4%. Subsequently, we report the performance of various separate and domain adaptation methods, with CPAEnhancer achieving a mAP of 36.2%. Within the domain adaptation category, DAGL-Faster demonstrated the best performance, achieving a mAP of 36.7%. In the union method, our proposed approach achieved a mAP of 37.8% using only foggy images for training, significantly outperforming other methods and demonstrating its effectiveness and superiority in foggy conditions.

To present performance results across various categories, Table 3 summarizes the quantitative outcomes of different object detection methods on the Foggy Cityspaces dataset. The table includes the mAP for each method across various target categories, including cars, motorcycles, buses, bicycles, pedestrians, cyclists, trains, and trucks. Our model demonstrates exceptional performance, achieving a precision of 0.451 in Bus detection and resulting in an overall mAP of 37.8%, significantly surpassing other methods. This outcome indicates that the proposed approach exhibits superior detection capabilities under complex foggy conditions, particularly

Method		Training for detection head	mAP(%)
Baseline	YOLOv8[32]	Foggy images only	32.4
Separate	AODNet[9]	Foggy images only	32.8
	FFANet[10]	Foggy images only	34.1
	CPAEnhancer[11]	Foggy images only	36.2
Domain adaptation	DAGL-faster[15]	Clean and foggy images	36.7
	SWDA[33]	Clean and foggy images	34.3
	LODS[34]	Clean and foggy images	35.8
	R-YOLO[14]	Clean and foggy images	34.9
Union	CF-YOLO[35]	Foggy images only	35.1
	CDNet[36]	Foggy images only	35.5
	DR-YOLO[26]	Foggy images only	34.5
	RDMNet[27]	Clean and foggy images	18.5
	TogetherNet[37]	Clean and foggy images	20.3
	Ours	Foggy images only	**37.8**

Table 2. Comparison of performance with state-of-the-art methods on the Foggy Cityspaces dataset, with the best results in bold font.

Method	Car	Mcycle	Bus	Bicycle	Person	Rider	Train	Truck	mAP(%)
YOLOv8	0.607	0.176	0.413	0.262	0.364	0.399	0.152	0.215	32.4
AODNet	0.578	0.173	0.442	0.261	0.327	0.378	0.269	0.196	32.8
FFANet	0.623	0.15	0.421	0.281	0.374	0.411	0.227	0.241	34.1
CPAEnhancer	0.621	0.198	0.423	0.324	0.38	0.424	0.262	0.26	36.2
DAGL-Faster	0.49	0.3	0.4	0.42	0.36	0.47	0.22	**0.28**	36.7
SWDA	0.44	0.3	0.36	0.35	0.3	0.42	0.33	0.25	34.3
LODS	0.48	**0.33**	0.39	**0.37**	0.34	**0.45**	0.19	0.27	35.8
R-YOLO	0.565	0.25	0.397	0.367	0.394	0.424	0.192	0.202	34.9
CF-YOLO	0.626	0.165	0.432	0.299	0.37	0.378	0.304	0.231	35.1
CDNet	0.626	0.195	0.424	0.338	0.377	0.399	0.306	0.177	35.5
DR-YOLO	**0.64**	0.151	0.411	0.31	0.39	0.414	0.235	0.211	34.5
RDMNet	0.51	0.05	0.25	0.16	0.23	0.2	0.01	0.08	18.5
TogetherNet	0.53	0.1	0.21	0.17	0.27	0.22	0.03	0.08	20.3
Ours	0.634	0.167	**0.451**	0.309	**0.4**	0.437	**0.365**	0.26	**37.8**

Table 3. Performance evaluation for each object class on Foggy Cityscapes. Bold indicates the best performance.

