Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.065267

ARTICLE

Deep Learning-Based Algorithm for Robust Object Detection in Flooded and Rainy Environments

Pengfei Wang[1,2,3], Jiwu Sun[2], Lu Lu[1,4], Hongchen Li[1], Hongzhe Liu[2], Cheng Xu[2] and Yongqiang Liu[1,*]

[1]Big Data Center, Ministry of Emergency Management, Beijing, 100013, China
[2]Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing, 100101, China
[3]School of Cyberspace Science and Technology, Beijing Jiaotong University, Beijing, 100044, China
[4]Smart Mining Business Department, Beijing Anxin Entrepreneurship Information Technology Development Co., Ltd., Beijing, 100013, China

*Corresponding Author: Yongqiang Liu. Email: yongqiangliu@sina.com

Received: 08 March 2025; Accepted: 09 May 2025; Published: 03 July 2025

ABSTRACT: Flooding and heavy rainfall under extreme weather conditions pose significant challenges to target detection algorithms. Traditional methods often struggle to address issues such as image blurring, dynamic noise interference, and variations in target scale. Conventional neural network (CNN)-based target detection approaches face notable limitations in such adverse weather scenarios, primarily due to the fixed geometric sampling structures that hinder adaptability to complex backgrounds and dynamically changing object appearances. To address these challenges, this paper proposes an optimized YOLOv9 model incorporating an improved deformable convolutional network (DCN) enhanced with a multi-scale dilated attention (MSDA) mechanism. Specifically, the DCN module enhances the model's adaptability to target deformation and noise interference by adaptively adjusting the sampling grid positions, while also integrating feature amplitude modulation to further improve robustness. Additionally, the MSDA module is introduced to capture contextual features across multiple scales, effectively addressing issues related to target occlusion and scale variation commonly encountered in flood-affected environments. Experimental evaluations are conducted on the ISE-UFDS and UA-DETRAC datasets. The results demonstrate that the proposed model significantly outperforms state-of-the-art methods in key evaluation metrics, including precision, recall, F1-score, and mAP (Mean Average Precision). Notably, the model exhibits superior robustness and generalization performance under simulated severe weather conditions, offering reliable technical support for disaster emergency response systems. This study contributes to enhancing the accuracy and real-time capabilities of flood early warning systems, thereby supporting more effective disaster mitigation strategies.

KEYWORDS: YOLO; vehicle detection; flood; deformable convolutional networks; multi-scale dilated attention

1 Introduction

Flooding under extreme weather conditions poses a major threat to human society, which can not only cause serious property damage but also lead to immeasurable life casualties and property damage. With the rapid development of smart city systems and emergency response technology, object detection technology under complex meteorological conditions has become an important research direction in the field of computer vision [1]. Extreme rainfall also increases the risk of flooding; in the face of such a serious situation, improving the accuracy and real-time disaster warning system is particularly important [2]. However, most of the traditional object detection methods are designed and optimized in conventional environments,

and there is a lack of targeted optimization solutions for the problems of target feature degradation, scale variability, and background interference that occur under complex meteorological conditions [3]. For example, in heavy rainfall and flooding scenarios, the quality of images captured by the camera is often poor due to poor lighting conditions, low visibility, and reflections on the water surface, which makes the subsequent object detection task extremely difficult [4]. In addition, the partial or complete occlusion of vehicles caused by flooding also increases the difficulty of object detection.

In recent years, the rapid development of deep learning technology has provided new ideas and methods to solve the above problems. Convolutional neural networks (CNN), as a representative of them, have shown excellent performance in object detection tasks, however, when facing extreme weather conditions such as flooding and rainfall, traditional CNNs are limited by a fixed geometric sampling structure, which makes it difficult to effectively deal with a series of problems triggered by harsh environments [5]. The existence of these problems requires researchers to develop novel object detection algorithms that are more adaptable to complex weather conditions.

Aiming at the challenge of object detection under flooding and rainfall weather, this paper proposes a YOLOv9 optimization model that combines an improved Deformable Convolutional Network (DCN) and a Multi-scale Dilation Attention mechanism (MSDA). Under severe weather conditions, traditional convolutional operations are difficult to effectively handle the object detection task due to image blurring, reflections, low contrast, etc. DCN ensures high detection accuracy by dynamically adjusting the position of the sampling points so that the network can flexibly locate the key features of the target, and MSDA achieves multi-scale feature learning through different expansion rates and flexibly adjusts the sensory field size to adapt to targets at different scales, effectively coping with complicated background noise, low contrast, and partial occlusion problems. This method aims to improve the performance of YOLOv9 in bad weather conditions and provide more robust technology support for object detection in complex environments.

The main contributions of this paper are as follows:

(1) Deformable Convolutional Network: integrating an improved deformable convolutional layer and modulation mechanism in YOLOv9, the network can adaptively adjust the sampling positions based on the input feature map. This approach allows the model to more accurately capture changes in target morphology in bad weather and adjust the sampling point locations by learning the offsets.
(2) Multi-scale Dilation Attention: increasing the receptive field by different expansion rates without reducing the resolution enables the model to effectively identify multi-scale targets, especially in flooding scenarios where the target changes due to waterlogging and occlusion presentation.

2 Related Work

Object detection in flooded and rainy environments, especially in the fields of autonomous driving and disaster emergency response, faces challenges such as image quality degradation, dynamic noise interference, and degradation of target features. In recent years, with the rapid development of deep learning technology, many researchers have been devoted to improving the robustness and accuracy of object detection algorithms under complex meteorological conditions.

In terms of rain removal techniques, Yang et al. [6] proposed an algorithm based on multi-scale feature extraction, which improves the network's ability to generalise to real images by incorporating spatial information and enhances the performance of existing rain removal algorithms by using multi-scale extraction. Zhou et al. [7] proposed PARDNet, which reduces the depth of the network by using two parallel self-networks for extracting the rain pattern and reduces the risk of overfitting. It also expands the receptive field by using null convolution so that it can capture more contextual information about the rain pattern.

Wang et al. [8] proposed RCDNet which is an interpretable rain pattern convolutional dictionary network that removes the rain pattern while maintaining good interpretability by introducing a dynamic rain pattern kernel inference mechanism. Gao et al. [9] proposed a contextual aggregation-based detail restoration de-raining network that aimed at effectively removing rain patterns from images and recovering background details. Through a comprehensive benchmark analysis of existing single-image rain removal algorithms, Yu et al. [10] proposed a rain removal method that is robust against adversarial attacks, combining multi-scale feature extraction and spatial attention mechanisms to enhance the model's ability to recognise rain patterns of different scales and directions. Although the above methods achieved excellent results in image rain removal, they were not validated in combination with downstream tasks such as object detection, which could not highlight their improved effectiveness for rainy day object detection tasks.

ARODNet proposed by Qiu et al. [11] is an adaptive rainy-day image enhancement and object detection network for autonomous driving applications under severe weather conditions. The method significantly improves the rain removal effect and background detail recovery by combining multi-scale feature extraction, spatial attention mechanism, and adaptive enhancement strategy, which enhances the accuracy of object detection. Zhou et al. [12] proposed SSDA-YOLO, a semi-supervised domain adaptive YOLO network for addressing the challenges in cross-domain object detection tasks, which achieves more effective knowledge transfer between different domains. Liu et al. [13] proposed Image-Adaptive YOLO, which significantly improves the performance of object detection in complex environments, such as rainy and foggy days, by introducing an image-adaptive mechanism that dynamically adjusts the model parameters to adapt to the image features under different weather conditions. Hu et al. [14] proposed DAGL, a fast domain adaptive Faster R-CNN (Region-based Convolutional Neural Networks) framework for vehicle object detection under severe weather conditions, which achieves more accurate vehicle detection under different weather conditions. Although the above literature has improved the detection performance in flooding and rainy day object detection through various methods, its effectiveness in dealing with extreme occlusion and complex background interference still needs to be improved.

Image denoising, as a classical problem in the field of underlying vision, has long faced the challenge of how to effectively remove noise while preserving texture details, which also has an impact on cloudy and rainy day target detection tasks. Zhang et al. [15] proposed a method called Robust Deformed Denoising Convolutional Neural Network (RDDCNN) to address the problem that the convolution operation during image denoising may change the original distribution of noise in the noise-containing image, thus increasing the training difficulty. Quan et al. [16], considering the limitations of traditional denoising algorithms in dealing with real image noise, proposed a hybrid Transformer-CNN architecture for real image denoising, aiming to combine the advantages of the two models to achieve better denoising results. Zhang et al. [17] learned a deep CNN denoiser prior to improve the efficiency of prior knowledge utilisation of deep learning models for image denoising, which showed good performance in image denoising and provided new ideas for image restoration. Ren et al. [18] addressed the problem that existing image denoising methods cannot adaptively adjust the denoising intensity, and proposed a deep network based on adaptive consistency prior to image denoising, which can be adaptively adjusted according to the image content, thus improving the denoising effect.

Currently, there is a lack of effective dedicated network models for object detection in complex environments, especially in accurately locating disaster victims or damaged materials. Despite the progress in the field of weather robust object detection, the existing methods still face challenges in dealing with extreme weather, such as low contrast, background noise interference, and partial occlusion, which make it difficult to meet the needs of practical disaster relief. By introducing DCN and MSDA, this study aims to improve the model's object detection capability under extreme conditions, provide more accurate and reliable technical

support for emergency rescue, enhance the model's understanding and adaptability to complex scenes, and improve the accuracy and robustness of target recognition under severe weather.

3 Methodology

3.1 YOLOv9 Network Structure

YOLOv9 (You Only Look Once Version 9) [19], as a single-stage object detection framework, achieves significant breakthroughs in model architecture design and gradient propagation optimisation by introducing innovative mechanisms such as Programmable Gradient Information (PGI) and Generalized Efficient Layer Aggregation Network (GELAN) and other innovative mechanisms, a significant breakthrough was achieved in the design of the model architecture and the optimisation of gradient propagation, and the structure of its network model is shown in Fig. 1. The model continues the core advantages of real-time detection of YOLO series, and at the same time, through multi-level feature fusion and lightweight structural design, it reaches a new technological height in the balance between detection accuracy and inference efficiency, providing a better solution for object detection tasks in complex scenarios.

By decoupling the gradient path from the feature extraction path, YOLOv9 achieves fully differentiable gradient information programming in a single-stage detector for the first time, which provides a new research direction for the optimization theory of deep networks. However, its dynamic computational graph mechanism may increase the memory consumption in the training phase, and the sensitivity to hyperparameters (e.g., PGI branch weight coefficients) still needs to be further explored. Future work can focus on model compression and hardware adaptation to promote its wide application in embedded systems.

In this study, YOLOv9 is chosen as the basic framework, mainly based on its innovative PGI and GELAN designs. PGI effectively alleviates the gradient degradation problem of small target detection in bad weather by explicitly retaining shallow gradient information through assisted supervised branching, while the heterogeneous convolutional combination of GELAN enhances feature diversity while reducing the number of parameters, providing more flexible underlying architectural support for subsequent integration of DCN and MSDA modules. Provides a more flexible underlying architecture support for the subsequent integration of DCN and MSDA modules. In contrast, the core improvements of YOLOv10 or YOLOv11 (e.g., dynamic label assignment or quantization-friendly design), although optimized in terms of speed, do not directly address the problem of geometrical variations and multiscale noise under complex meteorological conditions. In addition, the gradient path decoupling mechanism of YOLOv9 allows modular embedding of DCN and MSDA without disrupting the original gradient flow. For example, DCN's offset learning relies on stable gradient backpropagation, while YOLOv9's PGI effectively avoids the problem of offset learning failure due to network depth. If directly migrating to YOLOv10 or YOLOv11, its dynamic label assignment strategy may conflict with DCN's geometric adaptive sampling, requiring additional tuning costs.

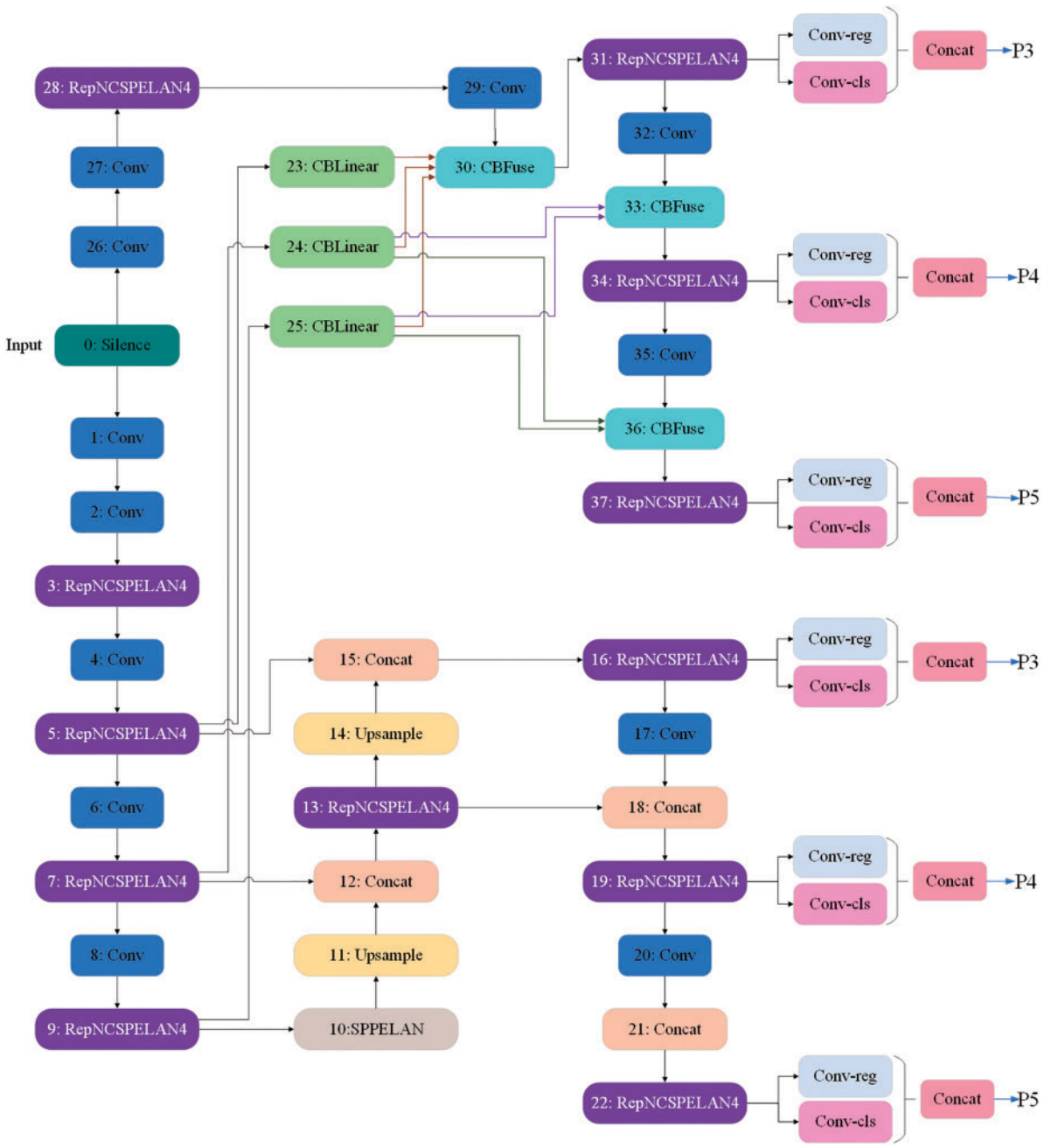

Figure 1: YOLOv9 network structure diagram

3.2 Overall Network Structure

In complex and dynamically changing environments, such as rainy and flooding scenes, traditional target detection methods often face many challenges. These conditions not only degrade the image quality but also introduce a lot of noise and ambiguity, which greatly affects the detection performance. To address this challenge, this paper proposes a method based on the combination of improved DCN, MSDA, and YOLOv9 framework. The network structure is shown in Fig. 2, with the improvements in the orange part (DCN) and the yellowish part (MSDA), which are all highlighted with red dashed boxes in the figure. By