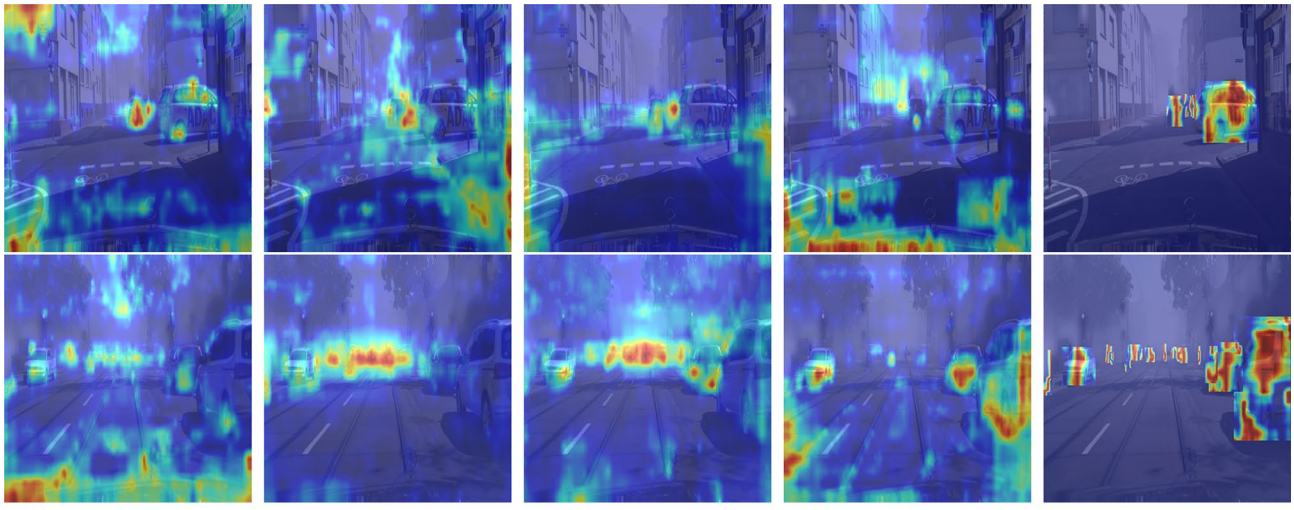
in critical target recognition, thereby further validating its effectiveness in practical applications. Examples of detection results from DSNet are presented in Fig. 10, illustrating the model's ability to accurately identify nearby objects even when obscured by fog. Additionally, we conducted heatmap visualizations, which are detailed in Fig. 11.

Although the proposed DSNet model demonstrates exceptional performance under most weather conditions, it still exhibits certain limitations in extremely dense fog conditions. Specifically, Fig. 12a illustrates a case where a traffic sign was mistakenly identified as a car. Despite the absence of occlusion, the model may misclassify the traffic sign due to the visual similarities between traffic signs and cars under dense fog conditions. Fig. 12b and c further demonstrate the model's performance degradation in heavy fog environments. In Fig. 12b , the motorcycle was not correctly detected, and in Fig. 12c , the pedestrian in the background was not detected. Dense fog significantly reduces image details, especially the visibility of distant objects, thereby impairing the model's detection capability. These failure cases highlight the impact of reduced image quality on model performance in low-visibility conditions. While the model performs well under most scenarios, further optimization is needed to enhance its robustness and accuracy in extreme weather conditions, such as heavy fog, to improve its performance in complex environments.

To validate our model's capability in real foggy scenes, we conducted experiments on the RTTS dataset, with results presented in Table 4. The table presents the precision for five object categories: buses, cars, bicycles, motorcycles, and pedestrians. Notably, our method exhibited exceptional performance across all categories, achieving a precision of 0.676 for the bicycle category and an overall mAP of 73.4%, thereby outperforming all comparison methods. This result indicates that the proposed approach provides enhanced accuracy and robustness for object detection in real foggy environments, thus making it suitable for practical applications.

(a) YOLOv8 **(b)** AODNet **(c)** FFANet **(d)** CPAEhancer **(e)** Ours

Fig. 10. We visualize the detection results of our DSNet on the Foggy Cityspaces dataset and compare them with some other methods.



(a) YOLOv8 **(b)** AODNet **(c)** FFANet **(d)** CPAEhancer **(e)** Ours

Fig. 11. Visualization of feature maps for DSNet based on the Foggy Cityspaces dataset. The feature maps exhibit stronger color distributions and finer localization, indicating enhanced target detection accuracy.