embedding the DCN and MSDA modules into the different layers of YOLOv9, the network can efficiently incorporate multilevel features. The orange section (DCN) demonstrates the integration of the DCN in the early and middle layers of the network to enable adaptive sampling grid adjustment strategies and feature amplitude modulation techniques. Replacing some of the standard convolutional layers in the YOLOv9 backbone network enhances the robustness of the model to target deformation, blurring, and reflective noise in flood/rain scenarios by dynamically learning the sampling point offsets (Δp_n) and feature modulation scalars (Δm_k). This enhances the model's adaptability to target deformation and noise interference, especially when dealing with problems triggered by harsh environments. The yellowish section (MSDA) highlights the location of the MSDA module, which is located in the deeper layers of the network and aims at integrating information from different scales to ensure that the model can handle multi-scale targets effectively. By introducing convolutional layers with different expansion rates to capture multi-scale contextual information and weighting this information through the attention mechanism, the multi-scale target features are captured adaptively to effectively solve the problems of waterlogged occlusion and scale diversity, which enables the model to have higher detection accuracy in flooding and rainy day scenarios.

In YOLOv9's base feature extraction network, some of the standard convolutional layers are replaced with deformable convolutional layers. This allows the network to adaptively adjust the position of the sampling grid according to the input feature map, enabling the network to better capture the detailed features of the target object even under low-contrast or blurred image conditions. To further improve the spatial support region modulation capability of the model, a modulation mechanism is further introduced. Each sampling point can not only be positionally adjusted by applying an offset, but also the amplitude of the input features can be modulated by a learned modulation scalar. This approach allows the network to ignore background information that may interfere with the classification decision and focus on the target object itself. Considering the large differences in target sizes that may occur in flooding scenarios, this paper adopts a multi-scale testing strategy by introducing a multi-scale inflationary attention module for dealing with targets at different scales and enhancing the focus on the target's key features, employing multiple convolutional layers with different inflation rates to capture the multi-scale contextual information, and weighting the obtained information through the attention mechanism to ensure that the model can work effectively at different scales. By embedding the DCN and MSDA modules into the YOLOv9 framework, the network not only improves the robustness and accuracy of the model under complex meteorological conditions but also demonstrates its great potential in practical applications. This innovative design not only enhances the model's ability to adapt to geometric transformations and multi-scale target identification but also provides strong technical support for its wide range of applications in areas such as disaster emergency response.

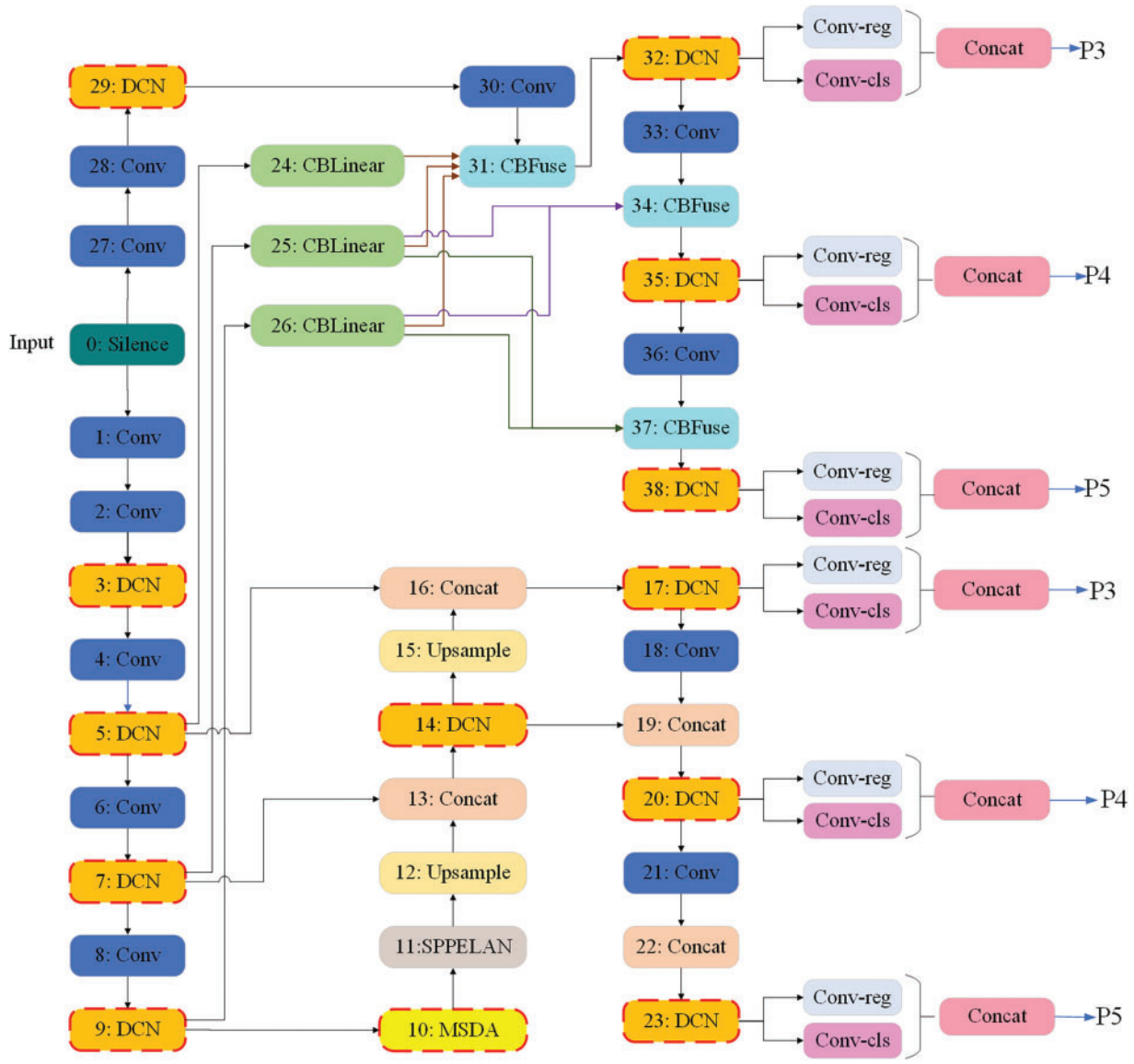

Figure 2: Structure of YOLOv9 optimisation model combining improved DCN and MSDA mechanisms

3.3 Deformable Convolutional Networks

Conventional CNNs have inherent limitations in handling geometric transformations such as object scale, pose, viewpoint, and partial deformation due to their fixed geometric structure [20]. Aiming at the problems of significant changes in target shape due to rain blurring or reflections, serious interference of target recognition by background noise, and the presence of a large number of targets of different sizes that are difficult to capture accurately in the vehicle detection task in rainy and flooded scenarios, we propose to apply the improved Deformable Convolutional Networks (DCN) [21] to the solution in the YOLOv9 network framework. By introducing deformable convolutional layers and modulation mechanisms, the adaptive adjustment ability of the model to geometric transformations is realised, which effectively solves the limitations of the traditional convolutional operation in dealing with irregular shapes and dynamic environment changes.

For the problem of target shape change, DCN dynamically adjusts the position of the sampling grid by learning the offset, which enables the network to more accurately capture the morphological changes of the target object due to bad weather conditions and improves the detection accuracy. Secondly, to solve the problem of background noise interference, we further introduce a modulation mechanism based on DCN, which not only allows adjusting the positional offsets of the sampling points but also regulates the input feature amplitude according to different spatial locations, thus reducing the influence of background information on the prediction results.