Performance of detectors in other adverse weather conditions

To validate the robustness of our method, we conducted experiments using the DAWN and rRain datasets. Table 5 presents the performance comparison derived from the DAWN dataset, encompassing average precision (mAP) results across various weather conditions, including fog, rain, sand, and snow. Our model demonstrates strong stability across diverse weather scenarios, achieving an overall mAP of 53.5% and surpassing existing methods such as YOLOv8 (50.6%) and AODNet (50.9%). Notably, our approach exhibits exceptional precision in sandy conditions, achieving a precision of 52.5%. These results highlight the adaptability and effectiveness of the proposed method in varied environments. Figure 13 illustrates the detection results of our DSNet on the DAWN dataset, along with visualizations of other comparative methods. Each row corresponds to a specific weather condition (fog, rain, sand, or snow), while each column represents a detection method. It is evident that our approach consistently delivers more accurate detection results across various adverse weather conditions.

Additionally, Table 6 presents a performance evaluation for each object category within the rRain dataset. Our model performs effectively across all categories, notably achieving a precision of 0.305 in the bicycle category, which contributes to an overall mAP of 30.7%, surpassing several comparative methods. These results further validate the superiority of our approach in complex and variable environments, underscoring its effectiveness in practical application scenarios.

(a) (b) (c)

Fig. 12. Our method exhibits cases of false positives and false negatives.

Method	Bus	Car	Bicycle	Mcycle	Person	mAP(%)
YOLOv8	0.604	0.837	0.616	0.676	0.780	70.3
AODNet	0.616	0.828	0.579	0.683	0.775	69.6
FFANet	0.621	0.840	0.595	0.679	0.783	70.3
CPAEnhancer	0.602	0.843	0.652	0.687	0.786	71.4
CDNet	0.614	0.833	0.626	0.678	0.780	70.6
CF-YOLO	0.619	0.849	0.631	0.697	0.787	71.6
DR-YOLO	0.614	**0.856**	0.622	**0.721**	**0.808**	72.4
RDMNet	0.46	0.75	0.28	0.47	0.69	53
TogetherNet	0.47	0.75	0.26	0.48	0.69	52.9
Ours	**0.643**	0.855	**0.676**	0.702	0.792	**73.4**

Table 4. Performance evaluation for each object class on RTTS. Bold indicates the best performance.

Method	Fog	Rain	Sand	Snow	All
YOLOv8	59.2	46.0	49.9	54.1	50.6
AODNet	59.3	47.7	48.6	57.1	50.9
FFANet	61.5	49.9	49.6	54.1	51.7
CPAEnhancer	61.3	50.1	48.8	53.9	51.3
CDNet	**63.6**	49.0	46.9	70.9	51.5
CF-YOLO	62.9	48.1	47.2	**74.3**	51.8
DR-YOLO	62.2	**54.4**	48.9	72.6	52.1
RDMNet	50	44	32.6	50	36.1
TogetherNet	45.6	45.9	27	51.9	31.9
Ours	63.3	51.3	**52.5**	72.1	**53.5**

Table 5. Performance comparison based on the DAWN dataset. All indicates the combined map value. Bold indicates the best performance.

DSNet aims to enhance object detection performance, demonstrating exceptional results not only in foggy conditions but also in other adverse weather scenarios, such as rain. This success can be attributed to its modular architecture, which integrates the MCA, DFN, and HPAT modules, thereby enhancing the model's robustness and adaptability. The DFN module effectively extracts information from diverse environmental backgrounds through multi-scale feature fusion. Meanwhile, the HPAT module employs channel attention and window self-attention mechanisms to activate more pixel regions, enhancing feature interaction and improving the recognition of critical features obscured by adverse conditions. Furthermore, the MCA module optimizes feature transfer and fusion, significantly enhancing feature extraction efficiency. By training under various weather conditions, DSNet learns a broader range of features and patterns, thereby strengthening its adaptability in high-noise or low-contrast environments. The model's design also facilitates seamless integration with existing

Fig. 13. The detection results of our DSNet on the DAWN dataset, along with visualizations of some comparative methods, are presented below.

Method	Bus	Car	Bicycle	Motorcycle	Person	mAP(%)
YOLOv8	0.198	0.650	0.228	0.121	0.233	28.6
AODNet	0.213	0.625	0.252	0.107	0.223	28.4
FFANet	**0.246**	0.647	0.234	0.108	0.225	29.2
CPAEnhancer	0.216	0.640	0.276	0.073	0.230	28.7
CDNet	0.190	0.655	0.214	**0.162**	0.218	28.8
CF-YOLO	0.187	**0.684**	0.249	0.097	0.230	28.9
DR-YOLO	0.229	0.670	0.196	0.108	**0.239**	28.8
RDMNet	0.151	0.546	0.227	0.086	0.159	23.4
TogetherNet	0.125	0.540	0.212	0.047	0.123	20.9
Ours	0.239	0.663	**0.305**	0.107	0.222	**30.7**

Table 6. Performance evaluation for each object class on rRain. Bold indicates the best performance.

detection algorithms, ensuring real-time performance, even under resource constraints. Collectively, these factors contribute to the enhanced object detection capabilities of DSNet across diverse weather conditions.

Performance of detectors in clear weather

To evaluate the performance of DSNet under clear weather conditions, we conducted experiments on the KITTI dataset, with the results presented in Table 7. It can be observed that most compared methods experience a

Method	Person	Bicycle	Car	Truck	mAP(%)
YOLOv8	0.854	**0.943**	0.98	**0.991**	94.2
AODNet	0.853	0.94	0.979	0.987	94
FFANet	0.851	0.924	0.981	0.99	93.7
CPAEnhancer	0.86	0.938	0.978	0.988	94.1
CDNet	0.843	0.942	0.978	0.989	93.8
CF-YOLO	0.88	0.925	0.977	0.989	94.3
DR-YOLO	0.819	0.904	0.966	0.987	91.9
RDMNet	0.71	0.83	0.96	0.98	87.1
TogetherNet	0.68	0.85	0.95	0.968	86.2
Ours	**0.865**	0.938	**0.983**	**0.991**	**94.4**

Table 7. Performance evaluation for each object class on KITTI. Bold indicates the best performance.

Module	V0	V1	V2	V3	V4	V5
Baseline	✓	✓	✓	✓	✓	✓
DFN		✓		✓	✓	✓
MCA			✓		✓	✓
HPAT			✓	✓		✓
mAP(%)	32.4	34.1	34.9	35.9	36.4	**37.8**

Table 8. Ablation study results on the Foggy Cityspaces dataset.

Module	mAP (%)
PA	36.0
CA	36.7
PA+CA	**37.8**

Table 9. Details ablation study of the DFN module. Bold indicates the best performance.

performance drop under clear conditions compared to the baseline. In contrast, only CF-YOLO and our proposed method show improvements. Notably, DSNet achieves the highest mAP across multiple categories, reaching an overall mAP of 94.4%, outperforming all other models. These results demonstrate that DSNet not only maintains robust performance in adverse weather conditions but also exhibits excellent detection capability under clear weather scenarios.

Ablation study

In this section, we conducted ablation experiments on the Foggy Cityspaces dataset to thoroughly evaluate the contributions of each module in DSNet. The experimental results are summarized in Table 8, illustrating the mAP performance of the model under various module combinations. The baseline model achieved a mAP of 32.4%. Following the introduction of the DFN module, the mAP increased to 34.1%, demonstrating its effectiveness in learning dehazing features. The further addition of the MCA module resulted in an mAP of 34.9%, underscoring its critical role in optimizing feature transfer. Upon integrating the HPAT module, the model's performance significantly improved, as HPAT further refines the image. In version V4, the mAP rose to 36.4%, while the final version V5, which consolidates all modules, achieved a mAP of 37.8%, demonstrating the significant effects of module synergy. These results indicate that each module positively influences object detection performance at various levels, particularly under adverse weather conditions, where their combination substantially enhances the model's detection capability. The ablation study not only validates the effectiveness of our model design but also offers important theoretical support for future research.

To evaluate the contribution of CA and PA to the model's performance, we conducted an ablation study on the DFN module, with the results presented in Table 9. When both CA and PA are utilized simultaneously, the model achieves the highest mAP. This outcome demonstrates that combining these two attention mechanisms within the DFN module effectively learns and optimizes dehazing features, thereby significantly enhancing the model's performance.