Conventional CNNs are limited by a fixed geometric sampling structure (e.g., regular grid points), which makes it difficult to effectively model geometric transformations (e.g., scale, gesture, deformation, etc.) of an object in an image. DCN is a convolutional neural network that can adapt to geometric transformations and it enhances the ability of CNNs to model geometric transformations by introducing two new modules: deformable convolution and deformable RoI pooling. These modules are based on the idea of adding spatially sampled positional offsets and learning these offsets from the target task without additional supervision.

(1) Deformable convolution: The sampling position of the standard convolution is defined by a fixed grid R. As an example, for a 3 × 3 convolution kernel, the corresponding grid is R = {(−1, −1), (−1, 0), . . ., (1, 1)}. For the output position p_0, the formula is shown in Eq. (1):

$$y(p_0) = \sum_{p_n \in \mathcal{R}} w(p_n) \cdot x(p_0 + p_n) \tag{1}$$

where $w(p_n)$ denotes the weight parameter at a sample point p_n of the convolution kernel and $x(p_0 + p_n)$ denotes the input feature map at the position $p_0 + p_n$, deformable convolution adds a learnable offset Δp_n to this for each sample point, the formulae are as shown in Eq. (2):

$$y(p_0) = \sum_{p_n \in \mathcal{R}} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \tag{2}$$

where Δp_n is learned from the input features by an additional convolutional layer. Since the offsets may be fractional, the feature values are computed by bilinear interpolation:

$$x(p) = \sum G(q, p) \cdot x(q) \tag{3}$$

where $G(\cdot,\cdot)$ is the bilinear interpolation kernel function. Fig. 3 illustrates a comparison of the sampling positions of the standard and deformable convolutions.

Fig. 4 shows a schematic of the operation of 3 × 3 deformable convolution. In standard 2D convolution, the input feature map x is first sampled using a regular grid R, and then the sampled values are weighted and summed by the weights w. The input feature map x is then summed by the weights w. In deformable convolution, the input feature map x is then summed by the weights w. Whereas in deformable convolution, this regular grid R is augmented with offsets $\{\Delta p_n | n = 1, \ldots, \mathrm{N}\}$, where $N = |R|$ denotes the number of grid points. This means that the originally fixed sampling positions are now irregular and with offsets $p_n + \Delta p_n$. Since these offsets Δp_n are usually in fractional form, the implementation is done by bilinear interpolation as shown in Eq. (3).

As can be seen in Fig. 3, these offsets are obtained by applying a convolutional layer to the same input feature map. This convolutional layer has the same spatial resolution and expansion coefficient as the current convolutional layer (e.g., also a 3 × 3 kernel with an expansion coefficient of 1 in Fig. 3). The output offset field has the same spatial resolution as the input feature map, while the channel dimension $2N$ corresponds to N 2D offsets. The convolution kernel that generates the output features and offsets is learned simultaneously during training.

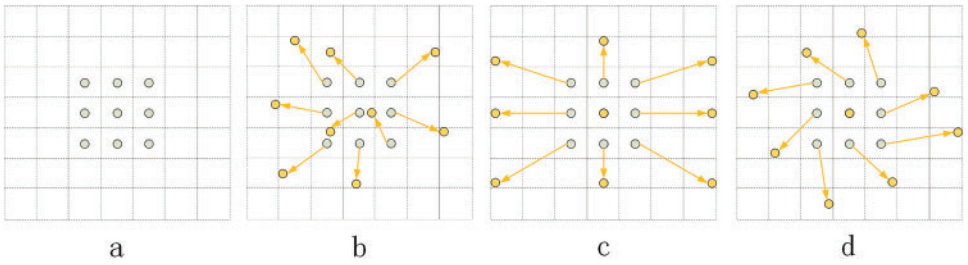

Figure 3: DCN convolution structure diagram. **(a)** the regular sampling grid of the standard convolution (indicated by grey dots), where the sampling positions of a 3 × 3 standard convolution kernel on the input feature map are shown; **(b)** the deformed sampling positions in the deformable convolution (indicated by yellow dots), as well as the additional offsets (indicated by yellowish arrows), which show the the regular sampling grid of the standard convolution is adapted by the learned offsets, making the sampling more flexible and able to adapt to the different geometric transformations of the input; **(c, d)** are examples of the special case of **(b)**, showing how the deformable convolution can be generalised to handle transformations such as scale changes and rotations

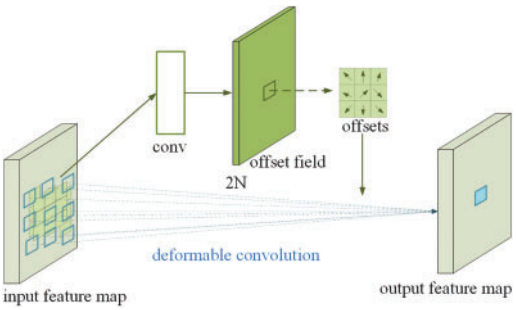

Figure 4: 3 × 3 deformable convolutional map

(2) Deformable RoI pooling: Conventional RoI pooling divides the region of interest (RoI) into a fixed grid of $k \times k$ spatial cells. Deformable RoI pooling introduces an offset Δp_{ij} for each cell to achieve adaptive partial localisation, especially for objects with different shapes, where the offsets are learned from the RoI features through a fully connected layer and normalised by the dimensions of the RoI to maintain scale invariance. The schematic diagram of deformable RoI pooling is shown in Fig. 5 and the equation is shown in Eq. (4):

$$y(i,j) = \sum_{p \in \text{bin}(i,j)} x\left(p_0 + p + \Delta p_{ij}\right)/n_{ij} \tag{4}$$

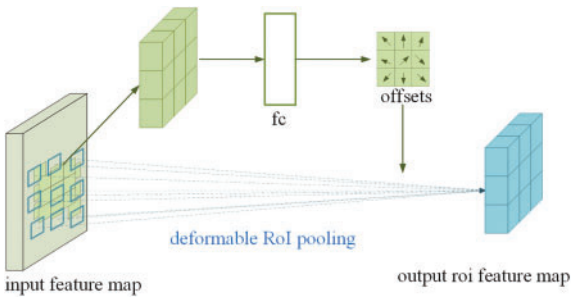

Figure 5: Illustration of 3 × 3 deformable RoI pooling

Both modules are lightweight, require only a few parameters and computations to learn offsets, can easily replace common modules in existing CNNs, and can be easily trained end-to-end with standard back-propagation. In addition, the modulation mechanism introduced in the deformable convolution module, which learns both the offset Δp_k and the modulation scalar $\Delta m_k \in [0,1]$ for each sampling point as shown in Eq. (5), endows the network with more fine-grained control over spatial support regions. This mechanism not only allows the model to learn and apply offsets to adjust the spatial sampling locations of the input features but also introduces the concept of feature amplitude modulation, whereby the input features at each sampling point can be scaled up or down according to their corresponding modulation scalars. This modulation process is achieved through additional learning parameters that determine the feature weights at each sampling point or location.

$$y(p) = \sum_{k=1}^{K} w_k \cdot x\left(p + p_k + \Delta p_k\right) \cdot \Delta m_k \tag{5}$$

The modulation scalar is generated via a Sigmoid function that allows the network to dynamically suppress or enhance the contribution of specific sample points. This design significantly improves the robustness of the model to complex backgrounds. In extreme cases, if the modulation scalar at a particular location is set to zero, the features at that location will not have any effect on the final output, meaning that the network can effectively ignore information from irrelevant or insignificant spatial locations. This modulation mechanism therefore provides another degree of freedom for the network to modulate its spatial support region more flexibly and precisely, thereby optimising the focus on the features of the target object and reducing interference from background noise or other non-critical information.

By combining offset learning with feature amplitude modulation, the deformable convolutional network not only enhances its adaptability to geometric transformations but also improves the model's ability to focus on target objects, which in turn improves the performance of visual tasks such as object detection and instance segmentation. This dual modulation strategy enables the network to capture object features of interest more accurately in complex image environments while suppressing unnecessary details, demonstrating the model's high flexibility and robustness in processing natural scenes.