To further validate the advantages of the MCA module in feature transmission, fusion, and inference speed, as shown in Table 10, the inclusion of the MCA module leads to a significant increase in FPS, from 44.6 to 78.1. This improvement demonstrates that the MCA module not only optimizes computational efficiency but also accelerates the inference process, thereby enhancing model performance. Furthermore, detection accuracy is also improved, indicating that the MCA module plays a crucial role in enhancing feature fusion and information

Module	Inference Time (s)	FPS	mAP (%)
Without MCA	0.0224	44.6	35.9
With MCA	0.0128	**78.1**	**37.8**

Table 10. Ablation study on the impact of mca module in Foggy Cityscapes dataset. Bold indicates the best performance.

Module	mAP(%)	Inference time (s)	FPS
Channel interleave	36.3	13.3	75.1
Channel random	36.6	13.5	74.1
Channel reverse	37	15.3	65.4
Channel shuffle	**37.8**	12.8	**78.1**

Table 11. Performance comparison of different channel operations. Bold indicates the best performance.

γ	0	0.25	0.5	0.75
mAP (%)	36.6	35.9	**37.8**	37.0

Table 12. Ablation study on the overlapping ratio of IOAB. Bold indicates the best performance.

flow, enabling the model to achieve faster processing while maintaining or even boosting detection performance. These results provide compelling evidence of the importance of the MCA module for real-time object detection tasks, particularly in foggy environments.

To validate the effectiveness of the Channel Shuffle operation, we designed several alternative channel rearrangement strategies for comparative experiments. Specifically, Channel Reverse rearranges channels by completely reversing their order; Channel Random generates a fully random channel arrangement for each input image; and Channel Interleave adopts a fixed-interval alternating strategy to interleave channels. In contrast, Channel Shuffle employs a grouped shuffling approach that reorganizes and interweaves channels, enabling more efficient feature recombination and fusion. As shown in Table 11, Channel Shuffle achieves the highest mAP and the FPS among all methods. These results demonstrate that Channel Shuffle effectively facilitates inter-channel information interaction, significantly reduces redundant feature accumulation, and enhances feature extraction efficiency, particularly under hazy conditions.

In the IOAB module, we introduced a constant γ to regulate the degree of overlap in the cross-attention mechanism. To examine the influence of different overlap ratios on image detail recovery, we evaluated model performance using γ values ranging from 0 to 0.75, as presented in Table 12. Notably, when $\gamma = 0$, the module reduces to a standard Transformer block. Experimental results reveal that the model achieves optimal performance at $\gamma = 0.5$. However, when γ is set to 0.25 or 0.75, the model's performance either stagnates or declines. This observation suggests that an inappropriate overlap ratio may weaken feature interactions between adjacent windows, adversely impacting the recovery of image details.

Efficiency analysis

Table 13 compares the real-time performance of DSNet with that of other methods. DSNet achieved a frame rate of approximately 78.1 frames per second (FPS) on an RTX 3090 GPU. Although DSNet did not attain the highest frame rate, it maintained commendable real-time performance while achieving the best object detection mAP. This indicates that DSNet not only retains robust real-time capabilities in dynamic environments but also prioritizes detection accuracy, making it an ideal choice for applications that require a combination of efficiency and precision. These results underscore the effectiveness of DSNet in delivering reliable performance in practical applications, particularly in scenarios where timely decision-making is crucial.

Conclusion

In this study, we propose DSNet, an innovative model specifically designed for object detection under adverse weather conditions. DSNet efficiently extracts dehazing features through the DFN module, optimizes feature transmission and fusion with the MCA module, and restores image details using the HPAT module, thereby significantly improving object detection accuracy in foggy environments. Experimental results demonstrate that DSNet outperforms existing methods across multiple datasets, including Foggy Cityscapes, RTTS, DAWN, rRain, and KITTI. particularly under low-visibility foggy conditions, where its detection accuracy surpasses that of current approaches. This paper offers a detailed introduction to the design of DSNet and validates its superior performance through comparative experiments with SOTA methods, ablation studies, and visualization analyses. The design of DSNet demonstrates significant potential for practical applications, particularly in autonomous driving and intelligent video surveillance. In autonomous driving scenarios, DSNet restores image details through the HPAT module, reducing false and missed detections caused by impaired visibility, thereby