In complex meteorological conditions, such as flooding and rainy environments, the object detection task faces many challenges, including, but not limited to, image quality degradation due to low contrast, blurring, and reflections etc. The DCN has demonstrated significant advantages in addressing these challenges through its unique mechanism design.

3.4 Multi-Scale Dilated Attention

When the object detection task is performed under flooding and rainy conditions, the background clutter in the image increases, the target object may be partially occluded, and the environmental factors can seriously affect the visibility and clarity of the target, which increases the detection difficulty. Traditional CNN-based methods may suffer from performance degradation due to the lack of sufficient contextual information when dealing with such complex scenes. In contrast, the MSDA [22] mechanism can effectively capture multi-scale features, which can improve the model's focus on key regions, enhance recognition capabilities, and improve the applicability of the model in different environments, which is crucial for improving the accuracy and robustness of object detection.

In traditional convolution-based neural networks, downsampling or convolutions with large step sizes are often used to increase the sensory field and reduce the computational cost. However, these methods lead to a decrease in feature map resolution, which affects model performance in tasks such as object detection

and semantic segmentation. MSDA first divides the channels of the feature map into different heads and then performs self-attention operations with different swelling rates within a fast window around the query. As shown in Eq. (6), for a query at position (i, j) on the original feature graph, the SWDA operation sparsely selects keys and values for the self-attention operation in a sliding window of size $w \times w$ centered on (i, j). In addition, to further exploit the information within the RoI, MSDA simultaneously captures contextual semantic dependencies at different scales. Different heads are set with different expansion rates r, enabling the capability of multi-scale representation learning.

$$X = \mathrm{SWDA}\left(Q, K, V, r\right) \tag{6}$$

where Q, K and V represent query, key and value matrices, respectively. The rows of each matrix represent individual query/key/value feature vectors. For the expansion rate $r \in N^{+}$ is used to control the degree of sparsity. In particular, for position (i, j), the corresponding component x_{ij} of the output X is defined by Eq. (7):

$$x_{ij} = Attention\left(q_{ij}, K_r, V_r\right) = Softmax\left(\frac{q_{ij}K_r^T}{\sqrt{d_k}}\right)V_r \tag{7}$$

where q_{ij} represents the feature vector at position (i, j) in the query matrix, and d_k denotes the dimensional size of the key vector, i.e., the dimensionality of the key vector space. K_r and V_r represent the keys and values selected from the feature maps K and V.

The overall architecture of multiscale inflated attention is designed to efficiently fuse multiscale feature representations, as shown in Fig. 6, where the model consists of an overlapping tokeniser, an overlapping downsampler, an MSDA block, and a Multihead Self Attention (MHSA) block. Each MSDA block contains Depthwise Separable Convolution (DSConv), Multi-scale Sliding Window Dilated Attention operation (SWDA) and MLP (Multilayer Perceptron). By default, a kernel size of 3 × 3 is used and the expansion rates are set to r = 1, 2 and 3, and the attentional receptive field sizes of different heads are 3 × 3, 5 × 5, and 7 × 7. In practice, MSDA achieves multi-scale feature learning by using different expansion rates in different heads. This not only enhances the model's ability to recognise targets at different scales but also effectively reduces the redundant computation in the self-attention mechanism. In addition, the expressiveness of the model is further enhanced by splicing the outputs from different heads and feeding them into a linear layer.

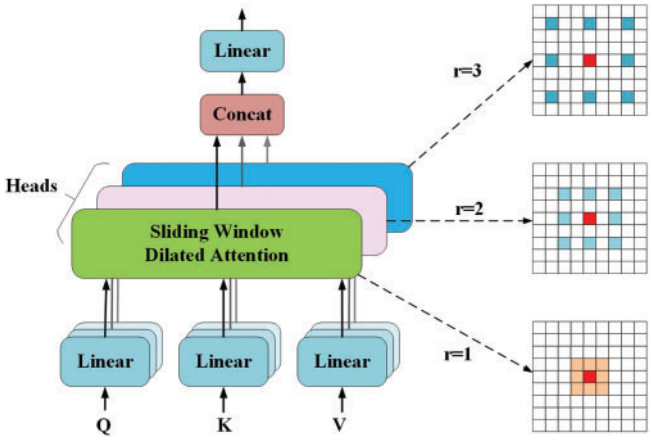

Figure 6: Multi-scale dilation attention

4 Experimental Setup and Dataset

4.1 Experimental Setup

To validate the effectiveness of the YOLOv9 optimisation model, combining the improved DCN and MSDA mechanisms for flood and rainfall weather object detection, the experimental setup is as follows:

This study uses an NVIDIA GeForce RTX 2080 GPU (Santa Clara, CA, USA) equipped with 12 GB GDDR6 memory. The system also includes an Intel Core i9-12900K CPU and 64 GB DDR4 memory to ensure efficient data loading and preprocessing. The model is built based on the PyTorch framework and utilises CUDA 12.1 for GPU acceleration, Ubuntu is chosen for the operating system and all code is executed in a Python 3.9 environment. Integrate OpenCV, NumPy, and other key-dependent libraries to improve the efficiency of image processing, numerical computation, and data enhancement.

The model was trained for 200 epochs, with the batch size set to 8 to balance the memory utilisation and training efficiency. For the added DCN layer, a method that helps to learn the effective offset quickly while keeping the model stable is used. To enhance the multi-scale feature learning capability, a feature imitation loss function is introduced to reduce the impact of background noise. The optimisation algorithm chooses stochastic gradient descent (SGD) with an initial learning rate of 0.001 and a cosine annealing strategy, with the weight decay coefficient set to 0.05 to prevent overfitting. For the loss function design, in addition to the standard object detection loss, a modulation scalar loss is added to help the model focus more on the target region and improve the overall performance. Together, these measures improve the robustness and accuracy of the model under complex meteorological conditions.

4.2 Dataset Selection

In exploring object detection algorithms applicable to flooding and rainfall weather conditions, this chapter focuses on the ISE-UFDS and UA-DETRAC datasets.

The ISE-UFDS dataset focuses on the task of vehicle object detection in flooding scenarios, which covers the challenges of low visibility, light changes, and reflections, and is a dataset specifically designed for vehicle hazard level detection in urban flooding scenarios, containing 20,152 images from real flooding sites that show vehicles in different flood risk levels and time periods, and is divided into training, validation and test sets in the ratio of 8:1:1. It can effectively simulate the common ground visual obstacles in flooding scenarios. Experiments using ISE-UFDS can help to evaluate and optimise the performance of the improved object detection algorithm in dealing with adverse conditions, thus improving the applicability of the model in real flooding environments.

UA-DETRAC [23] is an authoritative benchmark dataset jointly constructed by the Beijing Institute of Technology, the University of Technology Sydney, and other research institutes, designed for vehicle detection and multi-target tracking tasks from the perspective of low-altitude UAVs (Unmanned Aerial Vehicle). The dataset systematically covers the multi-dimensional challenges in real environments, including key issues such as dynamic changes in illumination, target scale diversity, dense occlusion, and dynamic background interference, by capturing aerial videos in complex urban traffic scenes. Its core objective is to provide a standardised evaluation platform for the computer vision field and to promote the application of UAV perception technology in real-world scenarios such as intelligent transportation and urban surveillance.

Although UA-DETRAC was not originally designed to focus on flooding or extreme weather conditions, its rich vehicle types and accurate labeling make it ideal for testing the robustness of object detection algorithms. What's more, the research in this chapter extends the UA-DETRAC dataset by synthetic techniques to introduce specific rainfall weather factors, such as raindrop interference [24], named UA-Rain,

and some of the synthetic sample images are shown in Fig. 7, to further enrich the diversity of the training samples and to enhance the model's ability to generalise in complex environments.

Figure 7: Sample image of the synthetic rainy dataset

The combination of ISE-UFDS and UA-DETRAC can enhance the effectiveness of object detection algorithms for research on multiple levels. Firstly, the unique challenges provided by ISE-UFDS contribute to the development of algorithms that can work stably under extreme visual conditions, while UA-DETRAC complements the large amount of data on real-world application scenarios and enhances the algorithms' adaptation to the real world. Secondly, the combination of the two can build a more comprehensive training framework that covers not only the analysis of static images but also the understanding of dynamic scenes, which is especially crucial for achieving an efficient and accurate vehicle detection system. This domain fusion approach to data integration also promotes technical exchange and innovation between different domains, promotes algorithmic innovation in application development, and improves the detection performance of the model.

5 Comparative Analysis of Experimental Results

5.1 Evaluation Indicators

In this study, to quantify the performance of vehicle detection algorithms in flooding scenarios, three standard performance evaluation metrics are used: Precision, Recall, and mean Average Precision (mAP) [25]. The precision rate aims to measure the proportion of vehicle targets identified by the model that are correct, as shown in Eq. (8), while the recall rate is used to assess the proportion of all present vehicles that the model can successfully identify, regarding Eq. (9). In addition, to comprehensively consider the balanced relationship between precision and recall, the metric F1-score is introduced, which can mitigate the effects of category imbalance to a certain extent and provide a balanced reflection of the overall performance of the algorithm, which is calculated as shown in Eq. (10). As a comprehensive performance metric widely used in object detection tasks, mAP comprehensively evaluates the detection capability of the model under various difficulty levels by considering the average accuracies under different Intersection over Union (IoU) thresholds. Especially in the flooding scenario, as some vehicles may have blurred contours due to flooding,

which increases the difficulty of detection, at this time, mAP better reflects the stability and accuracy of the model in this complex environment, and its specific calculation method follows Eqs. (11) and (12).

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{10}$$

$$AP = \sum_{i=1}^{n} (Precision_k \times (Recall_k - Recall_{k-1})) \tag{11}$$

$$mAP = \frac{\sum_{j=1}^{k} AP_j}{k} \tag{12}$$

where TP, FP, FN denote the number of true cases, false positive cases, and false negative cases, respectively, $Precision_k$ is the precision value when the recall rate is $Recall_k$, n is the number of points of all possible recall rates, and k is the number of APs in each category.

5.2 Comparative Analysis of Models

In this study, an improved model is constructed by introducing the DCN module and MSDN module to YOLOv9, and a comprehensive comparative performance evaluation is conducted. The experimental results are presented in Table 1. In the following sections, we will conduct an in-depth analysis of each model's performance across multiple dimensions. This detailed analysis aims to underscore the notable advantages offered by the improvements introduced in this study, specifically in the context of object detection tasks oriented towards flood and rainfall weather conditions.

Table 1: Comparison of model effects on the ISE-UFDS dataset

Model	Precision	Recall	F1-Score	mAP
Mask R-CNN	72.5%	76.3%	74.4%	79.6%
Faster R-CNN	69.0%	73.2%	71.0%	78.2%
Cascade R-CNN	70.0%	75.0%	72.4%	78.9%
YOLOv5	69.1%	75.6%	72.2%	75.8%
YOLOv9	69.1%	81.5%	74.8%	79.1%
Ours	70.0%	81.6%	75.4%	79.6%

Precision is an important measure of a model's ability to correctly identify. According to the data in Table 1, Ours achieved 70% in terms of accuracy, which is an improvement over YOLOv9's 69.1%. This improvement demonstrates that by integrating the DCN and MSDN modules, it is possible to make the YOLOv9 base network effective in reducing the false detection rate and improving the reliability of the detection results. In addition to this, the accuracy of the improved model exceeds that of other traditional

object detection models such as Faster R-CNN [26], Cascade R-CNN [27] and YOLOv5, which further validates its excellent performance in high-precision detection.

The recall rate reflects the proportion of all actual targets successfully detected by the model. Ours achieves 81.6% on this metric, which is higher than the base model, YOLOv9, and significantly ahead of other traditional models. The Ours model can maintain a high recall rate while effectively controlling the false alarm rate to ensure the accuracy and practicality of the detection results. This balance not only improves the model's ability to detect targets in complex environments but also enhances its robustness in coping with the task of vehicle detection under flooding and rainfall weather conditions. Especially when facing extreme weather conditions, the Ours model can accurately identify and locate a variety of vehicles by its feature extraction capability and multi-scale detection mechanism, and maintains a high level of detection accuracy even in harsh environments such as low visibility and reflective interference.

As one of the key indicators for comprehensively evaluating model performance, the F1-score of the Ours model is 75.4%, which is the highest among the compared models, indicating that the Ours model strikes a good balance between precision and recall, and can effectively cope with the challenge of object detection in complex scenarios. The F1-score of YOLOv9 is 74.8%, which is slightly lower than that of the Ours model, considering that the overall performance of the Ours model in terms of precision and recall is better. Slightly lower than the Ours model, and considering the combined performance of the Ours model in terms of precision and recall, its overall performance is much better. The F1-scores of other traditional models, such as Mask R-CNN, Faster R-CNN, and YOLOv5 are lower than those of the Ours model, which further validates the superiority of the Ours model in complex environments. The metric results fully demonstrate the advantages of the Ours model in handling object detection tasks of different categories and scales and its potential value in practical applications.

In terms of mAP, the Ours model achieves 79.6%, which is on par with the Mask R-CNN but significantly higher than the other models, reflecting the fact that the Ours model performs well in the task of object detection at different scales with high robustness and generalisation, and also highlights its superiority in the task of vehicle detection under flooding conditions. Under the challenges of low visibility and reflective interference, the Ours model achieves effective adaptation and accurate detection in complex environments. These improvements make the Ours model more adaptable and reliable in object detection tasks oriented to flooding and rainfall weather, providing strong technical support for practical applications.

Some of the experimental results are shown in Fig. 8. By comparing the results of the two rows, it can be observed that although the improved model shows relatively low confidence scores in some cases, there is an obvious omission in the detection results of the base YOLOv9 model, i.e., it fails to identify some of the vehicles in the scene. In contrast, the improved model proposed in this paper effectively ameliorates this problem and significantly reduces the occurrence of missed detections. This suggests that, despite the decrease in confidence under specific conditions, the approach in this paper improves the overall accuracy and completeness of object detection by optimising the detection mechanism and enhancing the model's adaptability to complex environments, and especially demonstrates superior performance when dealing with the challenges of flooding and rainfall weather conditions. Therefore, the improvement strategy is important for enhancing the robustness and reliability of the object detection system.

Figure 8: Experimental visualisation results. **(a)** demonstrates the object detection performance of the base YOLOv9 model. **(b)** presents the de-tection effect of the improved method proposed in this paper

The results obtained by training the UA-DETRAC dataset with the improved method are shown in Table 2 in comparison with the traditional model. Comparing the experimental results, it can be seen that by integrating the DCN module and the MSDA module, the proposed improved model can improve the training of different datasets and achieve a good recall rate while maintaining high precision, ensuring the accuracy and practicality of the detection results. On the UA-DETRAC dataset, the improved model achieves a mAP of 70.1%, which is 0.7 percentage points higher than YOLOv9. Despite the slight decrease in precision, the recall improves from 69.0% to 69.6%, showing the stronger robustness of the model in complex scenarios. This indicates that the proposed model is more adaptable and reliable in dealing with object detection tasks in different complex scenarios, which provides strong technical support for practical applications, especially in enhancing the effectiveness of the traffic warning system in times of disaster.