Method	Speed (s)	FPS
YOLOv8	0.0078	**128.2**
AODNet	0.0087	114.9
FFANet	0.0109	91.7
CPAEhancer	0.0283	35.3
DAGL-faster	**1.923**	0.52
CF-YOLO	0.0257	38.9
CDNet	0.0287	34.8
DR-YOLO	0.025	40
RDMNet	0.0403	24.8
TogetherNet	0.0314	31.9
Ours	0.0128	78.1

Table 13. Efficiency analysis. Bold indicates the best performance.

significantly improving pedestrian detection accuracy and safety. For instance, in dense fog, DSNet ensures that the vehicle system can accurately detect pedestrians and other obstacles at a distance, providing more reliable support for autonomous driving. Furthermore, DSNet's application in video surveillance systems demonstrates robust real-time performance. With the optimization provided by the MCA module, DSNet delivers clear video images even in adverse weather, ensuring timely responses to potential threats. Future work will focus on further optimizing the DSNet model to enhance its adaptability to other complex weather conditions, such as rain and snow, and exploring its potential across a broader range of real-time applications. This will not only advance technology in autonomous driving and intelligent surveillance but also open new avenues for object detection research under adverse weather conditions.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 15 January 2025; Accepted: 23 May 2025
Published online: 01 July 2025

References

1. Minaee, S. et al. Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**(7), 3523–3542 (2021).
2. Ye, M. et al. Deep learning for person re-identification: A survey and outlook. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**(6), 2872–2893 (2021).
3. Redmon, J. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016).
4. Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(6), 1137–1149 (2016).
5. Wang, C.-Y., Bochkovskiy, A. & Liao, H.-Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7464–7475 (2023).
6. Lin, T.-Y. et al. 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13. Springer, 740–755 (2014).
7. Everingham, M., Van Gool, L., Williams, C. K., Winn, J. & Zisserman, A. The pascal visual object classes (VOC) challenge. *Int. J. Comput. Vision* **88**, 303–338 (2010).
8. Deng, J. et al. IEEE conference on computer vision and pattern recognition. 248–255 (IEEE, 2009).
9. 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*, 4770–4778 (2017).
10. Qin, X., Wang, Z., Bai, Y., Xie, X. & Jia, H. FFA-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 11908–11915 (2020).
11. Zhang, Y., Wu, Y., Liu, Y. & Peng, X. CPA-enhancer: Chain-of-thought prompted adaptive enhancer for object detection under unknown degradations, arXiv preprint arXiv:2403.11220, (2024).
12. Liu, D. et al. Connecting image denoising and high-level vision tasks via deep learning. *IEEE Trans. Image Process.* **29**, 3695–3706 (2020).
13. Huang, S.-C., Le, T.-H. & Jaw, D.-W. DSNet: Joint semantic learning for object detection in inclement weather conditions. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**(8), 2623–2633 (2020).
14. 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).
15. Hu, M., Wu, Y., Yang, Y., Fan, J. & Jing, B. DAGL-faster: Domain adaptive faster R-CNN for vehicle object detection in rainy and foggy weather conditions. *Displays* **79**, 102484 (2023).
16. Ju, M., Zhang, D. & Wang, X. Single image dehazing via an improved atmospheric scattering model. *Vis. Comput.* **33**, 1613–1625 (2017).
17. Wang, W., Yuan, X., Wu, X. & Liu, Y. Fast image dehazing method based on linear transformation. *IEEE Trans. Multimedia* **19**(6), 1142–1155 (2017).
18. He, K., Sun, J. & Tang, X. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(12), 2341–2353 (2010).
19. Zhu, Q., Mai, J. & Shao, L. A fast single image haze removal algorithm using color attenuation prior. *IEEE Trans. Image Process.* **24**(11), 3522–3533 (2015).