Table 2: Comparison of models on the UA-DETRAC dataset

Model	Precision	Recall	F1-score	mAP
Mask R-CNN	71.1%	66.8%	68.9%	66.8%
Faster R-CNN	69.3%	70.5%	69.9%	68.5%
Cascade R-CNN	67.2%	66.1%	66.6%	66.1%
YOLOv5	71.0%	68.2%	69.6%	68.2%
YOLOv9	74.6%	69.0%	71.7%	69.4%
Ours	74.5%	69.6%	72.0%	70.1%

Meanwhile, to verify the effectiveness of the improved model with other vehicle detection network models in the vehicle detection task under normal scenarios, in this study, Kang et al. proposed an improved YOLO detector based on Type-1 fuzzy attentional mechanism (YOLO-FA) to address the detection accuracy of vehicle detection models in complex environments; and Wang et al. specifically targeted at the small-target vehicle The YOLOv5-NAM vehicle detection model designed by Wang et al. is trained on the UA-DETRAC dataset as a comparison, in which the mAP value of YOLO-FA is 70%, and that of YOLOv5-NAM is 50.8%,

which shows that the proposed method in this paper outperforms the current more excellent network models in the vehicle detection task.

To deeply evaluate the performance of different object detection models in the simulated raindrop environment, this study conducted detailed model comparison experiments on the UA-Rain dataset, and the experimental results are shown in Table 3.

Table 3: Comparison of models on the UA-Rain dataset

Model	Precision	Recall	F1-Score	mAP
YOLOv9	72.3%	65.3%	71.7%	66.2%
Ours	71.8%	67.5%	72.0%	67.7%

In terms of precision, the Ours model achieves 71.8%, slightly lower than YOLOv9's 72.3%. In terms of recall, the Ours model achieves 67.5%, which is somewhat higher than the 65.3% of YOLOv9. This shows that the Ours model is able to identify and locate the target more effectively while maintaining higher precision, reducing the phenomenon of missed detection and ensuring the comprehensiveness and reliability of the detection results. The F1-score of the Ours model is 72.0%, which is slightly higher than the 71.7% of YOLOv9. This result further validates that the Ours model strikes a good balance between precision and recall, and is able to achieve efficient and accurate object detection in complex rainy weather environments. Especially in the face of challenges such as low visibility and reflective interference, the Ours model is able to identify and locate various types of vehicles more accurately by virtue of its powerful feature extraction capability and multi-scale detection mechanism, and maintains high detection accuracy even under extreme weather conditions.

In terms of mAP, the mAP of the Ours model reaches 67.7%, which is significantly higher than the 66.2% of YOLOv9. This not only reflects that the Ours model performs well in different categories and scales of object detection tasks with high robustness and generalisation ability, but also highlights its superiority in flood and rainfall weather-oriented object detection tasks.

The results in Table 3 further validate the effectiveness of DCN and MSDA. Under simulated rainfall conditions, the mAP of the improved model improves by 1.5 percentage points (from 66.2% to 67.7%) compared to YOLOv9, while the recall also improves (from 65.3% to 67.5%), which demonstrates that the proposed module is able to effectively deal with the challenges of low visibility and reflective interference.

5.3 Analysis of Ablation Experiment Results

In order to deeply investigate the specific contributions of the DCN and MSDA modules in improving the model, this study conducts detailed ablation experiments on the ISE-UFDS dataset. By comparing the model performance under different configurations, it is possible to demonstrate the role of each module in improving the overall results.

As shown in Table 4, when only the DCN module is added, the model precision improves from 69.1% to 70.9% in YOLOv9, while the recall decreases from 81.5% to 78.7%. Despite the decrease in recall, the F1-score remains at 74.6%, which is the same as YOLOv9. In terms of mAP, it improves to 79.6%, which is 0.5 percentage points higher than YOLOv9. This indicates that the DCN module improves model accuracy while enhancing its robustness in complex environments, especially when dealing with challenges such as reflective interference. The DNC module can capture the deformation characteristics of the target more efficiently, thus improving the overall detection performance results. When the MSDA module is added, the precision

drops to 68.6%, but the recall rate increases to 81.6%. The F1-score is 74.5%, which is slightly lower than that of YOLOv9, however, the mAP is improved to 80.3%, which is 1.2 percentage points higher than that of YOLOv9, indicating that the MSDA module has a significant advantage in multiscale object detection, and is able to identify targets of different sizes efficiently, especially in flooding and rainfall weather conditions, where vehicles may show multiscale due to waterlogging, shading and other changes, the MSDA module can better adapt to these changes and improve the accuracy and comprehensiveness of detection. In addition to this, this study compares the inference speed of the improved model, and there are experimental results that show that the FPS metrics of the proposed method in this study are greatly improved compared to the basic YOLOv9 network, which can better satisfy the task of search and rescue of vehicular targets in flooding and rainfall scenarios.

Table 4: Ablation experiments on the ISE-UFDS dataset

Model	DCN	MSDA	Precision	Recall	F1-Score	mAP	FPS
YOLOv9	–	–	69.1%	81.5%	74.8%	79.1%	47
	√	–	70.9%	78.7%	74.6%	79.6%	53
Ours	–	√	68.6%	81.6%	74.5%	80.3%	60
	√	√	70.0%	81.6%	75.4%	79.6%	62

The DCN and MSDA modules are simultaneously integrated into the YOLOv9 network to evaluate the combined effect. The experimental results show that at this point, the model achieves 70% precision, 81.6% recall, 75.4% F1-score, and 79.6% mAP. Compared with using the DCN or MSDA modules alone, the joint use of the two modules not only achieves a balance in precision and recall, but also performs well in both F1-score and mAP, suggesting that the DCN and MSDA modules are functionally complementary, and together they enhance the model's ability to detect in complex environments. Especially in the face of object detection tasks under flooding and rainfall weather conditions, this combination can more comprehensively capture the morphological and scale changes of the target site and ensure the accuracy and reliability of the detection results.

Although on the surface, the performance improvement on certain datasets may seem small (e.g., 0.5%–1.5%), in real-world application scenarios, especially when performing critical tasks in flooded and rainy environments, this improvement may mean a significant reduction in missed detection cases, thus greatly improving the success rate of emergency rescue operations. In addition, through the ablation experiments, we found that the unique advantages that each of the DCN and MSDA modules brings to the model, together, result in a qualitative leap in the robustness and accuracy of the model under complex meteorological conditions.

For the results of the three groups of modular ablation experiments shown in Fig. 9, this study performs a visual comparative analysis in terms of both feature representation capability and detection efficacy. By comparison, it is easy to find that the adoption of the DCN module alone (column a) has a higher accuracy rate, but some missed detections occur. Comparatively, although the strategy of applying the MSDA module alone (column b) improves in reducing leakage detection, it does not reach the optimum in terms of accuracy. Relative to the results in columns a, b, when the two modules are applied together in the model (column c) is able to better detect vehicle targets in floods while maintaining accuracy. By combining the advantages of multi-scale data enhancement techniques and deformable convolutional networks, this combination

demonstrates excellent performance in the task of target detection in complex environments and provides new ideas and methods for subsequent research.

Figure 9: Qualitative results of ablation experiments. **(a)** the baseline model integrates only the deformable convolutional network (DCN) module. **(b)** independently adopts the multiscale feature aggregation (MSDA) module. **(c)** synergistically fuses the dual-module architecture of DCN and MSDA

6 Future Work

Continuous rainfall is accompanied by cloudy days, which are prone to low-light environments that may affect the target detection results. Although the improved model proposed in this paper demonstrates advantages in complex meteorological conditions such as flooding and rainy days, its performance in extremely low-light environments still needs to be further optimised. The next work needs to develop image enhancement algorithms specifically for low-light conditions, such as deep learning-based denoising and luminance adjustment methods, to improve the quality of the input image and thus indirectly enhance the detection accuracy of the model. Convolutional neural network architectures that are more adapted to low-light environments, such as the introduction of special activation functions or normalisation layers, are investigated and designed to better handle low-contrast and high-noise image data.

In addition, although this paper extends the UA-DETRAC dataset (named UA-Rain) by a synthetic technique to simulate rainfall weather conditions and validate the effectiveness of the model, there may be discrepancies between the synthetic data and the real-world data, which affects the performance of the model in real-world applications. The follow-up task requires the development of effective cross-domain adaptation strategies to enable the model to achieve better generalisation capabilities between synthetic and real data.