20. Dong, H., Pan, J., Xiang, L., Hu, Z., Zhang, X., Wang, F. & Yang, M.-H. Multi-scale boosted dehazing network with dense feature fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2157–2167 (2020).
21. Ren, W. et al. 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14. 154–169 (Springer, 2016).
22. Chen, Z., He, Z. & Lu, Z.-M. DEA-net: Single image dehazing based on detail-enhanced convolution and content-guided attention. In *IEEE Transactions on Image Processing* (2024).
23. Liu, W. et al. Image-adaptive yolo for object detection in adverse weather conditions. *Proc. AAAI Conf. Artif. Intell.* **36**(2), 1792–1800 (2022).
24. Kalwar, S., Patel, D., Aanegola, A., Konda, K. R., Garg, S. & Krishna, K. M. GDIP: Gated differentiable image processing for object detection in adverse conditions. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. 7083–7089 (IEEE, 2023).
25. Zhang, Q. & Hu, X. MSFFA-YOLO network: Multi-class object detection for traffic investigations in foggy weather. In *IEEE Transactions on Instrumentation and Measurement* (2023).
26. Zhong, F., Shen, W., Yu, H., Wang, G. & Hu, J. Dehazing and reasoning YOLO: Prior knowledge-guided network for object detection in foggy weather. *Pattern Recogn.* **156**, 110756 (2024).
27. Wang, X., Liu, X., Yang, H., Wang, Z., Wen, X., He, X., Qing, L. & Chen, H. Degradation modeling for restoration-enhanced object detection in adverse weather scenes. In *IEEE Transactions on Intelligent Vehicles* (2024).
28. Ding, X., Zhang, X., Ma, N., Han, J., Ding, G. & Sun, J. RepVGG: Making VGG-style convnets great again. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13733–13742 (2021).
29. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D. & Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1874–1883 (2016).
30. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L. & Timofte, R. SwinIR: Image restoration using Swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1833–1844 (2021).
31. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S. & Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10012–10022 (2021).
32. Jocher, G., Chaurasia, A. & Qiu, J. YOLO by ultralytics, https://github.com/ultralytics/ultralytics, aGPL-3.0 License (2023).
33. Saito, K., Ushiku, Y., Harada, T. & Saenko, K. Strong-weak distribution alignment for adaptive object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* 6956–6965 (2019).
34. Li, S., Ye, M., Zhu, X., Zhou, L. & Xiong, L. Source-free object detection by learning to overlook domain style. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8014–8023 (2022).
35. Ding, Q. et al. CF-YOLO: Cross fusion yolo for object detection in adverse weather with a high-quality real snow dataset. *IEEE Trans. Intell. Transp. Syst.* **24**(10), 10749–10759 (2023).
36. Zhang, Z.-D. et al. CDNet: A real-time and robust crosswalk detection network on jetson nano based on YOLOv5. *Neural Comput. Appl.* **34**(13), 10719–10730 (2022).
37. Wang, Y. et al. TogetherNet: Bridging image restoration and object detection together via dynamic enhancement learning. *Comput. Graphics Forum* **41**(7), 465–476 (2022).
38. Sakaridis, C., Dai, D. & Van Gool, L. Semantic foggy scene understanding with synthetic data. *Int. J. Comput. Vision* **126**, 973–992 (2018).
39. Li, B. et al. Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process.* **28**(1), 492–505 (2018).
40. Hassaballah, M., Kenk, M. A., Muhammad, K. & Minaee, S. Vehicle detection and tracking in adverse weather using a deep learning framework. *IEEE Trans. Intell. Transp. Syst.* **22**(7), 4230–4242 (2020).
41. Huang, S.-C., Hoang, Q.-V. & Le, T.-H. SFA-net: A selective features absorption network for object detection in rainy weather conditions. In *IEEE Transactions on Neural Networks and Learning Systems* (2022).
42. Geiger, A., Lenz, P., Stiller, C. & Urtasun, R. Vision meets robotics: The KITTI dataset. *Int. J. Robot. Res.* **32**(11), 1231–1237 (2013).

Author contributions

The first author, Zhiyong Jing, edited the manuscript and provided experimental assistance. The second author, Zhaobing Chen, was the designer of the methodology and an author of the paper. He was the main contributor to the writing of this paper. Yucheng Shi, the third author and Yufei Gao, the fourth author, provided valuable suggestions and assistance. Corresponding authors Lei Shi and Lin Wei provided comprehensive guidance.

Declarations

Competing interest

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.S. or L.W.

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.