7 Conclusion

In this study, a YOLOv9 optimisation model based on improved DCN and MSDA mechanisms is proposed, aiming to address the challenges of image blurring, dynamic noise interference, and scale variability faced by traditional object detection algorithms under flood and rainfall weather conditions. By integrating the adaptive sampling grid adjustment strategy, feature amplitude modulation technique, and multi-scale

contextual information capturing method, the robustness and accuracy of the model under severe weather conditions are significantly enhanced. The experimental results show that compared with Mask R-CNN, Faster R-CNN, Cascade R-CNN and the original YOLOv9 on the ISE-UFDS and UA-DETRAC datasets, the optimised model significantly improves several evaluation metrics, such as precision, recall, F1-score, and mAP, and in particular, it demonstrates stronger performance on the simulated rainfall weather dataset set shows stronger robustness and generalisation ability, providing a reliable technical guarantee for disaster emergency response. Future work will focus on improving the model performance under extreme low-light conditions and exploring more diversified data enhancement strategies.

Acknowledgement: Not applicable.

Funding Statement: This study was financially supported by the National Key R&D Program of China (No. 2022YFC3090603) and R&D Program of Beijing Municipal Education Commission (No. KZ202211417049). The authors are grateful to the anonymous reviewers for their comments and valuable suggestions.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Pengfei Wang and Jiwu Sun; methodology, Pengfei Wang and Jiwu Sun; software, Lu Lu; validation, Lu Lu, Hongchen Li and Hongzhe Liu; formal analysis, Cheng Xu; resources, Yongqiang Liu; data curation, Hongchen Li; writing—original draft preparation, Pengfei Wang; writing—review and editing, Hongchen Li; visualization, Jiwu Sun; project administration, Lu Lu; funding acquisition, Yongqiang Liu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Tahir NUA, Zhang Z, Asim M, Chen J, ELAffendi M. Object detection in autonomous vehicles under adverse weather: a review of traditional and deep learning approaches. Algorithms. 2024;17(3):103. doi:10.3390/a17030103.
2. Wang L, Qin H, Zhou X, Lu X, Zhang F. R-YOLO: a robust object detector in adverse weather. IEEE Trans Instrum Meas. 2022;72:1–11. doi:10.1109/tim.2022.3229717.
3. Zhang H, Xiao L, Cao X, Foroosh H. Multiple adverse weather conditions adaptation for object detection via causal intervention. IEEE Trans Pattern Anal Mach Intell. 2022;46(3):1742–56. doi:10.1109/TPAMI.2022.3166765.
4. Kumar D, Muhammad N. Object detection in adverse weather for autonomous driving through data merging and YOLOv8. Sensors. 2023;23(20):8471. doi:10.3390/s23208471.
5. Ding Q, Li P, Yan X, Shi D, Liang L, Wang W, 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. 2023;24(10):10749–59. doi:10.1109/tits.2023.3285035.
6. Yang J, Wang J, Li Y, Yao B, Xu T, Lu T, et al. Image deraining algorithm based on multi-scale features. Appl Sci. 2024;14(13):5548. doi:10.3390/app14135548.
7. Zhou N, Deng J, Pang M. Recovering a clean background: a parallel deep network architecture for single-image deraining. Pattern Recognit Lett. 2024;178(6):153–9. doi:10.1016/j.patrec.2024.01.006.
8. Wang H, Xie Q, Zhao Q, Li Y, Liang Y, Zheng Y, et al. RCDNet: an interpretable rain convolutional dictionary network for single image deraining. IEEE Trans Neural Netw Learn Syst. 2023;35(6):8668–82. doi:10.1109/TNNLS.2022.3231453.
9. Gao W, Zhang Y, Long W, Cui Z. A deraining with detail-recovery network via context aggregation. Multimed Syst. 2023;29(5):2591–601. doi:10.1007/s00530-023-01116-8.

10. Yu Y, Yang W, Tan YP, Kot AC. Towards robust rain removal against adversarial attacks: a comprehensive benchmark analysis and beyond. In: Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022 Jun 18–24; New Orleans, LA, USA. doi:10.1109/cvpr52688.2022.00592.
11. Qiu Y, Lu Y, Wang Y, Jiang H. ARODNet: adaptive rain image enhancement object detection network for autonomous driving in adverse weather conditions. Opt Eng. 2023;62(11):118101. doi:10.1117/1.oe.62.11.118101.
12. Zhou H, Jiang F, Lu H. SSDA-YOLO: semi-supervised domain adaptive YOLO for cross-domain object detection. Comput Vis Image Underst. 2023;229(4):103649. doi:10.1016/j.cviu.2023.103649.
13. Liu W, Ren G, Yu R, Guo S, Zhu J, Zhang L. Image-adaptive YOLO for object detection in adverse weather conditions. Proc AAAI Conf Artif Intell. 2022;36(2):1792–800. doi:10.1609/aaai.v36i2.20072.
14. 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. 2023;79(23):102484. doi:10.1016/j.displa.2023.102484.
15. Zhang Q, Xiao J, Tian C, Lin J, Zhang S. A robust deformed convolutional neural network (CNN) for image denoising. CAAI Trans Intell Technol. 2023;8(2):331–42. doi:10.1049/cit2.12110.
16. Quan Y, Chen Y, Shao Y, Teng H, Xu Y, Ji H. Image denoising using complex-valued deep CNN. Pattern Recognit. 2021;111:107639. doi:10.1016/j.patcog.2020.107639.
17. Zhang Y, Li K, Li K, Sun G, Kong Y, Fu Y. Accurate and fast image denoising via attention guided scaling. IEEE Trans Image Process. 2021;30:6255–65. doi:10.1109/tip.2021.3093396.
18. Ren C, He X, Wang C, Zhao Z. Adaptive consistency prior based deep network for image denoising. In: Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021 Jun 20–25; Nashville, TN, USA. doi:10.1109/cvpr46437.2021.00849.
19. Wang CY, Yeh IH, Liao HYM. YOLOv9: learning what you want to learn using programmable gradient information. In: Proceedings of the European Conference on Computer Vision; 2024 Sep 29–Oct 4; Milan, Italy. doi:10.1007/978-3-031-72751-1_1.
20. Bhatt D, Patel C, Talsania H, Patel J, Vaghela R, Pandya S, et al. CNN variants for computer vision: history, architecture, application, challenges and future scope. Electronics. 2021;10(20):2470. doi:10.3390/electronics10202470.
21. Zhu X, Hu H, Lin S, Dai J. Deformable convnets v2: more deformable, better results. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15–20; Long Beach, CA, USA. doi:10.1109/cvpr.2019.00953.
22. Jiao J, Tang YM, Lin KY, Gao Y, Ma AJ, Wang Y, et al. DilateFormer: multi-scale dilated transformer for visual recognition. IEEE Trans Multimed. 2023;25:8906–19. doi:10.1109/tmm.2023.3243616.
23. Wen L, Du D, Cai Z, Lei Z, Chang MC, Qi H, et al. UA-DETRAC: a new benchmark and protocol for multi-object detection and tracking. Comput Vis Image Underst. 2020;193(9):102907. doi:10.1016/j.cviu.2020.102907.
24. Huang SC, Hoang QV, Le TH. SFA-Net: a selective features absorption network for object detection in rainy weather conditions. IEEE Trans Neural Netw Learn Syst. 2022;34(8):5122–32. doi:10.1109/TNNLS.2021.3125679.
25. Zou Z, Chen K, Shi Z, Guo Y, Ye J. Object detection in 20 years: a survey. Proc IEEE. 2023;111(3):257–76. doi:10.1109/jproc.2023.3238524.
26. 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. 2016;39(6):1137–49. doi:10.1109/TPAMI.2016.2577031.
27. Cai Z, Vasconcelos N. Cascade R-CNN: delving into high quality object detection. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018 Jun 18–23; Salt Lake City, UT, USA. doi:10.1109/cvpr.2018.00644.